Programa de Pós-graduação em Engenharia de Produção e Sistemas

PPPRO

PERFIL MOTIVACIONAL DE INVENTORES DE PATENTES: UM MODELO CLASSIFICADOR APLICADO NO CONTEXTO DE UNIVERSIDADES PÚBLICAS BRASILEIRAS

Carlos Silva de Jesus

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Produção e Sistemas do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Cefet/RJ, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Produção e Sistemas.

Orientadores:

Profª. Cristina Gomes de Souza, D.Sc.

Prof. Douglas de Oliveira Cardoso, D.Sc.

RIO DE JANEIRO/RJ
OUTUBRO/2023

CEFET/RJ
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO

PERFIL MOTIVACIONAL DE INVENTORES DE PATENTES: UM MODELO CLASSIFICADOR APLICADO NO CONTEXTO DE UNIVERSIDADES PÚBLICAS BRASILEIRAS

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Produção e Sistemas do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Cefet/RJ, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Produção e Sistemas.

Carlos Silva de Jesus

Banca Examinadora:

Documento assinado digitalmente
gov.br CRISTINA GOMES DE SOUZA
Data: 20/10/2023 12:59:11-0300
Verifique em https://validar.iti.gov.br

Profª. Cristina Gomes de Souza, D.Sc. (Orientadora)

Documento assinado digitalmente
gov.br DOUGLAS DE OLIVEIRA CARDOSO
Data: 20/10/2023 11:21:30-0300
Verifique em https://validar.iti.gov.br

Prof. Douglas de Oliveira Cardoso, D.Sc. (IPT - Coorientador)

ALEXANDRE DE CARVALHO CASTRO:92958109704
Assinado de forma digital por ALEXANDRE DE CARVALHO CASTRO:92958109704
Dados: 2023.10.23 06:26:09 -03'00'

Prof. Alexandre de Carvalho Castro, D. Sc.

Documento assinado digitalmente
gov.br MARCO ANTONIO BARBOSA BRAGA
Data: 22/10/2023 15:54:34-0300
Verifique em https://validar.iti.gov.br

Prof. Marco Antonio Barbosa Braga, D. Sc.

Documento assinado digitalmente
gov.br MIRIAN PICININI MEXAS
Data: 20/10/2023 12:53:03-0300
Verifique em https://validar.iti.gov.br

Profª. Mirian Picinini Méxas, D. Sc. (UFF)

Documento assinado digitalmente
gov.br RAFAEL LIMA DE CARVALHO
Data: 20/10/2023 12:38:23-0300
Verifique em https://validar.iti.gov.br

Prof. Rafael Lima de Carvalho, D. Sc. (UFT)

RIO DE JANEIRO/RJ
OUTUBRO/2023

Ficha catalográfica elaborada pela Biblioteca Central do Cefet/RJ.

J58 Jesus, Carlos Silva de.
Perfil motivacional de inventores de patentes : um modelo classificador aplicado no contexto de universidades públicas brasileiras / Carlos Silva de Jesus. – 2023.
197 f.: il. color.

Tese (Doutorado) Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, 2023.
Bibliografia: f. 175-184.

Orientadores: Cristina Gomes de Souza, Dra. [e] Douglas de Oliveira Cardoso, Dr.

1. Engenharia de produção. 2. Pesquisadores – Motivação no trabalho. 3. Patentes. 4. Colaboração acadêmico-industrial - Brasil. 5. Propriedade intelectual. I. Souza, Cristina Gomes de (orient.). II. Cardoso, Douglas de Oliveira (orient.). III. Título.

CDD 338.9

Cefet/RJ - Sistema de Bibliotecas

Ficha catalográfica elaborada por: Samantha Andrade da Rosa – CRB7-5930

DEDICATÓRIA

Dedico este trabalho àqueles que dão especial contorno e sentido à sua realização.

Aos professores e orientadores: generosidade e referência.

Pais: colunas perpétuas.

Elaine: companhia que ilumina e alivia.

Deus...

"...Porque um menino nos nasceu, um filho se nos deu, e o principado está sobre os seus ombros, e se chamará o seu nome: Maravilhoso, Conselheiro, Deus Forte, Pai da Eternidade, Príncipe da Paz..." Isaías 09:27

AGRADECIMENTOS

A DEUS, digno de todo louvor e gratidão pois... “em Ti confiam os que conhecem o Teu nome...” Salmo 9:10

Aos meus pais pelo amor que sempre se traduziu em dedicação e carinho.

À minha família pela presença que fortalece e incentivos que ajudam a seguir.

À minha amada esposa e companheira de sonhos, desafios e aventuras, cujo brilho dos olhos me ilumina e o sorriso me encanta.

Às instituições acadêmicas e profissionais pelas quais passei e que proporcionaram experiências de grande aprendizado e formação.

À professora Vera Lúcia Izidoro, primeira e forte referência acadêmica.

À direção e equipe de profissionais acadêmicos e administrativos do Cefet/RJ, em especial aos colaboradores do PPPRO da DIPPG.

Aos meus orientadores, Profª. Cristina Gomes de Souza e Prof. Douglas de Olivera Cardoso cuja generosidade, dedicação e confiança foram indispensáveis nesta jornada.

Ao Jonatas Halliday Sant'anna do Nascimento pela colaboração no tratamento dos dados da pesquisa.

Aos professores e amigos Jurair Rosa de Paula Junior e Laura Silva de Assis pelo incentivo.

Aos professores Alvaro Chrispino e Roberta Dalvo Pereira da Conceicão pelas oportunas provocações intelectuais.

À direção e aos companheiros da UnED Petrópolis do Cefet/RJ. Em especial à Laice, Michele, Tadeu e Ueliton pelo apoio.

Às instituições e pesquisadores que compuseram a amostra deste trabalho, a quem esperamos ter trazido alguma contribuição para destacar a valiosa contribuição para a sociedade.

RESUMO

PERFIL MOTIVACIONAL DE INVENTORES DE PATENTES: UM MODELO CLASSIFICADOR APLICADO NO CONTEXTO DE UNIVERSIDADES PÚBLICAS BRASILEIRAS

Carlos Silva de Jesus

Orientadores: Profª. Cristina Gomes de Souza, D.Sc.

Prof. Douglas de Oliveira Cardoso, D.Sc.

O efeito das motivações de pesquisadores nos resultados das pesquisas e na consequente atividade de patenteamento carece de aprofundamento e maior sistematização. Esse tema ganha relevância no contexto brasileiro em que as universidades públicas são protagonistas no depósito de patentes. O objetivo deste estudo é desenvolver um modelo classificador a fim de subsidiar a identificação do perfil motivacional de inventores de patentes, relacionando suas motivações com o desenvolvimento de pesquisas e patenteamento. São realizadas revisões da literatura para identificar variáveis e fatores motivacionais, bem como conceitos, abordagens e métodos, a partir dos quais foi elaborado um *framework* teórico. Do levantamento na base de dados de patentes *Derwent* foi definida uma amostra composta pelas 12 universidades públicas mais bem ranqueadas pelo número de depósito de patentes e selecionados os seus 450 pesquisadores com pelo menos cinco pedidos de patentes. A amostra foi submetida a um questionário, além de coletadas as informações acadêmicas na plataforma de currículos *Lattes*, gerando um *dataset* com dados sociodemográficos, acadêmicos e sobre a atividade de patenteamento. Além da análise estatística foi realizada uma modelagem baseada em um *framework* teórico que forneceu parâmetros para o desenvolvimento de um algoritmo de classificação a fim de subsidiar e automatizar a identificação de perfis. O perfil obtido serviu à validação do modelo classificador proposto e ofereceu um panorama sociodemográfico e da atividade patentária dos inventores de patentes no Brasil. A amostra se revelou composta por pesquisadores de viés empreendedor, que reconhecem o valor e papel da estrutura e pesquisa universitária, mas apreciam a interação com a indústria. Têm baixa interação com o meio privado no processo de desenvolvimento da pesquisa, mas nutrem expectativas quanto ao licenciamento. São motivados majoritariamente por fatores intrínsecos à atividade de pesquisa, revelando que incentivos econômicos são menos relevantes comparados à superação de desafios intelectuais. São identificados como o meio institucional a que são vinculados e têm boa resposta a seus estímulos, sendo sensíveis a incentivos e orientações. O estudo busca oferecer um modelo de sistematização e diagnóstico, especialmente útil aos gestores de políticas públicas em inovação. Fornece também parâmetros para reformulação de práticas e incentivos institucionais na promoção das atividades de patenteamento. Além de oferecer um perfil de uma amostra engajada na atividade de patenteamento, o estudo amplia as fronteiras de discussão quanto aos fatores motivacionais afeitos à temática. São disponibilizados, por fim, o *dataset* com os dados originais, o modelo de questionário e os códigos-fontes desenvolvidos e aplicados na pesquisa.

Palavras-Chave: Inovação; Motivação; Patentes; Universidades, Árvore de Decisão; Aprendizado de Máquina.

RIO DE JANEIRO/RJ
OUTUBRO/2023

ABSTRACT

MOTIVATIONAL PROFILE OF PATENT INVENTORS: A CLASSIFICATION MODEL APPLIED IN THE CONTEXT OF BRAZILIAN PUBLIC UNIVERSITIES

Carlos Silva de Jesus

Advisors: Profª. Cristina Gomes de Souza, D.Sc.

Prof. Douglas de Oliveira Cardoso, D.Sc.

The effect of researchers' motivations on research results and the consequent patenting activity needs to be studied in greater depth and systematized. This topic gains relevance in the Brazilian context, where public universities are leading players in patent filing. The aim of this study is to develop a classification model to support the identification of the motivational profile of patent inventors, relating their motivations to the development of research and patenting. Literature reviews were conducted to identify variables and motivational factors, as well as concepts, approaches and methods, on the basis of which a theoretical framework was developed. From the survey of the Derwent patent database, a sample was defined comprising the 12 highest ranked public universities in terms of the number of patent applications, and its 450 researchers with at least five patent applications were selected. The sample was subjected to a questionnaire and academic information was collected from the Lattes curriculum platform, generating a dataset with socio-demographic, academic and patenting activity data. In addition to the statistical analysis, modeling was conducted based on a theoretical framework that provided parameters for the development of a classification algorithm in order to subsidize and automate the identification of profiles. The profile obtained served to validate the proposed classification model and provided a sociodemographic overview of the patenting activity of patent inventors in Brazil. The sample consisted of researchers with an entrepreneurial bias, who recognize the value and role of the university structure and research, but appreciate interaction with industry. They have low levels of interaction with the private sector in the research development process, but have expectations regarding licensing. They are mostly motivated by factors intrinsic to the research activity, revealing that economic incentives are less relevant than overcoming intellectual challenges. They are identified with the institutional environment to which they are affiliated and respond well to its incentives, being sensitive to incentives and guidance. The study seeks to offer a systematization and diagnostic model, especially useful for managers of public innovation policies. It also provides parameters for reformulating institutional practices and incentives to promote patenting activities. In addition to providing a profile of a sample engaged in patenting activity, the study broadens the boundaries of discussion on the motivational factors related to the subject. Finally, the dataset with the original data, the questionnaire model and the source codes developed and applied in the research are made available.

Keywords: Innovation; Motivation; Patents; Universities, Decision Tree; Machine Learning.

RIO DE JANEIRO/RJ
OCTOBER/2023

Lista de Figuras

Figura 1- Delineamento de pesquisa. 29
Figura 2 - Plano e estrutura de pesquisa. 32
Figura 3 - Sistematização da estratégia de busca 50
Figura 4 - *WordCloud* com a ocorrência de termos. 51
Figura 5 - Tipologia de inventores 59
Figura 6 - Engajamento da atividade patentária 61
Figura 7 - Esquema de desenvolvimento do *framework* 64
Figura 8 - Ênfase nas dimensões na aferição do perfil do pesquisador 67
Figura 9 - *Framework* de aferição das motivações na atividade patentária. 72
Figura 10a - *Framework* no perfil individual 76
Figura 10b - *Framework* no perfil de grupo 76
Figura 11 - Arquitetura básica de *Web Scraping* 83
Figura 12 - Algoritmo *Web Scraping* em *Python* 85
Figura 13 - Etapas de coletas dos dados 86
Figura 14 - Processo de coleta dos dados 91
Figura 15 - Representação das fases do *Knowledge Discovery in Databases* (KDD). 123
Figura 16 - Fluxo *KDD* da pesquisa. 124
Figura 77 - Algoritmo genérico para construção de árvores de decisão 146
Figura 88 - Jogar tênis ou não? Um modelo de árvore de decisão 147
Figura 99 - Sequência do processo da validação cruzada. 149
Figura 20 - Perfil motivacional não balanceado da parcela respondente da amostra. 156
Figura 21 - Perfil motivacional balanceado da parcela respondente da amostra. 158
Figura 22 - Perfil motivacional da parcela não respondente da amostra. 159

Lista de Quadros

Quadro 1 - Questões, objetivos e métodos da pesquisa. 28

Quadro 2 - Principais informações, resultados e conclusões dos artigos. 50

Quadro 3 - Fatores motivacionais identificados na literatura selecionada 57

Quadro 4 - Dimensões estruturantes, variáveis e categorias do *framework* 66

Quadro 5 - Descrição das categorias do *framework* 67

Quadro 6 - Indicadores para aferição do perfil do pesquisador 71

Lista de Tabelas

Tabela 1 - Resultados da aplicação de filtros de busca 44

Tabela 2 - Descrição dos estudos incluídos na revisão (n= 41) 53

Tabela 3 - *Ranking* de universidades públicas por pedidos de registros de patentes. 88

Tabela 4 - Amostra por intervalo de nascimento, gênero e local de nascimento. 97

Tabela 5 - Fatores motivacionais por ordem de importância. 114

Tabela 6 - Dados originais da natureza motivacional sem a padronização 131

Tabela 7 - Dados da natureza motivacional balanceados via *z-score* 132

Tabela 8 - Dados ranqueados da natureza motivacional 132

Tabela 9 - Dados balanceados via mínimo-máximo da natureza motivacional 133

Tabela 10 - Dados da relação entre normas e valores, sem o balanceamento. 135

Tabela 11 - Dados da relação entre normas e valores balanceados via *z-score.* 136

Tabela 12 - Dados ranqueados da relação entre normas e valores 137

Tabela 13 - Dados mm da relação entre normas e valores 137

Tabela 14 - Atributos binários convertidos para variáveis numéricas 141

Tabela 15 - Conversão do atributo binário relativo a gênero. 141

Tabela 16 - Apuração da acurácia obtida nos atributos alvos. 153

Tabela 17 - Classificação não balanceada da parcela respondente da amostra. 155

Tabela 18 - Classificação balanceada da parcela respondente da amostra. 157

Tabela 19 - Classificação da parcela não respondente da amostra. 159

Tabela 20 - Medidas de avaliação da classificação da parcela não respondente. 160

Lista de Gráficos

Gráfico 1 - Internacionalização educacional da amostra..... 98
Gráfico 2 - Distribuição por áreas de pesquisa 99
Gráfico 3 - Números de pedidos de patentes. 99
Gráfico 4 - Gráfico de dispersão da produção acadêmica 100
Gráfico 5 - Gráfico função de distribuição cumulativa empírica da produção acadêmica..... 101
Gráfico 6 - Correlação entre indicadores de patenteamento e publicação científica 102
Gráfico 7 - Local de nascimento = local do bacharelado?..... 103
Gráfico 8 - Local do bacharelado = local do mestrado? 103
Gráfico 9 - Local do mestrado = local do doutorado? 104
Gráfico 10 - Fluxo de migração acadêmica de pesquisadores (n=131). 106
Gráfico 11 - Fluxo de migração acadêmica de pesquisadores (n=450). 107
Gráfico 12 - Dados do processo de patenteamento..... 108
Gráfico 13 - Percepção em relação à colaboração universidade-indústria..... 109
Gráfico 14 - Distribuição de importância do grupo fator Extrínseco/Econômico 114
Gráfico 15 - Distribuição de importância do grupo fator Misto/Reputação 115
Gráfico 16 - Distribuição de importância do grupo fator Intrínseco/Desafio..... 115
Gráfico 17 - Distribuição de importância comparativo dos grupos de fatores..... 116
Gráfico 17 - Distribuição dos dados originais da natureza motivacional sem padronização.... 131
Gráfico 18 - Distribuição dos dados balanceados da natureza motivacional via *z-score*..... 132
Gráfico 19 - Distribuição dos dados ranqueados quanto à natureza motivacional 133
Gráfico 20 - Dados da natureza motivacional balanceados via mínimo-máximo..... 134
Gráfico 21 - Dados da relação entre normas e valores, sem o balanceamento. 136
Gráfico 22 - Distribuição dos dados da relação entre normas e valores via *z-score*. 136
Gráfico 23 - Distribuição dos dados ranqueados quanto à relação entre normas e valores.... 137
Gráfico 24 - Distribuição dos dados mm quanto à relação entre normas e valores..... 138
Gráfico 25 - Conversão em contagem por intervalo de ano de nascimento. 142
Gráfico 26 - Representação gráfica da acurácia obtida nos atributos alvos..... 154

Lista de Abreviaturas

BDA	*Bayh-Dole Act*
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CIP	Classificação Internacional de Patentes
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COVID	*Corona Virus Disease*
CSV	*Comma-Separated Values*
CT&I	Ciência, Tecnologia e Inovação
DCBD	Descoberta de Conhecimento em Base de Dados
DII	*Derwent Innovations Index*
ETT	Escritório de Transferência de Tecnologia
HTML	*Hyper Text Markup Language*
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
INPI	Instituto Nacional da Propriedade Industrial
ISI	*Institute for Scientific Information*
KDD	*Knowledge Discovery in Databases*
MDIC	Ministério do Desenvolvimento, Indústria e Comércio Exterior
NIT	Núcleo de Inovação Tecnológica
P&D	Pesquisa e Desenvolvimento
PSF	*Python Software Foundation*
RCI	Relação Ciência-Indústria
STD	*Self-Determination Theory*
TAD	Teoria da Autodeterminação
UFC	Universidade Federal do Ceará
UFMG	Universidade Federal de Minas Gerais
UFPB	Universidade Federal da Paraíba
UFPE	Universidade Federal de Pernambuco
UFPR	Universidade Federal do Paraná
UFRGS	Universidade Federal do Rio Grande do Sul
UFRJ	Universidade do Rio de Janeiro
UFRJ	Universidade Federal do Rio de Janeiro
UFS	Universidade Federal de Sergipe
UFSC	Universidade Federal de Santa Catarina
UFV	Universidade Federal de Viçosa
UNICAMP	Universidade Estadual de Campinas
USP	Universidade de São Paulo

SUMÁRIO

Capítulo I – INTRODUÇÃO 16
I.1 Contextualização do tema 16
I.2 Justificativa e relevância 21
I.3. Questões de pesquisa 24
I.4 Objetivos da Pesquisa 25
I.4.1 Objetivo Geral 25
I.4.2 Objetivos Específicos 25
I.5. Originalidade e Contribuição 25
I.6 Delineamento metodológico 28
Capítulo II – REVISÃO SISTEMÁTICA 35
II.1. *Background* 35
II.1.1. Cooperação ciência-indústria 35
II.1.2. Atividade patentária 39
II.2 Revisão sistemática da literatura 42
II.2.1 Estratégia de Busca 42
II.2.2 Critérios de inclusão e exclusão 44
II.2.3. Caracterização dos estudos 51
II.2.4. Identificação dos fatores motivacionais 54
II.2.5 Identificação dos perfis motivacionais 57
II.3 Construção do *framework* 63
II.3.1. Definição de dimensões, variáveis e categorias 64
III.3.2. Relação entre fatores e variáveis 69
II.4. Considerações finais do capítulo 77
CAPÍTULO III – METODOLOGIA PARA CONSTRUÇÃO DO *DATASET* 81
III.1 Conceituações básicas 81
III.1.1 A técnica *Web Scraping* 82
III.1.2 A linguagem *Python* 84
III.2 Construção do *dataset* 85

II.2.1 Definição da amostra 85
III.2.2 Coleta dos dados 90
III.2.2.1 Coleta através da Plataforma *Lattes* 92
III.2.2.2 Coleta através de questionário online 93
III.2.3 Estruturação do *dataset* 95
CAPÍTULO IV – PERFIL SOCIODEMOGRÁFICO, PATENTEAMENTO E FATORES MOTIVACIONAIS DOS PESQUISADORES BRASILEIROS 96
IV.1 Perfil sociodemográfico 96
IV.2 Posição quanto ao patenteamento 108
IV.3 Posição quanto aos fatores motivacionais 112
CAPÍTULO V – DESENVOLVIMENTO E APLICAÇÃO DO MODELO DE CLASSIFICAÇÃO DO PERFIL MOTIVACIONAL 120
V.1 Considerações sobre a amostra e organização dos dados 120
V.2 O processo de descoberta de conhecimento em base de dados 122
V.3 Pré-processamento dos dados 125
V.3.1 Parâmetros e marcadores para categorização da amostra 126
V.3.2 Redimensionamento de dados para a categorização 128
V.3.2.1 Balanceamento da camada Natureza motivacional 130
V.3.2.2 Balanceamento da camada relação entre normas/ valores pessoais 135
V.3.3 Conversão das variáveis categóricas em numéricas 139
V.4 A mineração de dados 142
V.4.1 O algoritmo de árvore de decisão 143
V.4.2 A avaliação do modelo classificador 148
V.5 Perfil motivacional 154
V.5.1. Classificação dos respondentes (n=131) 155
CAPÍTULO VI – REVISÃO E CONSIDERAÇÕES FINAIS 163
VI.1. Revisão dos objetivos e questões de pesquisa 163
VI.2. Originalidade das contribuições 169
VI.3. Principais desafios e limitações 171

VI.4. Possibilidades de avanços em estudos futuros 172

REFERÊNCIAS 175

APÊNDICE A – Questionário *on-line* 185

APÊNDICE B – Termo Circunstanciado de Livre Esclarecimento (TCLE) 190

APÊNDICE C – Repositório do *dataset* 193

APÊNDICE D – Códigos-fonte 194

Capítulo I – INTRODUÇÃO

I.1 Contextualização do tema

Estudos na temática de propriedade intelectual e inovação têm destacado o crescimento dos pedidos de depósito de patentes pelas universidades (ZHOU, 2018; BACKS *et al*., 2019). Como agentes e ambientes de interação da ciência, tecnologia e inovação (CT&I), as universidades têm nos artigos científicos e, mais recente e crescentemente, no depósito de patentes, meios de divulgação científica e indicadores de transferência de tecnologia (GUSMÃO, 2002; MUELLER e PERUCCHI, 2014, GAO *et al*.,2014).

O patenteamento está no escopo da literatura sobre *intellectual property right theory*, em que Haase *et al.* (2005, p. 357) destacam a proteção e exploração comercial de invenções nas universidades, concluindo que “as patentes se mostram como um instrumento de proteção efetivo que oferece possibilidades múltiplas para a transferência de conhecimento e tecnologia”.

Efetivamente, nos últimos anos, as universidades vêm sendo incentivadas a assumir uma atuação mais empreendedora, bem como, contribuir para o desenvolvimento social, econômico e cultural da região onde atuam (COMPAGNUCCI e SPIGARELLI, 2020; HARRISON e TUROK, 2017). Essa atuação vem se firmando como a terceira missão das universidades (COMPAGNUCCI e SPIGARELLI, 2020; BACKS *et al*., 2019), consolidando a importância dessas instituições para promoção da inovação e desenvolvimento tecnológico (DEGL’INNOCENTI *et al*., 2019; FREEL *et al*., 2019).

As universidades – produtoras e disseminadoras de conhecimento – têm sido reconhecidas como catalisadoras da inovação e promotoras do desenvolvimento tecnológico (DEGL’INNOCENTI *et al*., 2019; FREEL *et al*., 2019). O tema é analisado por Skute *et al*. (2019) e também por Nicotra *et al.* (2021) como resultado do empreendedorismo acadêmico, temática trazida por Etzkowitz e Zhou (2017) à luz do modelo de Tríplice Hélice.

O modelo de tríplice hélice remonta a um cenário de interação mútua dos elementos presentes no trinômio universidade, indústria e governo. É nos espaços das três hélices (universidade, indústria e governo) que se configura um processo dinâmico que

conjuga conhecimento, consenso e inovação (ETZKOWITZ e ZHOU, 2017). Neste ambiente de relação ciência-indústria (RCI) as universidades públicas têm assumido papel singular também na alocação de recursos públicos e na atração do capital privado, estimulando ao desenvolvimento de parcerias econômicas e científicas e mediando interesses e créditos nos benefícios da pesquisa (BEAUDRY e KANANIAN, 2013; ERIKSON *et al*.,2015; IRYNA e PODGORNA, 2018).

Observa-se, assim, o aumento significativo do patenteamento das universidades ao redor do mundo (ZHOU, 2018; BACKS *et al*., 2019). Trata-se de um tópico complexo que varia de acordo com cada país, instituição e área do conhecimento. Neste contexto, as universidades têm envidado esforços para criar um ambiente institucional favorável e sensível a tal movimento (ROTHAERMEL *et al*., 2007; NICOTRA *et al*., 2021).

As patentes têm três principais finalidades: proteger a propriedade intelectual; servir de fonte de tecnologia da informação e, finalmente, servir como indicador de ciência, tecnologia e inovação (LOWE; 2006; GARNICA e TORKOMIAN, 2009; BARROS e SOUZA, 2010).

No Brasil, a Lei Nº 9.279, de 14 de maio de 1996 estabelece direitos e obrigações relativos à propriedade industrial. Criado em 1970, cabe ao Instituto Nacional da Propriedade Industrial (INPI), autarquia federal vinculada ao Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC), a aplicação dos dispositivos da referida lei, concedendo e garantindo os direitos de propriedade industrial no Brasil. Dois são os tipos de patentes, a de invenção e o de modelo de utilidade com validades de 20 e 15 anos, respectivamente. Para ser patenteável, ainda de acordo com a lei brasileira da propriedade industrial, é necessário o atendimento dos requisitos de novidade, não anterior a divulgação de resultados; resultar de atividade inventiva, não se valendo de resultados óbvios; e ter aplicação industrial (BRASIL, 1996).

Ainda no cenário brasileiro, a Lei da Inovação - Lei nº. 10.973/2004 (posteriormente alterada pelo Código Nacional de Ciência, Tecnologia e Inovação - Lei nº 13.243/2016) atendeu à necessidade de maior estímulo à interação das universidades com o meio produtivo na medida em que determinou que as instituições científicas e tecnológicas instituíssem Núcleos de Inovação Tecnológica (NIT). A tais núcleos ou similares cabe, conforme o texto legal, a gestão da política institucional de inovação visando à

prospecção, articulação e negociação da produção científica acadêmica, sobretudo no que tange “à proteção das criações, licenciamento, inovação e outras formas de transferência de tecnologia” (BRASIL, 2004). A Unicamp, precursora nos NITs com a agência Inova Unicamp, é o caso clássico nacional de quanto é valido às universidades o investimento em patentes e no licenciamento de novas tecnologias via centros e núcleos de assistência à pesquisa (PESQUISA FAPESP, 2016).

Iniciativas legais têm estimulado o cenário acadêmico, resultando no aumento do número de pedidos de patentes de universidades brasileiras. O setor acadêmico brasileiro respondeu por 19,5% do total de pedidos de PCT (Tratado de Cooperação em Matéria de Patentes) com prioridade brasileira durante o período 2002-2012 (SILVA *et al*., 2017). As universidades têm se destacado entre os principais solicitantes de patentes residentes no INPI com seis instituições no topo do *ranking* dos 10 maiores cessionários no período 2000-2014.

Quanto aos pedidos de patentes de invenção feitos por residentes no Brasil, dados mais recentes, divulgados pelo Instituto Nacional de Propriedade Industrial, referentes ao ano de 2019 (INPI, 2020), mostram a predominância das universidades brasileiras. Dentre os 20 maiores depositantes de patentes, 17 são universidades. A Petrobras, que é a empresa com maior quantidade de depósitos de patentes do país, figura na quinta colocação desse *ranking*, atrás de quatro universidades, cada uma delas com mais de 60 depósitos realizados em 2019.

Conforme citado por Fischer *et al*. (2019), o registro de patentes por universidades em países desenvolvidos é marginalmente relevante (cerca de 4% nos EUA e 6% na Europa), enquanto no Brasil essas instituições assumem um papel estratégico. Um dos fatores que contribuem para esse fenômeno é que a maioria dos pesquisadores (aproximadamente 70%) que trabalham com pesquisa e desenvolvimento (P&D) no país está nas universidades (MCT&I, 2023). Esse cenário também difere da realidade vivenciada nos países desenvolvidos, onde grande parte desses pesquisadores trabalha na iniciativa privada (MCT&I, 2023).

Apesar dos avanços observados e do aumento do número de depósitos de patentes nos últimos anos, o processo de transferência de tecnologia entre universidade e empresa no Brasil ainda está em construção. Conforme destacado por Paranhos *et al*.

(2018), o país adotou políticas para estimular a relação universidade-empresa com bastante atraso. Muitos NITs são recentes e ainda estão em fase de estruturação.

Há outros desafios a serem superados: necessidade de mecanismos que permitam um processo adequado de interação entre universidade-empresa (RAPINI, 2007); alto nível de burocracia nas universidades públicas (FREITAS *et al*., 2013); alta rotatividade de pessoal e falta de profissionais qualificados para atuar nos NITs, especialmente para tratar de questões relacionadas à transferência de tecnologia (TORKOMIAN, 2009; QUERIDO *et al*., 2011; PARANHOS *et al*., 2018); falta de recursos financeiros (RYAN, 2014); resistência e conflitos de culturas e demandas envolvendo o relacionamento entre universidades e empresas (CLOSS *et al*., 2013, PARANHOS *et al*., 2018; LEITE *et al*., 2019); imaturidade das tecnologias geradas pelas universidades, em geral não submetidas às etapas de prova de conceito e prototipagem (PARANHOS *et al*., 2018, LEITE *et al*., 2019). Outro aspecto trazido por Garcez e Moreira (2017) é o elevado *backlog* (tempo entre o depósito e a concessão da patente) no Brasil, que também desestimula o patenteamento e dificulta as negociações devido à falta de segurança jurídica quanto à concessão da proteção almejada.

Esse cenário resulta em uma baixa eficácia dos pedidos de patentes feitos por universidades no Brasil. Muitos pedidos de patentes são abandonados, não concedidos e não explorados comercialmente (QUERIDO *et al*., 2011). O licenciamento de tecnologias ainda é limitado segundo Paranhos *et al*., (2018), fazendo com que os inventores não tenham incentivo suficiente para despender esforços adicionais no desenvolvimento de novas tecnologias que possam ser comercializadas (JENSENE e THURSBY, 2001).

Com o objetivo de melhorar esse cenário, um novo marco legal - conhecido como Código de Ciência, Tecnologia e Inovação - entrou em vigor recentemente, com o objetivo de reduzir os obstáculos legais e burocráticos e dar maior flexibilidade às instituições brasileiras que atuam nesse sistema (RAUEN, 2016). Os efeitos dessa nova regulamentação, entretanto, só serão conhecidos no futuro.

Contudo, de acordo com Póvoa e Rapini (2020), o protagonismo das universidades brasileiras no sistema de patentes é incomum e consiste em um interessante e oportuno objeto de estudo. O aumento da quantidade de patentes por parte das universidades brasileiras tem despertado o interesse da academia. Por exemplo:

Assunção (2000) e Barbastefano *et al*. (2010) apresentaram um panorama dos depósitos das universidades brasileiras. Castro *et al*. (2007) abordaram como a Agência Inova da Unicamp produz e dissemina as informações tecnológicas geradas na universidade. Amadei e Torkomian (2009) também focaram no patenteamento das universidades públicas paulistas, enquanto Serafini *et al.* (2011) se concentraram nas universidades do Nordeste do Brasil. Querido, Lage e Vasconcellos (2011) criaram um índice para avaliar o desempenho das universidades e respectivos núcleos de inovação tecnológica tomando como base a concessão das patentes em relação ao número de depósitos.

Em estudos mais recentes, Soares, Torkomian e Nagano (2020) exploraram como a qualidade dos regulamentos universitários e o desenvolvimento econômico regional impactam nos novos pedidos de patentes e acordos de licenciamento. Dias e Kannebley Junior (2021) analisaram como o desempenho científico e o patenteamento são influenciados pelos diferentes estilos de organização dos laboratórios de pesquisa das universidades. Fischer *et al*., (2019) abordaram a evolução da colaboração universidade-indústria sob uma perspectiva do desenvolvimento tecnológico, ressaltando o papel das universidades nesse contexto. Viana *et al*. (2018) exploraram a transferência de tecnologia entre universidade-empresa com foco na propriedade intelectual e instrumentos legais voltados para esse tipo de transferência de tecnologia. Ainda versando sobre o patenteamento das universidades brasileiras, Cavalheiro *et al*., (2019) analisaram especificamente o portfólio de patentes da Universidade Federal Fluminense. Conforme pode ser observado, diversos estudos têm sido realizados analisando diferentes aspectos do patenteamento em universidades do país.

Nesse sentido, Tijssen (2018) sinaliza que o patenteamento das universidades é resultado de um conjunto de fatores individuais, institucionais e contextuais. No caso do presente estudo, o foco são os pesquisadores. Eles são os principais atores desse processo, sendo necessário compreender o seu perfil e as motivações que os levam a patentear seus inventos (MUELLER e PERUCCHI 2014; GAO *et al*. 2014).

Os pesquisadores, cujo perfil é tão diverso quanto suas formações, motivações e influências, mantêm vínculo com universidades e dialogam com a iniciativa privada cujos interesses comerciais, como analisa Póvoa (2010), nem sempre estão alinhados

aos objetivos dos ambientes acadêmicos, que, financiados por recursos públicos, têm o condão de gerar benefícios e prestar contas à sociedade.

Devido ao emergente protagonismo dos pesquisadores tanto na captação de recursos quanto na escolha do objeto de análise e meio de divulgação e, principalmente, em definir a qual interesse atender no desenvolvimento da pesquisa, estudos contemporâneos têm ressaltado a relevância de mapear e conhecer o perfil e motivações dos pesquisadores das universidades no Brasil (OLIVEIRA e VELHO, 2010; SOUZA *et al*.;2011; PERUCCHI e MUELLER, 2014; CATIVELLI e LUCAS, 2016) e no mundo (LAM, 2011; EJERMO e KÄLLSTRÖM, 2016; BURHAN *et al*., 2017; BLIND *et al*., 2018).

I.2 Justificativa e relevância

Apesar dos ambientes universitários desempenharem o papel legal e social no estímulo, organização e gestão da geração e transmissão do conhecimento científico, a disposição de publicar ou patentear o resultado de uma pesquisa, envolve decisões do pesquisador/inventor em alinhamento com as propostas institucionais.

Deste ponto, como salientam Perucchi e Mueller (2014), emerge o dilema pessoal do pesquisador, ora focalizando os benefícios de visibilidade e prestígio acadêmico na publicação de artigos científicos, ora almejando à oportunidade de requisitar o patenteamento e garantir o domínio da inventividade para si. O referido dilema pode ser considerado superado, conforme relação entre o patenteamento e o número de publicações observada por Stephen *et al.* (2005) e a complementaridade de ambos os fenômenos, como também constatado por Souza *et al.* (2011).

Identificar o perfil e as motivações dos pesquisadores em patentear é importante para melhor compreender o desempenho desses atores (RYAN, 2014; RYAN e BERBEGAL-MIRABENT, 2016) e para estabelecer políticas capazes de estimular o seu engajamento em atividades que resultem em transferência de conhecimento e tecnologia (OLAYA-ESCOBAR *et al*., 2020). Essa é uma temática complexa, segundo Tissen (2018), pois se orienta por valores e contextos diversos com reflexo em comportamentos. Não deve, portanto, ser subestimada (OLAYA-ESCOBAR *et al*., 2020) e precisa ser analisada a partir de uma perspectiva multidimensional.

Na literatura internacional são encontrados diversos estudos com foco nos fatores que influenciam a motivação dos pesquisadores. Embora possa ser identificado um amplo conjunto de fatores, esses estão relacionados e podem ser agrupados em aspectos intrínsecos e extrínsecos, geralmente associados a questões tais como reconhecimento, reputação, satisfação pessoal, cultura organizacional, regulação institucional, infraestrutura disponível, retorno material e financeiro (OWEN-SMITH e POWELL, 2001; ACKERS, 2005; BALDINI *et al*., 2007; MOUTINHO *et al*., 2007; BLIND e GAUCH, 2009; D'ESTE e PERKMANN, 2010; SAUERMANN e COHEN, 2010; OLIVEIRA, 2011; BOLLI e SCHLÄPFER, 2013; BURHAN *et al*., 2017; CHANG *et al*., 2017; SCHMID e WANG, 2017; e outros).

Por sua vez, outros estudos buscam analisar e propor categorias de perfis de pesquisadores que variam conforme suas motivações e comportamentos. Como exemplos podem ser citados os trabalhos de Stephan e Levin (1992); Stephan e Everhart (1998); Deci e Ryan (2000); Ryan e Deci, (2000a; 2000b); Gagné e Deci (2005); Göktepe (2008); Blind e Gauch (2009); Göktepe-Hulten e Mahagaonkar (2009); D'Este e Perkman (2011); Lam (2011); Bolli e Schläpfer (2013); Bentley *et al*. (2015); Tijssen (2018) e Blind *et al. (*2018).

Em se tratando do contexto brasileiro, também existem alguns estudos voltados para identificação dos fatores que influenciam no patenteamento e características dos pesquisadores que figuram como inventores das patentes de universidades no país. Dentre esses se têm os trabalhos desenvolvidos por Barbieri (1999); Pinheiro (2001); Oliveira e Velho (2009); Oliveira (2011); Souza *et al.* (2011); Moura (2012); Closs *et al*. (2013); Perucchi e Mueller (2014); Cativelli e Lucas (2016); Pereira *et al*., (2019); Oliveira *et al*. (2020).

Barbieri (1999) analisou o papel do inventor independente ou individual, identificando suas tipologias, filiações e superação de desafios em ações de P&D, concluindo que os inventores brasileiros carecem de incentivos e têm motivações difusas. Pinheiro (2001) tratou do perfil, motivações, dificuldades e o comportamento dos inventores-empreendedores. A pesquisa teve abrangência local, destacando os inventores de uma feira anual de novas tecnologias em Porto Alegre/RS sem considerar o ambiente das instituições de pesquisas públicas.

Analisando, ainda que introdutoriamente, o perfil dos inventores acadêmicos, Oliveira e Velho (2009) buscaram ampliar a compreensão da atividade de patenteamento nas universidades públicas paulistas, destacando as principais características de seus inventores. O trabalho de Oliveira (2011), ainda que tenha focalizado o universo das universidades do Estado de São Paulo, é um dos estudos mais completos no Brasil que tem no seu escopo as patentes e o perfil motivacional dos inventores com um profundo e consistente embasamento teórico e histórico.

Moura (2012) buscou conhecer as relações de ciência e tecnologia a partir da coautoria e coinvenção, bem como, a motivação para pesquisar, identificando variações de motivos e interesses conforme a área de desenvolvimento da pesquisa. Closs *et al*. (2013) também investigaram as relações universidade-indústria a fim de identificar o que motiva e o que desencoraja os pesquisadores acadêmicos a transferir tecnologia para as empresas a partir de um estudo envolvendo a PUCRS, PUC-Rio, USP e IFMG.

Os trabalhos de Souza *et al.* (2011) e Perucchi e Mueller (2014), abordaram o perfil dos pesquisadores das universidades brasileiras com foco em características objetivas a partir de dados disponíveis na Plataforma Lattes. Cativelli e Lucas (2016) buscaram caracterizar as invenções oriundas de patentes nas universidades públicas brasileiras. Quanto ao perfil dos pesquisadores, a análise se deteve na área de formação, atividades desempenhadas, quantidades e códigos da Classificação Internacional de Patentes (CIP) dos documentos na base do INPI. Além de constatar o crescimento do número de depósitos de patentes e o protagonismo de instituições paulistas, o estudo ressalta a maturidade dos inventores como nível na carreira e desempenho acadêmico.

Pereira *et al*., (2019), por sua vez, buscaram identificar os fatores que estimulam e inibem a produção de patentes nas universidades brasileiras a partir da percepção dos núcleos de inovação tecnológica. Em um estudo recente, Oliveira *et al*. (2020) apresentaram os principais fatores associados à motivação dos pesquisadores para patentear. Esse estudo, entretanto, foi restrito a duas universidades do Estado de Pernambuco, além de não ter como foco identificar uma tipologia de perfil motivacional dos pesquisadores.

Com base na revisão da literatura sobre o assunto, observa-se que os estudos que versam sobre o tema no país são escassos. Alguns tencionam traçar o perfil dos pesquisadores considerando características mais objetivas, tais como formação,

produtividade, vínculos institucionais, fontes de financiamento, linhas de pesquisa e dados geográficos. Os que buscam discorrer sobre as motivações, por sua vez, têm o foco em análises mais pontuais ou restritas a uma pequena amostra de pesquisadores ou com base na percepção de outros atores que não os próprios pesquisadores. Não obstante a relevância e contribuição de todos esses estudos, observa-se que ainda existe um grande campo de pesquisa a ser explorado em relação à temática em tela no Brasil.

Têm-se, assim, uma lacuna acadêmica, como apontado por Souza *et al.* (2011), Perucchi e Mueller (2014), Cativelli e Lucas (2016), de estudos cujo escopo dê conta da identificação do perfil social, comportamental e motivacional dos pesquisadores, como também sinalizado por Stephan *et al.* (2005). Por fim, uma análise da atividade de patenteamento a partir das motivações do inventor patenteador, como aponta Souza *et al.* (2011), ainda resta não atendida, além da ausência de verificação de como tais motivações afetam os resultados das pesquisas desenvolvidas e a consequente atividade de patenteamento.

I.3. Questões de pesquisa

Diante da diversidade de estudos sobre os fatores que motivam os pesquisadores a patentearem os resultados de suas pesquisas e dos perfis desses pesquisadores, conforme apontado na literatura internacional, surge a primeira questão de pesquisa:

Q1. *Como o conhecimento sobre esse tema pode ser organizado e sistematizado de modo a relacionar os fatores que influenciam no patenteamento com os perfis dos pesquisadores que atuam nas universidades?*

Considerando que os pesquisadores lidam com contextos institucionais diversos e que suas motivações e comportamento são influenciados por vários fatores intrínsecos e extrínsecos, emerge a segunda questão de pesquisa:

Q2. *Quais fatores influenciam e qual o perfil dos pesquisadores das universidades públicas brasileiras?*

Considerando que é importante conhecer o perfil motivacional dos pesquisadores a fim de se ter subsídios para uma melhor definição de políticas públicas e políticas institucionais mais apropriadas para promover o engajamento desses atores nas atividades de patenteamento, tem-se a terceira questão de pesquisa:

Q3. *Como identificar as motivações e o perfil dos pesquisadores das universidades?*

São essas três questões de pesquisa que norteiam o presente trabalho e que direcionam a definição dos objetivos e delineamento das escolhas metodológicas.

I.4 Objetivos da Pesquisa

I.4.1 Objetivo Geral

O objetivo geral da pesquisa é desenvolver um modelo que permita identificar e classificar o perfil motivacional de inventores patenteadores, considerando a relação das suas motivações com o desenvolvimento de pesquisas e patentes a partir de dados coletados, empregando técnicas e ferramentas computacionais.

I.4.2 Objetivos Específicos

Busca-se, assim, o atendimento dos seguintes objetivos específicos e complementares:

- Construir um *framework* de modo a organizar e sintetizar os fatores que influenciam na motivação para patentear, bem como, relacionar esses fatores com os perfis dos pesquisadores das universidades.
- Identificar os inventores de patentes das universidades públicas brasileiras e levantar suas características e os fatores que influenciam na motivação para patentear.
- Estruturar e aplicar uma metodologia baseada em ferramentas computacionais na composição de um modelo classificador que permita a identificação do perfil dos inventores de patentes.

I.5. Originalidade e Contribuição

A análise que visa identificar e compor o perfil motivacional de pesquisadores considera a capacidade inventiva e arcabouço intelectual do indivíduo como um importante insumo da pesquisa que pretende entender e propor soluções de interesse social (BEAUDRY e KANANIAN, 2013; BOLLI e SCHLÄPFER; 2013; BOH *et al*., 2014 e; LAWSON e STERZI 2014; ZI e BLIND, 2015; ZEMTSOV *et al*., 2016; SVAGZDIENE e KUKLYTE, 2016; IRYNA e PODGORNA, 2018).

A hipótese da presente investigação é que as motivações determinam o perfil do pesquisador e influenciam a atividade de pesquisa acadêmica e, consequentemente, o processo de patenteamento. São, portanto, passíveis de identificação e classificação

na medida em que orientam os pesquisadores cujo comportamento tem efeitos no desenvolvimento e resultado das pesquisas, sobretudo as patenteáveis pelos centros acadêmicos e no cumprimento das suas finalidades no processo de inovação tecnológica.

Conforme mencionado anteriormente, a literatura internacional apresenta estudos sobre fatores motivacionais e perfis de pesquisadores. Faz-se necessária uma consolidação dos conteúdos desses trabalhos. Essa literatura, por vezes, se apresenta fragmentada e dispersa, abordando contextos pontuais e específicos. Nesse sentido, a construção de um *framework* que é uma representação abstrata de conceitos e métodos representa um avanço ao buscar sintetizar e integrar o conhecimento existente sobre o assunto.

Em se tratando da realidade brasileira, têm-se uma lacuna na literatura nacional quando comparada à tendência internacional de melhor conhecer o perfil e entender as motivações dos pesquisadores em patentear os resultados de suas pesquisas. Souza *et al.* (2011), Perucchi e Mueller (2014) e Cativelli e Lucas (2016) reforçam e existência dessa lacuna apontando serem necessários mais estudos cujo escopo dê conta da identificação quanto ao perfil social, comportamental e motivacional dos pesquisadores como também aponta Stephan *et al.* (2005).

É no atendimento de tais lacunas que residem a relevância e originalidade da presente pesquisa, que se desenvolve objetivando ampliar e aprofundar a investigação dos trabalhos citados, incrementando os dados e considerando outras variáveis a fim de identificar e relacionar as motivações pessoais com a atividade patentária, elaborando, enfim, um pretendido perfil motivacional de inventores patenteadores.

De forma complementar, a estruturação de uma metodologia com aplicação de ferramentas computacionais que permita identificar o perfil dos pesquisadores, desenvolvendo um modelo classificador baseado em algoritmo de árvore de decisão também reforça a originalidade e contribuição desse trabalho.

Cumpre destacar que o emprego de ferramentas computacionais leva em conta que as informações pessoais e acadêmicas coletadas, em parte, se originam de produções científicas pulverizadas em meios eletrônicos de registros disponibilizados pelos próprios pesquisadores e, no caso das patentes, em bases de dados eletrônicas disponíveis para consulta. Outra parte resulta da aplicação de instrumentos de

pesquisa, sobretudo, levantamento de dados via questionário (Apêndice A). Assim, ainda quanto à originalidade e ineditismo da presente pesquisa, tem-se a aplicação de ferramentas computacionais também na coleta, armazenamento e, principalmente, no tratamento dos dados disponibilizados/ coletados. No escopo do trabalho, considerada a natureza qualitativa/ quantitativa dos dados e o seu volume visando à definição de um perfil comportamental/motivacional, dá-se, ao fim o desenvolvimento de um modelo de classificação via algoritmo classificador baseado em árvore de decisão.

Na literatura não foram identificados trabalhos similares que tenham adotado uma metodologia que permitisse fazer a identificação do perfil dos pesquisadores de forma automatizada a partir da inserção de dados. Embora para efeito deste estudo, o modelo classificador seja aplicado e validado na identificação do perfil dos pesquisadores que patenteiam nas universidades brasileiras, sua aplicação pode ser reproduzida em outras universidades, públicas ou não, nacionais ou no exterior.

Portanto, além da contribuição teórica visando maior compreensão conceitual sobre essa temática, a presente pesquisa também traz como resultado um produto - o modelo subsidiado por um algoritmo classificador em si e seus instrumentos de coleta de dados - que podem ser aplicados em outros contextos institucionais, fornecendo parâmetros e referências para a elaboração de políticas mais assertivas com potencial de promover maior engajamento dos pesquisadores em atividades de patenteamento e transferência de tecnologia.

Afinal, conforme apontado por Faria *et al*. (2018), pouca atenção tem sido dada ao funcionamento interno das universidades no Brasil em termos de incentivo aos seus recursos humanos, notadamente aos professores universitários. Segundo os autores, existem poucas análises formais, por exemplo, sobre como as universidades podem melhor motivar o desempenho do corpo docente nas áreas de pesquisa e desenvolvimento.

Portanto, o ineditismo e contribuição da pesquisa estão, respectivamente, em preencher as lacunas apontadas e, adicionalmente, em disponibilizar modelos, ferramentas e base conceitual e de dados para estudos visando à atração e retenção de recursos humanos nas universidades, além de oferecer variáveis e conhecimento para o aprimoramento da atuação dos NITs em mediar e promover o avanço das pesquisas acadêmicas.

I.6 Delineamento metodológico

Uma pesquisa apresenta elementos estruturantes que dão sustentação ao desenvolvimento do estudo, indicando o percurso a ser seguido até o alcance dos objetivos propostos. Esses elementos são: as questões de pesquisa; os objetivos; e os métodos, que devem se configurar de forma articulada, coerente e consistente. O Quadro 1 apresenta esses elementos que são a base da presente pesquisa.

Questão de pesquisa	Objetivo específico	Método
Como o conhecimento sobre esse tema pode ser organizado e sistematizado de modo a relacionar os fatores que influenciam no patenteamento com os perfis dos pesquisadores que atuam nas universidades?	Construir um framework de modo a organizar e sintetizar os fatores que influenciam na motivação para patentear, bem como, relacionar esses fatores com os perfis dos pesquisadores das universidades.	Revisão sistemática da literatura
Quais fatores influenciam e qual o perfil dos pesquisadores das universidades brasileiras?	Identificar os inventores de patentes das universidades públicas brasileiras e levantar suas características e os fatores que influenciam na motivação para patentear.	Levantamento na base de patente *Derwent*; *Web Scraping* na base de currículos Lattes; *Survey* com aplicação de questionário *on-line*;
Como identificar as motivações e o perfil dos pesquisadores das universidades?	Estruturar e aplicar uma metodologia baseada em ferramentas computacionais na composição de um modelo classificador que permita a identificação do perfil dos inventores de patentes.	Modelagem baseada e algoritmo de classificação

Quadro 1 - Questões, objetivos e métodos da pesquisa.

Fonte: Elaborado pelo autor.

No que se refere à tipologia do presente estudo, pode-se dizer que o ambiente de pesquisa é institucional tendo o enfoque de observação no inventor, ou conjunto de inventores, ligado a diferentes instituições de pesquisa. Assim, a finalidade da pesquisa é aplicada já que pretende contribuir para o aprimoramento do ambiente de pesquisa universitário.

Para alcançar os objetivos propostos, conforme esboçado na Figura 1, pretende-se adotar o seguinte delineamento de pesquisa: a) natureza exploratória e descritiva; b) abordagem qualitativa com tratamento quantitativo dos dados coletados; e c) estratégias de investigação múltiplas abrangendo revisão sistemática da literatura,

levantamento em bases de patentes, *web Scraping* na base de currículos *Lattes*, *survey* com aplicação de questionários *on-line* e modelagem baseada em árvore de decisão com a estruturação de um algoritmo de classificação.

A investigação se dará em duas etapas de abordagens, inicialmente com caráter exploratório ao levantar informações e proporcionar esclarecimento e delimitação ao tema, permitindo o mapeamento do cenário de patentes e pesquisadores vinculados às universidades. Importa em preparação para a segunda etapa, descritiva que, por sua vez, pretende realizar o levantamento e tratamento dos dados que culminarão na caracterização das motivações e perfil dos inventores patenteadores (SEVERINO, 2007; MARCONI e LAKATOS, 2010; GIL, 2011).

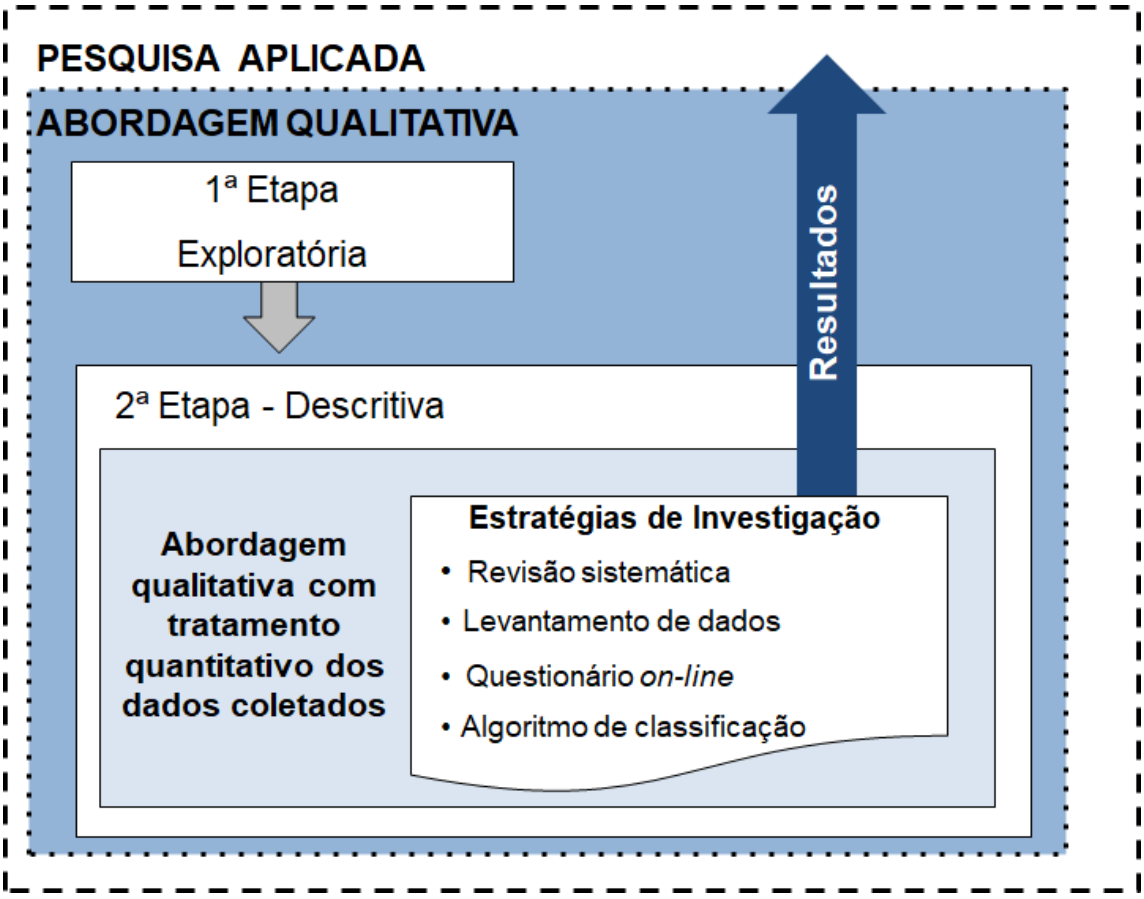


Figura 1- Delineamento de pesquisa.
Fonte: Elaborado pelo autor.

O presente estudo, que busca identificar, compreender e tecer entendimento sobre motivações dos pesquisadores está identificado com o que Severino (2007) prefere tratar como abordagem qualitativa, considerando que o termo abordagem faz melhor referência a fundamentos epistemológicos e menos a especificidades metodológicas. Martins (2006) caracteriza a avaliação qualitativa como própria à descrição,

compreensão e interpretação de fatos e fenômenos. Tal tratamento concorda com a definição de método dada por Lakatos e Marconi (2003) e Gil (2011), como um caminho para se chegar a um determinado fim, um conjunto de procedimentos intelectuais e técnicos adotados para se atingir o conhecimento, interpretar fenômenos e atribuir significados, como ora se pretende.

Considerando o vasto cenário de pesquisadores e amplo banco de dados pretendidos, é necessário o emprego de tratamento quantitativo com a aplicação de técnicas e ferramentas computacionais a fim de determinar e analisar a amostra que se pretende obter. Para tanto o desenvolvimento da pesquisa, da concepção do seu escopo ao tratamento e análise de dados, demandou a atuação de equipe multidisciplinar com colaboração nas áreas de engenharia de produção e ciência da informação.

Bogdan e Biklen (2010) assinalam que a pesquisa qualitativa se processa, dentre outros, via revisão bibliográfica a partir de pesquisas anteriores, documentos, manuais e informações organizacionais. Tais aspectos servem bem a presente proposta de investigação, pois sugerem a observação de fatos ou fenômenos cujas causas e comportamento se desejam conhecer, buscando compará-los com a finalidade de descobrir as relações existentes entre eles e, finalmente, verificar aspectos que possam servir à caracterização pretendida quanto aos inventores.

Tal metodologia, conjugando componentes qualitativos e quantitativos, de acordo com Gil (2011), permite a familiaridade com o problema e o acesso a gama de fenômenos, sendo mais ampla, facilitando a busca por dados dispersos e outros, históricos, já organizados.

Como comenta Yin (2010), o objeto de pesquisa pode ser um indivíduo, indivíduos, evento ou entidade, cujas decisões, programas ou processos são passíveis de análise. Tal estratégia de pesquisa, de cunho predominantemente descritivo, enfoca uma situação específica que se distingue como única em suas características. Ressalta-se, porém, que se trata de um fenômeno da vida real, um caso específico que serve à abstração necessária à pesquisa.

A apreciação de eventos contemporâneos, ainda segundo Yin (2010), tem incremento de benefício quando associado a outras técnicas que conferem validade e colaboram para o rigor da pesquisa. A conjugação de diferentes fontes de pesquisa potencializa o contato com aspectos históricos e comportamentais, além de minimizar eventuais

problemas de precisão e validade de método e de resultados (MARTINS, 2006; YIN, 2010).

Conforme comentado anteriormente, a estratégia de investigação faz uso de diversos métodos, os quais se encontram detalhados ao longo do trabalho. Cada um dos métodos empregados foi selecionado visando melhor atender a cada um dos objetivos estabelecidos conforme mostrado no quadro 1.

Os trabalhos nos quais a presente pesquisa busca referência metodológica são os desenvolvidos por Blind *et al. (*2018); Lam (2011) e Göktepe (2008), além das provocações acadêmicas oriundas dos estudos de Cativelli e Lucas (2016), Moura (2012), Souza *et al*. (2011) e Owan e Nagaoka (2008).

Além do delineamento e abordagens já explicitadas, a Figura 2 apresenta a estrutura e plano metodológico geral da pesquisa, bem como exibe os desdobramentos e consequente caminho metodológico a fim de alcançar os objetivos em diferentes e sequenciais capítulos do relatório da tese.

O trabalho tem em seu relatório a disposição em capítulos que configuram a estruturação em eixos temáticos que, por vezes, podem assumir subdivisões em formato comum em artigos científicos, modelo adotado por considerar a vantagem de organização do material visando à publicação dos resultados da pesquisa.

Salienta-se que o plano metodológico do trabalho é único, porém os capítulos podem conter metodologia própria, nele descrita e com resultados relativos à respectiva fase de pesquisa, contribuindo, portanto, na parte para o alcance dos objetivos específicos e, no todo, concorrendo para a obtenção do objetivo geral da pesquisa.

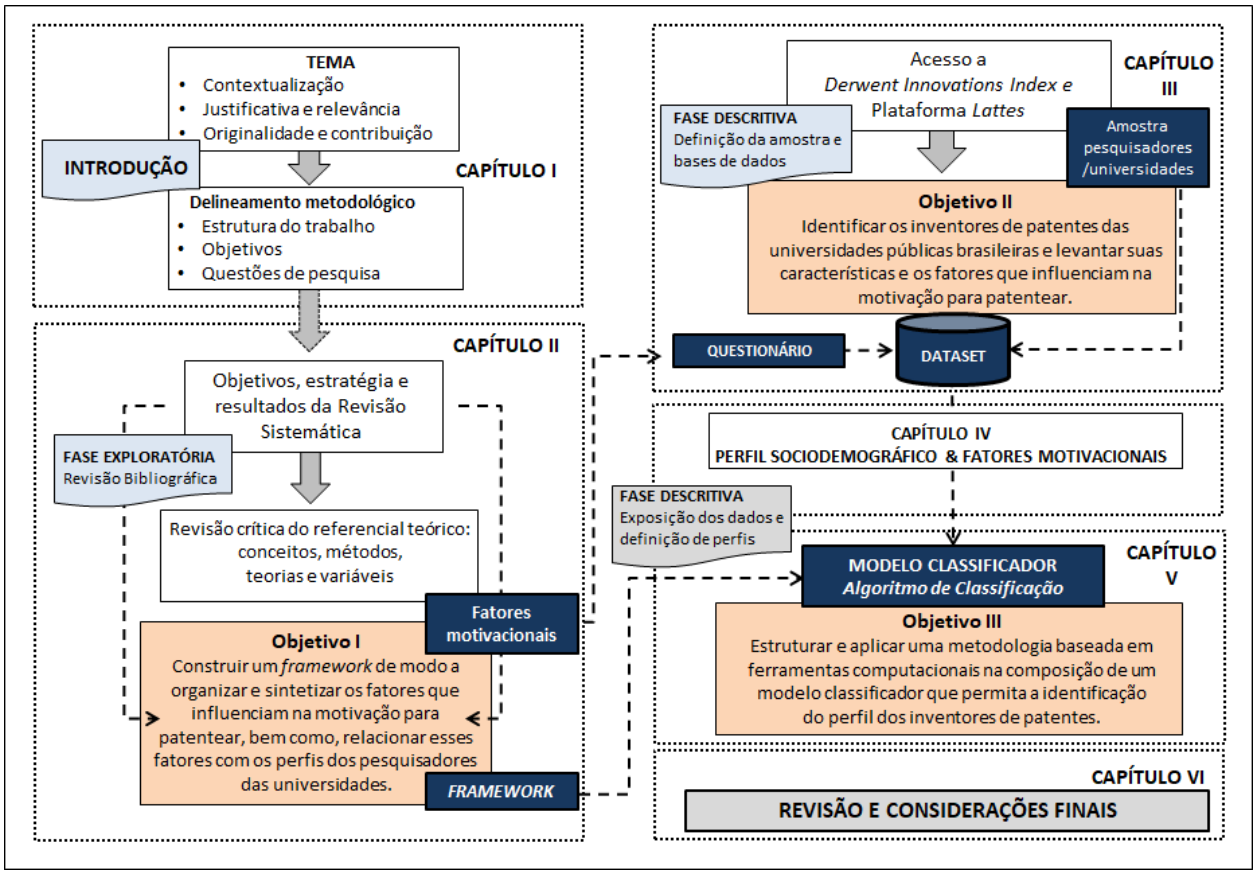


Figura 2 - Plano e estrutura de pesquisa.
Fonte: Elaborado pelo autor.

Registra-se que o projeto de pesquisa contendo a proposta, objetivos, metodologia, instrumentos de pesquisa e público-alvo foi submetido à apreciação do Comitê de Ética em Pesquisa em Seres Humanos do Centro Universitário Arthur Sá Earp Neto/ Faculdade de Medicina de Petrópolis/ Hospital de Ensino Alcides Carneiro – UNIFASE, indicado pela Plataforma Brasil. O projeto detalhado juntamente com o Termo de Conhecimento Livre e Esclarecido (Apêndice B) foram submetidos em 22/11/2019 via Plataforma Brasil tendo o parecer final número 3.807.130 emitido em 23/01/2020 com a aprovação da pesquisa.

I.7. Estrutura do trabalho

O trabalho se estrutura em seis capítulos. Conforme sinalizado anteriormente, cada capítulo remete a uma parte fundamental do trabalho, sendo de alguns apresentam objetivos, metodologias e resultados próprios. Embora cada capítulo se apresente como uma parte independente, se pretende um inter-relacionamento entre eles de

modo a se alcançar o objetivo geral do trabalho, bem como, responder as questões de pesquisa.

O **primeiro capítulo**, em tela, introduz e justifica o tema de pesquisa, apresentando as motivações do estudo e sua relevância, originalidade e condições de ineditismo. Aponta ainda as questões de pesquisa que orientam os objetivos, escopo e metodologia, definindo as fronteiras conceituais e estruturais do relatório de pesquisa.

O **segundo capítulo**, de caráter predominantemente conceitual, se dá no desenvolvimento de revisões da literatura, pretendendo alcançar o primeiro objetivo de pesquisa em colecionar e revisar artigos científicos indexados em bases de dados visando à construção do *framework* teórico. Organiza, discute e relaciona os conceitos coletados e obtêm definições, metodologias, teorias e outras informações pertinentes à proposta e, principalmente, apresenta as variáveis e fatores motivacionais que, concertados, culminam no desenvolvimento de um modelo analítico (*framework*) em que se baseia a aplicação das soluções estatísticas e computacionais de classificação da amostra.

O **terceiro capítulo** expõe a metodologia para consolidar o universo e a amostra de pesquisa consoantes aos objetivos pretendidos. Conceitua as técnicas computacionais aplicadas nessa etapa de acesso e coleta dos dados de patentes e patentadores. Finalmente discorre sobre a elaboração e aplicação dos instrumentos a que a amostra é exposta visando à geração de um *dataset*.

O **quarto capítulo** parte das informações consolidadas no *dataset* para apresentar o perfil sociodemográfico de pesquisadores brasileiros, apresentando ainda o posicionamento da amostra quanto ao patenteamento e a atribuição de importância aos fatores motivacionais.

O **quinto capítulo** expõe os caminhos e estratégias para, à luz do *framework* teórico, desenvolver e aplicar um modelo de classificação de perfis motivacionais. Para tanto são empreendidas ações de organização, tratamento e correlação dos dados coletados com a aplicação das soluções computacionais na elaboração de um algoritmo de classificação.

No **sexto capítulo** é promovida a consolidação, revisão e considerações sobre os principais resultados. Busca considerar e atender às questões e aos objetivos de

pesquisa para então apresentar as contribuições, desafios, limitações e possibilidades de avanços em estudos futuros.

Capítulo II – REVISÃO SISTEMÁTICA

O presente capítulo apresenta a revisão sistemática e a estruturação do *framework. Buscando* cumprir o primeiro objetivo de pesquisa, organiza e sintetiza os fatores motivadores ligados ao patenteamento, relacionando-os com o perfil dos pesquisadores envolvidos em atividades patentárias. O capítulo inicia com considerações em torno das relações entre ciência-indústria e as atividades patentárias, visando introduzir o contexto no qual a pesquisa se desenvolve. Na sequência é apresentada a metodologia empregada na revisão sistemática com a descrição da estratégia de busca e dos critérios de inclusão e exclusão das publicações, bem como, a identificação dos fatores e perfis motivacionais encontrados na literatura. O capítulo finaliza com a elaboração da proposta do *framework*.

II.1. *Background*

A temática dessa pesquisa está inserida em um contexto que é afetado pelas relações de cooperação ciência-indústria (também chamadas de relações universidade-empresa ou academia-empresa), bem como, pelas atividades patentárias no ambiente dessas organizações.

II.1.1. Cooperação ciência-indústria

Além de seu tradicional papel na formação de pessoas (ensino) e geração de conhecimento (pesquisa), a transferência de conhecimento para a indústria e a sociedade vem se firmando como a terceira missão das universidades (BACKS *et al*, 2019; COMPAGNUCCI e SPIGARELLI, 2020). Esse é um fenômeno global (JONGBLOED *et al*., 2008) que está alinhado com o modelo da tríplice hélice que, segundo Etzkowitz e Zhou (2017), representa as três pás que interagem entre si mutuamente, combinando elementos presentes no trinômio universidade, indústria e governo. No contexto da inovação, ainda segundo os autores, “a tríplice hélice ‘organizada’ é uma ferramenta institucional invisível de crescimento econômico regional e de desenvolvimento social” (ETZKOWITZ e ZHOU, 2017, p. 30). Este cenário reflete a visão de desenvolvimento apontada há décadas por Schumpeter (1934) para quem a inovação se dá como resultado de um processo de gerar novas combinações a partir de ideias, recursos e forças existentes em interação (CROITORU, 2012; BOH *et al*. 2014).

Tal modelo de interação fica mais evidente à luz do potencial trazido na esteira do conceito Indústria 4.0, assim denominada a quarta revolução industrial com aplicação de novas tecnologias aos meios produtivos, onde a interação e a conexão são promotoras e também resultado do desenvolvimento, tendo por base recursos financeiros, econômicos, científicos e humanos que, para Lomachynska e Podgorna (2018), se traduz no potencial de inovação.

O modelo tríplice hélice tem sido analisado e utilizado em iniciativas em países como Suécia, Holanda, Estados Unidos, México, Canadá, China, Singapura e Coreia do Sul, dentre outros (GAO *et al*. 2014; SINGH *et al*. 2015; STEK e VAN GEENHUIZEN, 2015; SVAGZDIENE e KUKLYTE, 2016; STEZANO, 2018).

A relação ciência-indústria recebe novo contorno no conceito do empreendedorismo baseado no conhecimento, tema de estudo realizado por Moutinho *et al.*(2016), que considera a materialização da interação: pela cooperação em projetos ou consórcios de pesquisa; no desenvolvimento de empresas tipo *spin-off* que resultam de pesquisas de uma empresa originária ou derivada de atividades da pesquisa acadêmica; e na transferência de direitos de propriedade intelectual, incluindo patenteamento e licenciamento.

As diferentes modalidades de transferência de conhecimento dão conta que um número crescente de pesquisadores não colabora apenas entre si. Frequentemente seus vínculos se estendem à indústria, atraídos especialmente pela alta tecnologia ou condições de financiamento. Dentre os benefícios desta relação, Beaudry e Kananiam (2013) analisam que para a indústria há a redução dos custos em P&D e a apropriação de uma sólida base científica disponível nas universidades com vistas à inovação. Para o acadêmico, por sua vez, o financiamento privado permite uma maior exploração de áreas antes não planejadas e então demandadas pelo ambiente privado, permitindo assim uma exposição pessoal a um cenário que permite a aplicação de suas pesquisas, além da possibilidade de ampliar sua rede profissional e expor seus resultados. Para a universidade, os principais benefícios institucionais colecionados por Moutinho *et al. (*2016) são: a) obtenção de fundos para estudantes, sobretudo de pós-graduação; b) acesso ou aquisição de equipamentos específicos para pesquisas; c) agregar experiência em modelos de negócios; d) desenvolvimento de produtos e

serviços comercializáveis; e) captar recursos visando a expansão das atividades de pesquisa.

A presença de capital privado nas academias levanta a polêmica de considerá-las como prestadoras de serviço de valor agregado. Aqui se evoca a necessidade da mediação na transferência de conhecimento e tecnologia, em inglês, *Knowledge Technology Transfer (KTT)*, que resulta das interações do empreendedor acadêmico com as organizações privadas. Para Halilem *et al. (*2017), a transferência de recursos do mercado privado para o ambiente público universitário pode se concretizar mediante um processo de comercialização formal ou licenciamento, gerando receita direta para a universidade, ou num caminho informal, tanto de conhecimento quanto de recursos, via consultorias realizadas que são serviços de base individual que buscam resolver problemas concretos e têm o pesquisador como o maior beneficiário (MOUTINHO *et al*. 2016).

Quanto à interferência ou influência do capital privado nas pesquisas universitárias e na mediação da relação entre pesquisador e o mercado privado, como percebido por ainda por Halilem *et al. (*2017), há de se dosar o nível de controle pela entidade de pesquisa sobre a atuação do pesquisador. Neste sentido, Oliver e Sapir (2017) destacam a alteração na lógica acadêmica tradicional que passa a admitir que a transferência tecnológica decorra naturalmente da atividade universitária, cabendo aos escritórios de incentivos à inovação o despertar do potencial comercial do conhecimento gerado no meio universitário. Nesta direção colaboram Beaudry e Kananian (2013) ao identificar que pesquisadores que mantêm relação com o financiamento privado são mais produtivos e inovadores.

Ainda Oliver e Sapir (2017) refletem sobre o impacto da profissão de cientista ou acadêmico nas finalidades das pesquisas, se orientados pela busca ao conhecimento ou com viés puramente comercial. Os autores analisam que o chamado capitalismo acadêmico, que se vale da lógica de mercado visando o financiamento, pode, no limite, afetar os modos e finalidades do espaço público da pesquisa que deveria se pautar pela busca e compartilhamento do conhecimento. Na esteira no empreendedorismo acadêmico, contudo, apontam que o engajamento de pesquisadores em atividades comerciais, além do esperado aumento da atividade de patenteamento, não tem

necessariamente impacto negativo sobre o desenvolvimento da ciência pelo conhecimento e na atração de pesquisadores considerados tradicionais.

Ressalta-se que o nível de engajamento com a pesquisa, em processos formais e informais de transmissão do conhecimento e tecnologia, se dá menos pelo reconhecimento da propriedade da invenção do que pelo acesso a financiamentos e recursos para a pesquisa, importando, na fronteira, no fenômeno da mobilidade acadêmica que é a migração ou fuga do capital humano – da expressão em inglês *brain drain* (ACKERS, 2005) – para outras instituições ou mesmo regiões geográficas mais desenvolvidas ou próximas de centros econômicos (BOLLI e SCHLÄPFER 2013 e ZEMTSOV *et al*. 2016). Em contraposição, analisando um caso extremo da realidade, Schmid e Wang (2017), analisam o viés negativo do incentivo financeiro chinês em estabelecer cotas e fornecer estímulos financeiros como incentivo à inovação no país que, no primeiro momento, resultou em aumento nominal, mas que se revelou em uma produção científica tida como artificial e de qualidade insustentável.

A pesquisa de base acadêmica equilibrada em seus recursos e incentivos tem permitido que a universidade seja reconhecida pelo o que se denomina terceira missão e também como empreendedorismo acadêmico por Göktepe (2008), Erickson *et al*. (2015) e Singh *et al*. (2015), na medida em que a avaliam como importante ator no desenvolvimento econômico ao desenvolver suas atividades tradicionais de pesquisa e ensino e na colaboração com outras organizações, companhias, associações e indivíduos com foco no desenvolvimento regional.

Na tríplice relação não há protagonismos, mas ênfase na visão integrada e sistêmica. Como analisa Gao *et al*. (2014) as empresas são tomadas como tendo um papel de liderança na inovação já que operam no mercado, sendo a ele mais sensível. A universidade é considerada um apoio intelectual na medida em que proporciona pessoal altamente qualificado e com produção científica que se traduz em conhecimento para a indústria. Há ainda a atuação dos governos que, além do papel próprio de regular e controlar, é agente de estímulo e facilitador dos meios de cooperação.

Dado o crescente destaque da pesquisa universitária e a ampliação das relações da universidade com o meio privado e o interesse governamental em seus resultados, se fez necessária a criação de uma estrutura gestora de inovações que proteja

politicamente o ambiente acadêmico. Em artigo publicado em 1998, Henderson *et al*., considerando os efeitos do *Bayh-Dole Act (BDA)* nos EUA em 1980 que culminou num mecanismo de regulação e indução à produção científica universitária, já observavam que do aumento do patenteamento das pesquisas universitárias americanas decorria uma taxa elevada de transferência tecnológica para o setor privado e, consequentemente, aumento da taxa de retorno social à pesquisa acadêmica.

O *Bayh-Dole Act* se caracterizou numa quebra de paradigma, estabelecendo modos e meios da exploração da pesquisa financiada com recursos públicos. Como destacam Cruz e Souza (2015), tal evento oportunizou a criação e organização de escritórios de transferência tecnológica com vistas a estabelecer políticas institucionais de proteção à propriedade intelectual e facilitação da transferência de tecnologia via licenciamentos. Com o marco legal houve a garantia de proteção da propriedade de patentes pelas universidades como gestoras do investimento público, prática seguida por países europeus, com analisam Ejermo e Källström (2016).

II.1.2. Atividade patentária

Considerando a interação entre as pesquisas científicas e atividades industriais à luz da cooperação ciência-indústria, a atividade patentária, apesar de recente, foi por Gusmão (2002) tomada como uma das modalidades formais de transferência de tecnologia com maior crescimento e com consequências observadas na orientação de políticas de pesquisa e inovação. Para Svagzdiene e Kuklyte (2016), considerando o cenário europeu, os pedidos de patentes por um milhão de habitantes integra o conjunto de fatores que auxilia na identificação do estado atual e futuro da atividade inovadora. Focalizando o mercado privado, Yang e Minutolo (2016) avaliam que as patentes têm papel importante na obtenção de vantagem competitiva ao universo empresarial.

O patenteamento é, aliado à de produção de artigos, empregado como indicador tanto de produtividade científica quanto da contribuição das universidades na produção e transferência de conhecimento. Lewensohn (2015) alega que a sobrevivência de patentes depende das suas características em si, bem como daquelas do seu titular e inventor. Bolli e Schläpfer (2013) apresentam a falta de financiamento como um fator de impacto negativo em pesquisa e, por consequência, na atividade de patenteamento. Os temas da sobrevivência e manutenção das patentes trazem à luz a questão das

taxas cobradas durante o processo de solicitação e, principalmente, o pagamento durante o tempo em que a patente estiver em vigor. Considerando o cenário desfavorável de vários centros públicos de pesquisa, os inventores acabam por depender de financiamentos externos que, no limite, pode atrair preocupações quanto à comercialização e impacto na finalidade da pesquisa e, como comentam Oliver e Sapir (2017), uma alteração na lógica acadêmica pelo eventual tendencioso viés financeiro.

Rememorando a interação entre o universo da pesquisa com financiamento público em associação aos recursos privados, Natalicchio *et al*. (2014) ponderam que a falta de ativos nas organizações públicas de pesquisa é contrabalanceada pela oferta de conhecimento científico e de qualificação humana, recursos altamente atraentes às organizações privadas que, por sua vez, dispõem de diversificação tecnológica, agregando fatores que resultam em qualidades às patentes e, aprimoram, como aponta Beaudry e Kananian (2013), uma rede de colaboração que não se restringe à academia.

Neste sentido, Lawson e Sterzi (2014) evocam a questão da frequência do patenteamento por pesquisadores. Em regra, a maioria dos inventores patenteia apenas uma vez. Da obtenção do primeiro pedido de patente, que resulta em grande medida da qualidade da invenção, depende o sucesso e o relevante impacto quantitativo e qualitativo em patenteamentos futuros. Os classificados como ‘inventores em série’, assim denominados os que têm histórico em patentear, segundo Göktepe (2008), estão em grande medida envolvidos em atividades inventivas e têm vistas ao patenteamento de suas invenções.

Ljungberg *et al*. (2013) distinguem as patentes de origem acadêmicas, oriundas de pesquisas desenvolvidas em universidades com financiamento público e com inventores ligados à pesquisa básica e de longo prazo, das não-acadêmicas que se concentram nas tecnologias com retorno imediato e de curto prazo e com financiamento predominante da iniciativa privada.

Patentes acadêmicas têm número médio maior de citações do que as não-acadêmicas. A citação de patentes, representa relevante indicador de produtividade, prejudicado em grande medida, como destaca Mauleón *et al*. (2014), pela falta de padronização no registro de dados dos titulares e inventores. Os referidos autores destacam problemas

quanto à obtenção de dados de patentes por problemas de registros deficientes ou incompletos. Dados que poderiam aprimorar, pelo exemplo trazido pelos autores quanto à realidade espanhola, análises sobre a baixa representação do gênero feminino no cenário de patentes e da pesquisa em geral daquele país. Liaw *et al*. (2014) destacam a importância dos dados das referências de não-patentes que são documentos apresentados por ocasião do processo de patentes com finalidade de fundamentar a solicitação, indicar fontes e comprovar ineditismo.

O processo longo e demoradas questões que envolvem o processo do patenteamento, além de outras questões de sigilo, são consideradas por Zi e Blind (2015) como influenciadoras da baixa participação de inventores acadêmicos no processo de normalização de invenções.

Barbastefano *et al*. (2010, p.10), considerando a parceria entre universidades e empresas entre 1999 e 2008, verificaram que “a maioria das patentes são depositadas pelos autores em conjunto com as instituições universitárias”. Desse complexo contexto de interação e cooperação de valores e interesses acadêmicos, sociais e econômicos, muitas vezes conflitantes, cabe salientar que a pesquisa acadêmica é desenvolvida, em última análise, com forte protagonismo de indivíduos pesquisadores/inventores.

Os pesquisadores, com perfis diferentes conforme formações, motivações e influências, apesar de vinculados a centros de pesquisa acadêmica, dialogam com o meio privado com interesses, principalmente comerciais, que se contrapõem, como discute Póvoa (2010), aos dos ambientes universitários, que, por sua vez, são financiados por recursos públicos, tendo o condão de gerar benefícios à sociedade.

Apesar dos Escritórios de Transferência de Tecnologia (ETT) desempenharem o papel legal no estímulo, organização e gestão da geração e transmissão do conhecimento científico, a disposição de publicar ou patentear o resultado de uma pesquisa, envolve decisões de cada pesquisador/inventor em alinhamento com as propostas institucionais a que está vinculado. Neste ponto, como salientam Perucchi e Mueller (2014), emerge o dilema pessoal do pesquisador, ora focalizando os benefícios de visibilidade e prestígio acadêmico na publicação de artigos científicos, ora almejando à oportunidade de requisitar o patenteamento e garantir o domínio da inventividade para si. O referido dilema, contudo, não é unanimidade, principalmente pela relação observada por Stephen *et al*. (2005) entre o patenteamento e o número de publicações e a

complementariedade de ambos os fenômenos, como também constatado por Souza *et al*. (2011).

Aqui a discussão supera a escolha do meio de divulgação, dando lugar à presença de uma ambivalência com potencial de orientar e interferir na produção e na efetiva transferência tecnológica. Assim consideram Owen-Smith e Powell (2001) ao destacarem que a decisão por patentear os resultados de pesquisas universitárias resulta de diferentes fatores, como de concepções sobre os benefícios das patentes, do apoio institucional, do papel dos escritórios de transferências tecnológica, da noção de custos de licenciamento, do alinhamento com interesses comerciais, além da recorrente atenção à proteção da propriedade.

A esta altura, o protagonismo dos pesquisadores/inventores ganha maior contorno, sobretudo aqueles vinculados aos centros de pesquisas universitários que participam na captação de recursos - soberanos na escolha do objeto de estudo e tendo forte influência no meio de divulgação, mas, principalmente, em serem portadores de motivações e interesses que influenciam diretamente no desenvolvimento da pesquisa. Daí ressalta-se a necessidade de identificar o perfil e, sobretudo, as motivações dos pesquisadores engajados na atividade patentária.

Diante da diversidade de fatores e perfis encontrados nos diversos estudos publicados é que se desenvolve o presente trabalho, viabilizando o conhecimento de temas, métodos empregados, principais contribuições e lacunas não exploradas. A base teórica da presente investigação tem seu escopo na atividade patentária e nos fatores motivacionais relacionados a temática à luz da cooperação ciência-indústria.

II.2 Revisão sistemática da literatura

Essa seção apresenta os procedimentos e critérios adotados para a seleção dos artigos que integraram a revisão sistemática da literatura, a caracterização desse conjunto de artigos, a identificação dos fatores motivacionais e dos diferentes perfis encontrados na literatura.

II.2.1 Estratégia de Busca

A revisão sistemática da literatura se deu a partir de busca realizada em banco de dados de produções científicas internacionais disponível no endereço eletrônico www.periodicos.capes.gov.br. A base de dados consultada foi a *Web of Science*, eleita

em função da sua abrangência de indexação e característica multidisciplinar de publicações.

O levantamento inicial, visando uma maior familiaridade com a temática, acessou entre 2018 e 2019 publicações referentes ao período 1945-2018 a partir da definição de *strings* de busca formados por termos na língua inglesa aderentes aos temas de pesquisa associados por operadores lógicos booleanos inseridos na pesquisa por tópicos, como a seguir:

- *patent* AND researcher* AND (profile OR characterist* OR factor* OR motivation*)*
- *patent* AND Innovation AND "Intellectual Property" AND "Triple Helix"*
- *"Science-Industry Cooperation" AND (universit* OR patent*)*
- *profile AND patent* AND (researcher* OR inventor*)*

A utilização de diferentes combinações de termos associados a diversos operadores booleanos pretendeu atingir maior número de publicações em alinhamento com os temas de pesquisa e, ao mesmo tempo, estabelecer filtros específicos visando à coleta precisa e qualificada de conteúdo afeito aos temas.

A cada *string* de busca foram aplicados os seguintes filtros de elegibilidade: a) tipo de documento: artigos e revisões; b) período de publicação: considerou o período dos últimos 6 anos a contar desse levantamento inicial, 2013 a 2018 considerando a atualização do tema e o elevado número de artigos verificado nos anos mais recentes e c) filtro de categorias de publicação da base aderentes ao tema de pesquisa. A Tabela 1 apresenta a estratégia e mecanismos de busca e a seleção das publicações.

A aplicação dos filtros de elegibilidade resultou em um volume de artigos (n = 92) cujos elementos principais como título, autores, fonte de publicação, data de publicação, número de citações, resumo, idioma, palavras-chave e área de pesquisa e outros, foram exportados para um editor de texto. Procedeu-se a leitura dos resumos e demais elementos, inclusive referências bibliográficas, a fim de estabelecer análise crítica e selecionar aqueles documentos mais aderentes ao tema de pesquisa e objetivos propostos, importando na exclusão (n = 54) e inclusão (n = 38) de trabalhos submetidos à leitura completa.

***Strings* de busca**	**Resultado Inicial**	**Artigos e *Reviews***	**Agrupamento em lista marcada**	**Filtro de tempo (2013 a 2018)**	**Filtro de Categorias da *Web of Science*****	**Artigos para leitura completa**
patent AND researcher* AND (profile OR characterist* OR factor* OR motivation*)*	354	263	263	-		
patent AND Innovation AND "Intellectual Property" AND "Triple Helix"*	10	10	271	-	**Artigos para leitura dos Resumos**	**38**
"Science-Industry Cooperation" AND (universit OR patent*)*	3	3	274	-		
profile AND patent AND (researcher* OR inventor*)*	72	54	299	174	**92**	

*** Management ; Business ; Information Science Library Science ; Economics ; Computer Science Interdisciplinary Applications ; Planning Development ; Social Sciences Interdisciplinary ; Environmental Sciences ; Environmental Studies ; Law ; Public Administration ; Area Studies ; Education Scientific Disciplines ; Political Science ; Sociology.*

Tabela 1 - Resultados da aplicação de filtros de busca
Fonte: Elaborado pelo autor.

II.2.2 Critérios de inclusão e exclusão

Os 38 artigos (n = 38) foram lidos na íntegra em todos seus elementos, inclusive quanto às referências bibliográficas, a fim de selecionar aqueles alinhados ao escopo da revisão de literatura e ao atendimento dos critérios de inclusão e exclusão da referência definitiva, quais sejam:

- Abordagem e relevância: alinhamento com o tema de estudo;

- Caráter descritivo ou explicativo do tema de estudo;
- Análise dos resultados: hipóteses claras e discutidas à luz do referencial teórico.

Da análise alguns artigos (n = 10) foram excluídos por trazerem pouco ou nenhuma contribuição para o tema de pesquisa, já que versavam predominantemente sobre padronização, qualidades de patentes, pesquisa empreendedora e relação entre pesquisadores, temas colaterais que escapam às fronteiras imediadas da presente pesquisa.

Assim sendo, após a leitura completa, foram admitidos 28 artigos (n = 28) na composição da revisão bibliográfica, a seguir organizados em ordem alfabética:

Beaudry e Kananian (2013); Blind *et al*. (2018); Boh *et al*. (2014); Bolli e Schläpfer (2013); Burhan *et al*. (2017); Chang *et al*. (2017); Ejermo e Källström (2016); Erikson *et al*. (2015); Gao *et al*. (2014); Ge *et al*. (2016); Halilem *et al*. (2017); Lawson e Sterzi (2014); Lewensohn *et al*. (2015); Liaw *et al*. (2014); Ljungberg *et al*. (2013); Mauleón *et al*. (2014); Lomachynska e Podgorna (2018); Moutinho *et al*. (2016); Natalicchio *et al*. (2017); Oliver e Sapir (2017); Schmid e Wang (2017); Singh *et al*. (2015); Stek *et al*. (2015); Stezano (2018); Svagzdiene e Kuklyte (2016); Yang e Minutolo (2016); Zemtsov *et al*. (2016); Zi e Blind (2015).

O referencial bibliográfico dos artigos selecionados foi analisado, importando na admissão de outros artigos (n = 13) que foram admitidos e adicionados à revisão. Isso se deu, principalmente pelo observado alinhamento com o tema e objeto de estudo, sendo igualmente submetidos aos critérios de inclusão e exclusão já expostos. São organizados a seguir em ordem alfabética, perfazendo o total de artigos admitidos na revisão (n=41).

Ackers (2005); Baldini *et al*. (2007); Blind e Gauch (2009); Croitoru (2012); Cruz e Souza (2015); D'este e Perkmann (2011); Garnica e Torkomian (2009); Göktepe (2008); Lam (2011); Moutinho *et al*. (2007); Oliveira (2011); Owen-Smith e Powell (2001); Sauermann e Cohen (2010).

O Quadro 2 dispõe as principais informações, metodologias, resultados e conclusões que resultaram da leitura completa dos 41 artigos que corresponderam a esse levantamento realizado.

Autor, ano	Metodologia	Principais resultados e conclusões
Baldini *et al*., 2007	Pesquisa exploratória e descritiva tipo *survey*	Pesquisa realizada com pesquisadores italianos, concluindo que a motivação para o envolvimento em atividades pantentárias se dá visando maior prestígio e reputação, além da busca por incentivos para a pesquisa.
Beaudry e Kananian, 2013	Pesquisa exploratória e descritiva com revisão bibliográfica e *survey*	Pesquisadores envolvidos em pesquisas que resultam em patentes e que mantêm relação com o financiamento privado e colaboração ciência-indústria são mais produtivos e inovadores na geração de patentes do que aqueles cuja rede de colaboração se restringe à academia.
Blind *et al*., 2018	Pesquisa exploratória e descritiva com o empreso de revisão bibliográfica e *survey*	Introduz a Teoria da Autodeterminação e os três fatores de motivação "*puzzle*," "*ribbon*," e "*gold*.". Conclui que pesquisadores têm na publicação algo inerente à sua atividade, sem relevantes motivações. Quanto à normalização é impulsionada por motivações intrínsecas, não estimuladas por recompensas ou reconhecimento. O patenteamento, por sua vez, é estimulado pelas recompensas financeiras.
Boh *et al*., 2014	Pesquisa descritiva e explicativa com estudo de caso	Verifica a amplitude de conhecimento e a profundidade da especialização do pesquisador/inventor ao impactar o sucesso e valorização profissional corporativa.
Bolli e Schläpfer, 2013	Revisão bibliográfica e estudo empírico	Um fator para a mobilidade acadêmica é a busca por infraestrutura, autonomia e reconhecimento profissional, além de acesso a financiamentos e recursos para a pesquisa.
Burhan *et al*., 2017	Pesquisa exploratória a partir de revisão de literatura e questionários	A razão/motivação para patentear vai além de proteção à propriedade intelectual, envolvendo questões de reconhecimento profissional, interesses comerciais e relações institucionais.
Chang *et al*., 2017	Pesquisa exploratória e descritiva com emprego de modelos e cenários	O estágio de divulgação de pesquisas é determinado por fatores que influencia o benefício econômico, reputação dos pesquisadores e aumento da concorrência. Maiores benefícios são demonstrados quando a divulgação e patenteamento são realizados em estágios avançados/maduros da pesquisa.
D'este e Perkmann, 2011	Pesquisa exploratória e descritiva tipo *survey*	A pesquisa analisa o envolvimento de acadêmicos britânicos com a indústria, apurando que a relação não se orienta com vistas a objetivos monetários, visando, por fim, a promoção de pesquisas em si.
Ejermo e Källström, 2016	Pesquisa exploratória e descritiva com emprego de *survey* e econometria	Importante estímulo ao pesquisador/inventor é envolvê-lo no processo e resultado da comercialização a partir do seu reconhecimento de propriedade na invenção/inovação.
Erikson *et al*., 2015	Pesquisa descritiva com emprego de questionário	Analisa o empreendedorismo da pesquisa de base universitária que evoca o resultado das atividades de pesquisa e ensino como fator de desenvolvimento econômico.
Gao *et al*., 2014	Pesquisa exploratória a partir	Enfatiza o modelo de tríplice hélice em que empresas que lideram a inovação e as universidades como

	de análise bibliométrica	estruturas geradoras de capacitação, pesquisas e conhecimento apoiadas pelos governos e resultando em inovações panteadas.
Ge *et al*., 2016	Pesquisa descritiva com levantamento econométrico	Uso de dados patentes e citação de patentes para aferir dados de filiação e mobilidade de pesquisadores na realocação de capital intelectual visando inovação e empreendedorismo deve ser usado com cautela. Para este fim a rede *Linkedin* apresenta dados promissores.
Halilem *et al*., 2017	Pesquisa exploratória e descritiva tipo *survey*	A detenção exclusiva de propriedade pela universidade e ausência de partilha nos benefícios da pesquisa/patente se traduz em menor nível de engajamento pelo pesquisador acadêmico.
Lawson e Sterzi, 2014	Pesquisa exploratória e descritiva tipo *survey*	O treinamento e envolvimento inicial em atividades de patenteamento sinaliza propensão futura no aumento quantitativo e qualitativo de patentes.
Lewensohn *et al*., 2015	Pesquisa exploratória e descritiva com análise documental e *survey*	A sobrevivência de patentes é uma medida de produtividade. A duração de uma patente depende das características da patente, do inventor e do titular da patente.
Liaw *et al*., 2014	Pesquisa exploratória e descritiva tipo *survey*	Propõe um método para classificar periódicos acadêmicos que utilizam referências de não-patentes que são documentos que diversos utilizados como base para a solicitação de patentes.
Ljungberg *et al*., 2013	Pesquisa exploratória e descritiva tipo *survey*	Destaca a diferença entre patentes acadêmicas e não-acadêmicas, estas concentradas em tecnologias de retorno imediato e contando com mais citações, enquanto aquelas focadas em pesquisas básicas e de longo prazo com menos citações.
Mauleón *et al*., 2014	Pesquisa exploratória e descritiva tipo *survey*	A falta de padronização (nome e sexo) no registro de patentes dificulta uma análise mais precisa do gênero de inventores. As mulheres têm presença no patenteamento em temas ligados a humanas e biológicas, sendo escassas nas áreas da indústria e engenharia.
Lomachynska e Podgorna, 2018	Pesquisa descritiva a partir de estudo empírico com análise comparativa	Recursos financeiros, econômicos, científicos e humanos são elementos do potencial de inovação que contribuem significativamente para a competitividade da economia.
Moutinho *et al*., 2007	Pesquisa exploratória e descritiva tipo *survey*	Considerando as organizações públicas de pesquisa em Portugal, o trabalho analisa o comportamento de pesquisadores, apurando um baixo engajamento em atividades de patenteamento e licenciamento por falta de apoio à pesquisa e baixo conhecimento dos processos.
Moutinho *et al*., 2016	Pesquisa exploratória a partir de revisão de literatura e questionários	Dentre os principais benefícios da interação ciência-indústria está a valorização do conhecimento do pesquisador e a captação de recursos que ampliem a atividade de patenteamento e transferência tecnológica.
Natalicchio *et al*., 2014	Pesquisa exploratória e	Patentes desenvolvidas a partir de colaboração de parceiros com diversificação tecnológica são mais

	descritiva tipo *survey*	impactantes.
Oliver e Sapir, 2017	Pesquisa exploratória com revisão bibliográfica	Destaca a alteração da lógica acadêmica, admitindo que o engajamento de cientistas na comercialização de pesquisas não tem impacto necessariamente negativo na transferência do conhecimento científico e na propriedade intelectual.
Owen-Smith e Powell, 2001	Pesquisa exploratória com emprego de questionário e entrevistas	O trabalho destaca a relevância do conhecimento sobre os custos e benefícios pelos pesquisadores como incentivo e promotor da divulgação de novas invenções.
Sauermann e Cohen, 2010	Pesquisa exploratória e descritiva tipo *survey*	O artigo considera a motivação dos indivíduos como relevante para o engajamento em atividades de inovação. Os fatores extrínsecos – desafio intelectual –, são mais benéficos à inovação, quando comparados aos intrínsecos – recompensas monetárias.
Schmid e Wang, 2017	Pesquisa exploratória com revisão bibliográfica e entrevistas	O artigo apresenta o paradoxo chinês de auto investimento em P&D e baixa qualidade de inovação que resulta de uma intervenção estatal que estimula uma inovação artificial e pode contribuir para o aumento de plágios e corrupção acadêmica.
Singh *et al*., 2015	Pesquisa exploratória com revisão bibliográfica	Considerando as universidades da China, Coreia, Hong Kong, Singapura e Taiwan é destacado o aumento do investimento nos respectivos sistemas nacionais de inovação. Contudo a qualidade das patentes universitárias é de baixa qualidade comparada com o ambiente privado.
Stek *et al*., 2015	Pesquisa descritiva com levantamento econométrico	A atividade de patenteamento quando resulta da interação universidade-indústria, demonstrando a importância do investimento público e sua relevância para a indústria.
Stezano, 2018	Pesquisa exploratória com revisão bibliográfica	Considera que os centros tecnológicos são importante intermediários entres os atores da inovação, principalmente alavancando expectativas e visões, fortalecendo redes de colaboração, facilitando a aprendizagem e transferência de conhecimento e tecnologia. Apoiam, assim, a implementação de políticas tecnológicas, industriais e de inovação.
Svagzdiene e kuklyte, 2016	Pesquisa exploratória com revisão bibliográfica	Dentre os fatores que identificam o estado atual e futuro da atividade inovadora estão: percentual de investimento do PIB, capital intelectual e humano, pedidos de patentes, acesso à informação e tecnologia, investimento estrangeiro.
Yang e Minutolo, 2016	Pesquisa exploratória com revisão bibliográfica	Patentes desempenham papel importante na vantagem competitiva de empresas, sendo seu valor aumentado quando agregado a patentes relacionadas.
Zemtsov *et al*., 2016	Pesquisa descritiva com análise econométrica	Benefícios do capital privado combinado com financiamento público. Destaca a mobilidade e fuga de pesquisadores para regiões desenvolvidas em busca de recursos para pesquisa.
Zi e Blind, 2015	Pesquisa exploratória e	Pesquisadores são menos propensos a participar de processos de normalização devido, principalmente, à

	descritiva tipo *survey*	proteção da propriedade intelectual. A contradição se dá porque a normalização pressupõe a divulgação da pesquisa que para ser patenteada é necessário o sigilo, fundamental às intenções da aplicação e absorção pelo mercado dos resultados da pesquisa.
Ackers, 2005	Pesquisa exploratória com revisão bibliográfica	O mercado de pesquisa é móvel, devido ao fluxo de pesquisadores qualificados que se deslocam entre instituições e regiões por motivações diversas, inclusive por reconhecimento, acesso a melhores estruturas e condições.
Blind e Gauch, 2008	Pesquisa exploratória e descritiva tipo *survey*	Apresenta padrões quanto a motivações e barreiras para a normalização aplicáveis ao processo de patenteamento.
Croitoru, 2012	Resenha	Revisão crítica do livro *Theory of Economic Development: An Inquiry Into Profits, Capital, Credit,Interest, and the Business Cycle* que considerou a inovação como elemento fundamental para o desenvolvimento econômico.
Cruz e Souza, 2015	Pesquisa exploratória com revisão bibliográfica	Enquanto o *Bayh-Dole Act* permite o patenteamento pelas universidades americanas de pesquisas com recursos públicos, a lei de inovação tecnológica brasileira representa um estágio anterior, servindo como um estímulo à inovação e pesquisa aplicada nacional.
Garnica e Torkomian, 2009	Pesquisa exploratória com estudo multicaso e entrevistas	Considerando as universidades públicas do estado de São Paulo é notada maior organização e estrutura orientada para a gestão da propriedade intelectual, sobretudo, pela atuação dos NITs e a definição de marcos e políticas de incentivo.
Göktepe-Hultén, 2008	Pesquisa exploratória com estudo de caso	A Tese oriunda da Universidade *Lund* na Suécia baseando no que denomina de terceira missão da Universidade (relação com entidades empreendendo pesquisa visando patenteamento, empreendorismo e comercialização) busca compreender os aspectos comportamentais dos inventores e fatores que influenciam o patenteamento. Destacam-se do trabalho a tipologia dos inventores e as categorias de fatores externos e internos que motivam os pesquisadores acadêmicos no Patenteamento.
Lam, 2011	Pesquisa exploratória e descritiva. Estudo de caso com emprego de entrevistas e questionário	Baseando-se na Teoria da Autodeterminação o estudo realizado no Reino Unido emprega três conceitos [*'gold' (financial rewards), 'ribbon' (reputational/career rewards)* e *'puzzle' (intrinsic satisfaction)]* para examinar aspectos extrínsecos e intrínsecos dos pesquisados envolvidos no processo de comercialização de pesquisas.
Oliveira, 2011	Pesquisa exploratória e descritiva com revisão bibliográfica e entrevista semiestruturada	Tese da Unicamp com foco nas Univ. públicas de São Paulo, o trabalho oferece evidências do avanço destas instituições na gestão da propriedade intelectual como reação ao marco legal e recursos de fomento, principalmente pela FAPESP. Quanto ao levantamento de percepções e perfil do pesquisador, se trata de um levantamento preliminar realizado a partir de entrevista com 14 pesquisadores. Dentre as principais motivações

		se destacaram a buscar por reconhecimento dos pares, promoção acadêmica e recompensas financeiras.

Quadro 2 - Principais informações, resultados e conclusões dos artigos.
Fonte: Elaborado pelo autor.

O primeiro exercício da revisão é a coleção e relação das informações, metodologias, resultados e conclusões obtidas da leitura completa dos artigos. Ainda outros artigos que compõem o arcabouço teórico e metodológico da pesquisa integram a bibliografia final sem, contudo, comporem a revisão sistemática da literatura. A proposta de sistematização seguida no levantamento e os resultados da aplicação de filtros citados e a quantidade de artigos admitidos por etapa para a composição da referência teórica estão dispostos na Figura 3.

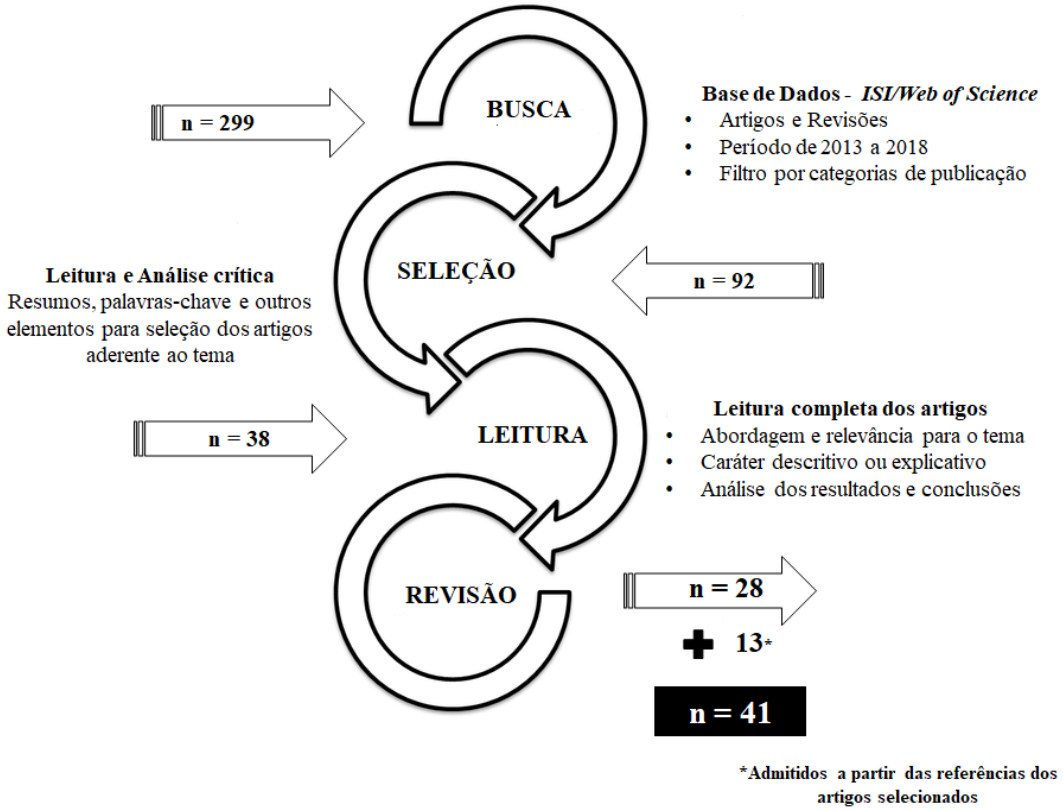


Figura 3 - Sistematização da estratégia de busca
Fonte: Adaptado de Valente (2017)

A fim de corroborar a demonstração da qualidade amostral dos artigos coletados foi utilizada a técnica de mineração de texto em conjunto com a visualização de termos, visando mapear a frequência nos trabalhos coletados e formar uma nuvem de palavras. Para esse empreendimento, foi utilizada a ferramenta *on-line* disponível em https://www.wordclouds.com/. Todos os títulos, resumos e palavras-chave dos artigos

obtidos na revisão foram organizados em um arquivo de texto em inglês cujo conteúdo foi submetido à ferramenta.

O resultado pode ser observado na Figura 4 que traz uma nuvem de palavras, servindo para informar e ilustrar o resultado da ocorrência de termos relacionados à pesquisa e presentes na bibliografia. Os 12 termos principais seguidos pelo número de ocorrências são os seguintes: *innovation* (n=76); *research* (n=68); *patente* (n=57); *patenting* (n=56); *academic* (n=48); *patents* (n=34); *university* (n=32); *researchers* (n=30); *inventors* (n=27); *factors* (n=25); *technology* (n=24); *universities* (n=23).



Figura 4 - *WordCloud* com a ocorrência de termos.
Fonte: Elaborado pelo autor a partir de wordcloud.com

II.2.3. Caracterização dos estudos

A Tabela 2 descreve as informações relevantes à pesquisa nos artigos coletados (n = 41). Observa-se que um pouco mais da metade dos artigos (n = 21) teve como campo de pesquisa a Europa. Apesar da pulverização de estudos no continente, a Alemanha (n = 3) e o Reino Unido (n = 3) têm o maior número de estudos. Neste cenário, destaque ainda para os estudos realizados no Brasil (n = 3).

Quanto aos métodos de pesquisa se destacam os estudos de caso e *Surveys*, este último se demonstrando recorrente e útil principalmente ao levantamento e evidenciação de dados quantitativos como os que resultam de estudos que mapeiam múltiplas variáveis.

Em relação ao público-alvo foi possível perceber que a terminologia empregada para nominar o grupo de pesquisa em vários estudos é, de fato, de menor relevância já que termos como inventores, pesquisadores, patenteadores, acadêmicos e cientistas são utilizados como similares na maior parte dos artigos observados. Contudo ressalta-se que em alguns estudos (n = 3) a denominação como inventores foi especialmente considerada.

Características dos estudos	n	%
Local de Realização		
Europa	21	51%
Ásia	6	15%
EUA	5	12%
Brasil	3	7%
Canadá	2	5%
México	1	2%
Rússia	1	2%
Israel	1	2%
Índia	1	2%
Tipo de estudo		
Estudo de caso	16	39%
Survey	15	37%
Revisão bibliográfica	10	24%
Público-alvo		
Pesquisadores	20	49%
Acadêmicos	10	24%
Cientistas	4	10%
Outros	4	10%
Inventores	3	7%
Ambiente de Pesquisa		
Misto	32	78%
Público	7	17%

Privado	2	5%
Aborda patenteamento		
Sim	38	93%
Não	3	7%
Apresenta base teórica (*background*) na apuração dos fatores motivacionais*		
Sim	3	7%
Não	38	95%

* Aqui se pretende observar se o trabalho apresenta teorias – por exemplo a *Self-Determination Theory* – na identificação/categorização na indicação de fatores emocionais.

Tabela 2 - Descrição dos estudos incluídos na revisão (n= 41)
Fonte: Elaborado pelo autor.

O ambiente de pesquisa misto é predominante (n = 32), considerando que os estudos coletados tomaram como universo tantas instituições de pesquisas públicas quanto as privadas, convindo ressaltar que os desenvolvidos especificamente em instituições privadas são escassos (n = 2). Um número expressivo de trabalhos (n = 38) aborda o tema afeito às patentes ou à atividade de patenteamento o que atesta a aderência da revisão ao tema proposto e, finalmente, a oportunidade de colecionar trabalhos em que as variáveis ou fatores motivacionais puderam ser observados.

Do conjunto dos trabalhos a maioria (n = 38) apresenta e desenvolve estudos abordando fatores afeitos ao comportamento, engajamento, capacitação, disponibilidade de insumos e condições infraestuturais em relação a pesquisadores, sem necessariamente apresentar uma base teórica quanto à origem do fator abordado. Não quer isto representar que tal fator não seja de caráter motivacional, aliás, quase todos os fatores considerados têm potencial de relação, por exemplo, com a *Self-Determination Theory (Teoria da Autodeterminação),* teoria presente na definição e tratamento dos fatores motivacionais nos trabalhos de Lam (2011) e Blind *et al*. (2018). Como exemplo de tal potencial se tem o estudo sueco desenvolvido por Göktepe (2008) que, apesar de não mencionar a referida teoria, faz distinção de fatores de base motivacional então classificados em internos e externos, abordagem similar à assumida na referida teoria que, por sua vez, classifica os fatores em intrínsecos e extrínsecos. A citada teoria, apesar de não servir à definição e/ou conceituação de fatores motivacionais, é principalmente importante no fornecimento de base para a

classificação e tratamentos dos mesmos, o que foi abordado na estruturação do *framework* proposto.

II.2.4. Identificação dos fatores motivacionais

Após o primeiro levantamento dos artigos conforme descrito na seção anterior, o passo seguinte se volta para identificação dos fatores que exercem influência sobre a motivação dos pesquisadores em patentear os resultados de suas pesquisas. Nesse sentido, duas observações se fazem necessárias.

A primeira refere-se ao fato de que a produção do conhecimento é dinâmica e novos estudos vão sendo desenvolvidos. Assim sendo foi feito um monitoramento dos artigos divulgados a partir de 2018, que pudessem ser incorporados na análise, fazendo parte da fundamentação teórica do presente trabalho.

A segunda observação é quanto ao fato de que o primeiro levantamento não se restringiu exclusivamente ao perfil e aos fatores motivacionais dos pesquisadores das universidades quanto ao patenteamento. Apesar da relevância desses trabalhos correlatos e para uma melhor compreensão conceitual sobre a temática desse estudo, a identificação dos fatores motivacionais considerou uma parte das publicações desse primeiro levantamento, acrescida de outras decorrentes de trabalhos que foram incorporados, seja devido ao surgimento de novos artigos/relatórios, seja por terem sido identificadas ao longo do desenvolvimento do trabalho.

A revisão da literatura complementar foi baseada em artigos e revisões indexados nas bases de dados da *Web of Science* e *Scopus*, escolhidas devido à sua gama de indexação, relevância e característica multidisciplinar das publicações. A seguinte consulta foi adotada considerando os campos de título, resumo e palavras-chave: [*patent * AND (researcher* OR inventor**) *AND motivation** *AND universit**]. Na Web of Science, foram encontrados 30 documentos, sendo 23 artigos ou revisões. No *Scopus*, foram encontrados 41 documentos, sendo 29 artigos ou revisões. No entanto, muitos artigos apresentaram conteúdo não relevante para os propósitos deste estudo. Excluindo publicações e duplicações não alinhadas, 7 artigos permaneceram com foco específico nas motivações ou perfil de pesquisadores universitários.

Novas pesquisas foram realizadas utilizando outras consultas mais abrangentes para obter um maior número de publicações: [patent* AND (researcher* OR inventor*) AND

(profile OR characterist* OR factor* OR motivation*) AND universit*]; and [patent* AND (researcher* OR inventor*) AND (motivation*)]. Outras publicações foram adicionadas a partir de pesquisas em outras bases de dados e nas referências de artigos selecionados. Com esses procedimentos, o número de publicações aumentou para 56.

Em uma segunda etapa foi feita a leitura completa das publicações selecionadas de acordo com os seguintes critérios: (i) abordagem e relevância: alinhamento com o tema do estudo; (ii) caráter descritivo ou explicativo do sujeito de estudo; e (iii) análise dos resultados: hipóteses claras discutidas à luz do *framework* teórico. Algumas publicações foram excluídas com base nessa análise mais detalhada de seu conteúdo, restando 25 publicações efetivamente utilizadas na elaboração do *framework*.

As publicações selecionadas foram divididas em dois grupos para a elaboração do *framework*. O primeiro grupo foi formado por publicações que abordavam modelos e apresentavam perfis de pesquisadores. Esse primeiro grupo de publicações teve maior influência na estruturação do *framework*, fornecendo sua base conceitual. O segundo grupo foi importante para ampliar e complementar a identificação de fatores motivacionais capazes de influenciar o comportamento dos pesquisadores.

Diante do exposto, tem-se que os fatores motivacionais foram identificados a partir de um conjunto de 24 artigos considerados mais aderentes em relação ao propósito da pesquisa, conforme relacionado no Quadro 3. Após a leitura dos estudos e suprimidas eventuais repetições de conteúdo e abordagens, os fatores motivacionais dos autores apresentados representam, pela maior relevância e abrangência, a coleção de fatores e motivos alinhados ao escopo do presente trabalho. Acrescenta-se que, para cada uma das publicações, foi observado se o estudo teve como base alguma teoria específica sobre motivação.

Autores	**Local**	**Fatores motivacionais**	**Base Teórica**
Ackers, 2005	Europa	Progressão na carreira; Busca por centros de excelência de pesquisa; Diferenciais de salário; Melhores instalações de pesquisa; Viagens e experiência cultural	Não
Baldini, 2011	Itália	Prestígio / reputação e troca de conhecimento; diversos incentivos, cuja importância varia de acordo com as características pessoais e o contexto.	Não
Baldini *et al*. 2007	Itália	Fundos para pesquisa; Equipamentos para laboratórios de pesquisa; Fundos de investigação; Fundos para bolsistas de pesquisa; Ganhos pessoais; Prestígio / visibilidade / reputação; Estímulos para a pesquisa; Troca de conhecimento com o mundo exterior; Desenvolvimento econômico e tecnológico	Não

Blind e Gauch 2009	Alemanha	Solução de problemas técnicos; Vantagem tecnológica; Reputação pessoal; Fontes de financiamento; comercialização	Não
Blind *et al*. 2018	Alemanha	Estabelecimento de Spin-offs; Cooperação com a indústria; Garantia da tecnologia; Publicação de resultados; Reputação do Pesquisador; Licenciamento	Sim
Bolli e Schläpfer 2013	Áustria, Alemanha e Suécia	Estabilidade profissional; Baixa mobilidade	Sim
Burhan *et al.*2017	Índia	Proteção da pesquisa; Fama e reconhecimento; Licenciamento e negociação; Sinalização para investidores e colaboradores; Pressão institucional e dos pares; Incentivo por recompensa financeira/ indicador de produção	Não
Chang *et al*. 2017	China	Benefício econômico; Reputação; Competição e colaboração	Não
Closs *et al*., 2013	Brasil	Auto-direção; estimulação; hedonismo; conquistas; poder; conformidade; tradição; benevolência; segurança; universalismo.	Não
D'este e Perkmann 2010	Reino Unido	Aplicabilidade da pesquisa; Renda da pesquisa; Informações sobre o problema; Feedback da indústria; pesquisa industrial; acesso a materiais e equipamentos; Rede de colaboração; conhecimento da pesquisa,	Não
Daniel e Alves, 2020	Portugal	Valorização curricular (prestígio / visibilidade / reputação), potencial de comercialização e/ou incentivos da universidade.	Não
Freitas e Nuvolari, 2012	Holanda	Proteção das invenções que serão comercializadas; motivos heterodoxos principalmente para sinalizar competências de pesquisa específicas; projetos alinhados às agendas de pesquisa dos patrocinadores públicos.	Não
Göktepe, 2008	Suécia	Solucionar questões de pesquisa; Recompensas pessoais e sociais; Benefícios ou recompensas financeiros; Empregabilidade e plano de carreira; Características pessoais; Ciclo de vida da carreira científica; Capital científico (imagem e reputação); Capital social (rede de colaboração); Disciplina científica e Relevância industrial; Recursos de pesquisa e acesso a fundos e recursos externos; Influência dos pares, modelos e regras de grupo; Localização, ambiente social e cultura; Estratégia e política da academia/universidade; Legislação de patentes; Núcleos de Inovação Tecnológica (NIT); Interesse político (Terceira Missão)	Não
Halilem *et al*. 2017	Canadá	Motivação Pecuniária	Não
Huang *et al*., 2019	Taiwan	Comercialização, aprendizagem, acesso a recursos e ensino.	Não
Hussler e Penin, 2012	France	Motivação proativa, em que os cientistas estão entusiasmados com as patentes; motivação reativa, em que os cientistas são mais relutantes em patentear, mesmo que sejam forçados a fazê-lo de qualquer maneira.	Não
Lach e Schankerman, 2008	EUA	Incentivos e royalties	Não
Lam 2011	Reino Unido	Financiamento para pesquisas; Aplicação e exploração de resultados de pesquisa; Troca / transferência de conhecimento; Curiosidade intelectual; Construir redes	Sim

		pessoais e profissionais; Geração de renda e oportunidades para estudantes; Renda pessoal	
Moutinho *et al*. 2007	Portugal	Proteção do conhecimento; Organização e objetivos da pesquisa; Financiamento e retornos da pesquisa; Relações e sinalizações de mercado para a indústria; Retornos sociais e financeiros	Não
Oliveira, 2011	Brasil	Elevar reputação, prestígio e reconhecimento; Benefícios financeiros; Gosto por desafio; Multidisciplinariedade; interesse pela aplicação do resultado	Não
Owen-Smith e Powell 2001	EUA	Proteção da pesquisa; Alavancagem da carreira; Fonte de renda; Solução de problemas; Treinamento	Não
Sauermann e Cohen 2010	EUA	Salário ou renda; Segurança no emprego; Desafio intelectual; Independência e autonomia; Oportunidade de avanço na carreira; Nível de responsabilidade; Contribuição para a sociedade	Não
Schmid e Wang 2017	China	Maior autonomia na pesquisa; Incentivo financeiro.	Não
Walter *et al*. (2018)	Alemanha	Incentivos financeiros diretos e indiretos	Não

Quadro 3 - Fatores motivacionais identificados na literatura selecionada
Fonte: Elaborado pelo autor.

A partir dos dados organizados no quadro 3 é possível perceber a predominância de estudos do continente europeu. Combinados com dados da tabela 2, é possível observar que a Teoria da Autodeterminação foi empregada nos estudos (n = 3) desenvolvidos por Lam (2011), Bolli e Schläpfer (2013) e Blind *et al*. (2018).

Ainda do quadro 3 se observa que o volume de variáveis observadas se avoluma em razão, principalmente, dos objetivos específicos dos estudos coletados e das diferenças terminológicas. Por exemplo, Göktepe (2008) prefere considerar o termo “benefícios financeiros” enquanto Lam (2011) emprega “renda pessoal” ou ainda Chang *et al*. (2017) adota o termo “benefício econômico” para designar a variável ou fator financeiro. O fator financeiro e/ou econômico seguido pelos fatores de pessoais e os ligados à carreira e vida acadêmica são abordados na maioria dos estudos observados, apontando e endossando, nesta leitura, a classificação trazida por Lam (2011), dentre outros, quanto às classes *Gold, Puzzle e Ribbon*, tratadas mais adiante.

II.2.5 Identificação dos perfis motivacionais

A leitura das publicações selecionadas também permitiu identificar alguns trabalhos voltados para definição de perfis motivacionais.

Dentre os estudos com foco em conhecer e entender o perfil e motivação de pesquisadores/inventores, considerando a profundidade e abrangência da abordagem, além da consistente metodologia, estão os desenvolvidos por Göktepe (2008); Lam

(2011); Blind *et al*. (2018), que foram as principias referências utilizadas como fundamentação conceitual para construção do *framework*. Outros trabalhos também podem ser citados tais como Blind e Gauch (2009); Bolli e Schläpfer (2013); Ge *et al*. (2016); Burhan *et al*. (2017); e Chang *et al*. (2017). Segue breve descrição desses trabalhos:

a) O estudo de Goktepe (2008)

O trabalho de Göktepe (2008), intitulado *Inside the Ivory Tower:InventorsePatents at Lund University*, resultou em uma tese apresentada na *Lund University*. Considera o ambiente universitário e o sistema patentário sueco e busca investigar por que e como os pesquisadores patenteiam, identificando os fatores internos e externos. Trata-se de uma pesquisa exploratória e estudo de caso por meio de técnicas qualitativas e quantitativas para a coleta de dados, como entrevistas e *survey* a que foram submetidos mais de 120 pesquisadores.

A partir de uma revisão bibliográfica, e sem se basear em uma teoria específica, o trabalho busca classificar os fatores motivacionais e categorizar os pesquisadores. A correlação das variáveis deste estudo com os demais colhidos na revisão considera os fatores internos como aqueles relacionados às características como habilidades, competências e experiências do pesquisador. Os externos, por sua vez, estão relacionados à estrutura institucional, legal e acesso a recursos materiais e emocionais. O trabalho tem relevante lista de fatores motivacionais que, apesar de não citar base teórica para tal arranjo, poderia ser associada às considerações trazidas na Teoria da Autodeterminação (TDA).

Göktepe (2008) apresenta ainda relevante categorização ou tipologia dos pesquisadores, adaptada na Figura 5, quanto à frequência do patenteamento. Distinguem os inventores seriais, aqueles com três ou mais patentes, dos ocasionais, com menos de três patentes. Quanto às parcerias, os inventores são distinguidos como ativos e passivos.

Quanto às parcerias, os inventores são distinguidos como ativos e passivos. Tem-se, por fim, que os **seriais ativos** têm 3 patentes ou mais e estão envolvidos em todos os aspectos do pedido de patente e na comercialização, desenvolvendo suas patentes através de *spin-offs*, mantendo assim sua filiação à universidade e estando cientes quanto ao potencial valor comercial da pesquisa na indústria. Os **seriais passivos**,

também com 3 patentes ou mais, e não participam diretamente do processo de solicitação da patente e compartilham os resultados com uma entidade/empresa parceira que solicita a patente e compensa o pesquisador pela transferência de propriedade intelectual.

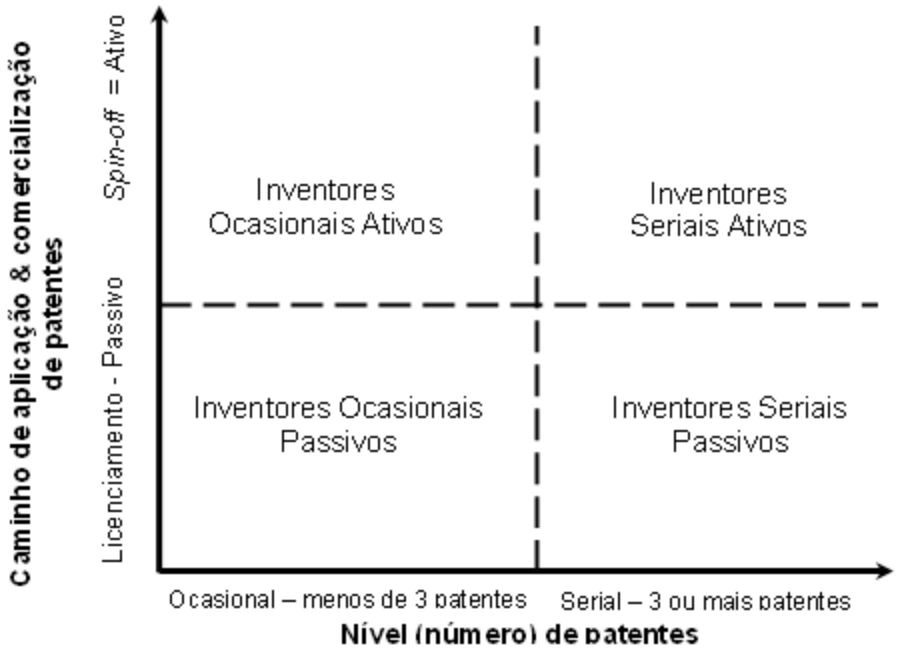


Figura 5 - Tipologia de inventores
Fonte: Adaptado de Göktepe (2008)

Os **ocasionais ativos** são pesquisadores com menos de três patentes, mas se utilizaram de *spin-offs*. Os **ocasionais passivos** também têm menos de três patentes e patenteiam em parcerias com colegas experientes, mormente inventores em seriais, se valendo de escritórios de transferência de tecnologia para registro de suas invenções.

Salienta-se que, considerando o escopo e metodologia do presente trabalho, são considerados passivos os pesquisadores que têm menos do que cinco patentes, enquanto ativos são os que têm cinco patentes ou mais.

b) O estudo de Lam (2011)

No artigo *What motivates academic scientists to engage in research commercialization: 'Gold', 'ribbon' or 'puzzle'?*, Lam (2011), com inspirações do trabalho de Göktepe (2008), desenvolve uma pesquisa exploratória e descritiva, empregando o estudo de caso para examinar aspectos das motivações de pesquisadores tendo como escopo a

comercialização de pesquisas. O estudo se baseou em 36 entrevistas individuais e um questionário *on-line* com 735 cientistas de cinco grandes universidades do Reino Unido. Citando a Teoria da Autodeterminação o estudo categoriza os fatores em internos e externos e também introduz outra teoria a partir das referências de Stephan e Everhart (1998) para qualificar as motivações em três categorias de motivadores ou fatores de engajamento: *Ribbon, Gold e Puzzle.*

O emprego da Teoria da Autodeterminação por Lam (2011) se dá a partir de pesquisas desenvolvidas por Deci e Ryan (2000); Ryan e Deci, (2000a; 2000b) e Gagné e Deci (2005), considerando que o comportamento e respostas dos indivíduos são influenciados por variáveis pessoais e outras afeitas ao desenvolvimento da atividade. Conclui que as motivações dos indivíduos diferem, sendo determinadas por fatores externos e internos, dando origem a uma ação. A Teoria da Autodeterminação postula que o comportamento humano se dá pelo atendimento de necessidades psicológicas básicas: a *competência*, *autonomia* e *vínculo*, conjugando *o porquê* com o *para quê* das atividades. O aprimoramento da teoria pelos autores tem demonstrado que da relação das motivações extrínsecas e intrínsecas depende a qualidade da experiência e o desempenho nas atividades desenvolvidas pelos indivíduos (Deci e Ryan, 2006).

Adicionalmente o trabalho apresenta categoria de motivações pecuniárias ou não, citando Stephan e Everhart (1998), divididas em *Ribbon, Gold e Puzzle* que, numa tradução livre em português resultaria nas expressões Fita, Ouro e Desafio. De forma conjugada, o fator *ribbon* daria conta de motivações de origem mista, envolvendo aspectos extrínsecos e intrínsecos já que conjuga efeitos no prestígio e reputação do inventor reconhecido por si e pelos pares em função da conquista na publicação do trabalho ou no registro da patente ou citações de patentes ou, ainda, no autoreconhecimento em desenvolver a pesquisa e observar os resultados. No presente trabalho a tradução *reputação* para *ribbon* é a adotada. Extrínseca é o fator *gold* que representa a busca por recompensa financeira ou benefício econômico-financeiro. Este fator decorre da pesquisa em si, mas não necessariamente importa em incremento pecuniário para o inventor, mas em financiamento da infraestrutura e insumos da pesquisa. O fator intrínseco, por sua vez, está no *puzzle* que representa a satisfação em resolver problemas e prover soluções criativas para questões e desafios inerentes à pesquisa.

A Figura 6 apresenta o modelo de engajamento na atividade comercial conceituado por Lam (2011) cujo escopo está na comercialização dos resultados de pesquisa, mas que bem serve aos estudos de atividades patentárias.

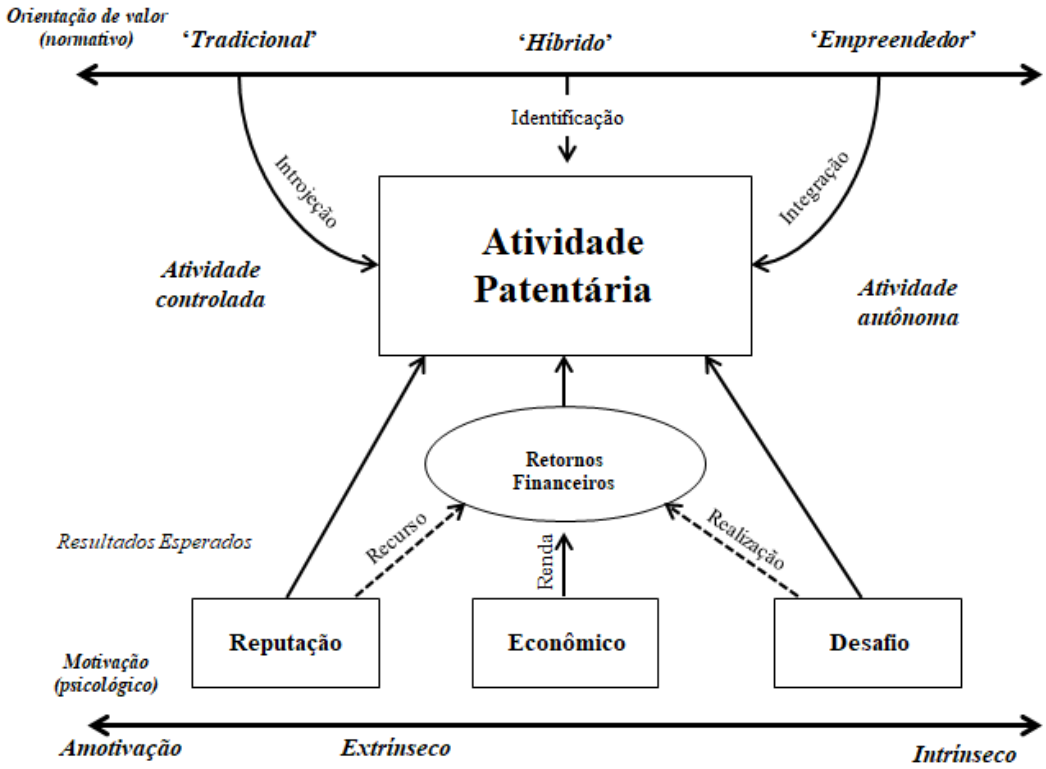


Figura 6 - Engajamento da atividade patentária
Fonte: Adaptado de Lam (2011)

c) O estudo de Blind *et al*. (2018) e Blind e Gauch (2009)

Blind *et al. (*2018) protagonizaram o estudo mais recente e que observa as motivações e barreiras de engajamento ao processo de publicação, patenteamento e normalização de pesquisas. No desenvolvimento de uma pesquisa exploratória e descritiva foram considerados 129 pesquisadores de um instituto federal de pesquisa na Alemanha. À semelhança de Lam (2011), o autor considera a Teoria da Autodeterminação e os três fatores (*puzzle, ribbon e gold*) na identificação das motivações. Uma importante contribuição do trabalho, além dos resultados mais atualizados que permitem melhor comparação, está uma vasta revisão dos estudos da Teoria da Autodeterminação. Quanto às patentes, os fatores considerados no estudo são mais ligados às dificuldades encontradas no processo de patentear do que necessariamente limitações ou interesses de ordem pessoal. Contudo, a relação de fatores analisados serve ao

mapeamento de impactos de processos na criação ou ampliação de estímulos pessoais ao pesquisador.

Por sua vez, o trabalho de Blind e Gauch (2009) é focado em conhecer os motivos, obstáculos e soluções no processo de normalização nas áreas de nanotecnologia e nanociência. Na análise das motivações está sua principal contribuição na medida em que os autores buscam entender as motivações pessoais de pesquisadores em participar da padronização das pesquisas. Não se baseia em teoria específica, mas destaca que as principais motivações se localizam entre encontrar a solução para um problema técnico, aumentar a vantagem tecnológica e acompanhar o desempenho de outros pesquisadores. Dentre as motivações de menor impacto está a redução do custo de transação e o aprimoramento da reputação social.

d) Outros estudos

Bolli e Schläpfer (2013) analisam o fenômeno da mobilidade considerando os efeitos do ambiente em economistas na Alemanha, Áustria e Suíça. Partindo do princípio de que o universo da pesquisa é altamente qualificado, móvel e de amplitude internacional, o trabalho identifica dentre os fatores da mobilidade a reputação, autonomia, acesso a fontes de pesquisa e ascensão na carreira. Apesar de não se deter na análise do perfil dos pesquisados e nem citar uma teoria para tal, o trabalho considera que as motivações extrínsecas desempenham um papel importante na publicação pelos pesquisadores da área, apresentando fatores que podem dar conta da mensuração de interesses e percepções dos pesquisadores.

Nesta mesma linha, o trabalho desenvolvido por Ge *et al*. (2016) investiga a mobilidade de inventores, considerando tal fenômeno como uma realocação de capital intelectual visando à inovação e empreendedorismo. Na busca pela motivação para tal, apesar de não citar base teórica para as motivações, se propõe a estabelecer comparação entre os dados de registros de patentes com os fornecidos na rede social *Linkedin*, para concluir que os dados coletados desta são mais confiáveis do que aqueles disponíveis nos registros.

Considerando o aumento da atividade de patenteamento Burhan *et al*. (2017) analisam o interesse de pesquisadores na Índia em patentear, apresentando um conjunto de motivações para o depósito de patentes organizado a partir de uma revisão na literatura do tema. O trabalho, também não citando teoria como base, faz distinção

entre pesquisadores juniores e seniores a fim de estabelecer comparação dos efeitos de fatores como: proteção, licenciamento, recompensas, reputação, colaboração e pressão organizacional. Os resultados demonstraram efeitos diferenciados. Nos seniores predominou a busca por licenciamento da pesquisa, colaboração e proteção da pesquisa, enquanto nos juniores predomina a colaboração, seguida pelas recompensas, fama e pressão organizacional.

Chang *et al*. (2017) consideram a influência de fatores como benefício econômico, valor para a reputação, diferencial com a concorrência e o estabelecimento de colaboração na divulgação de invenções. Também sem apresentar base teórica que sirva à mensuração das motivações, verifica que os benefícios, que também podem ser considerados fatores motivadores, são maiores quando a pesquisa é divulgada e a invenção patenteada em estágios mais avançados, provendo melhores resultados na transferência de tecnologia entre a indústria e universidade.

Considerando a análise sucinta dos trabalhos revisados e, em especial, os que nesta seção foram citados, uma pequena amostra empregou teorias para identificar e aferir motivações. Na maioria dos estudos os fatores na intenção de conhecer e mensurar comportamentos, tendências e motivos de pesquisadores ou outros profissionais são realizados a partir de fatores extraídos de revisões bibliográficas em literatura disponível. Nos poucos trabalhos que abordam aspectos da motivação a Teoria da Autodeterminação se apresenta como relevante em oferecer os parâmetros para classificação das motivações e categorização dos pesquisadores.

II.3 Construção do *framework*

A partir das abordagens, conceitos, resultados e metodologias obtidas na revisão sistemática da literatura, é desenvolvida a seguinte proposta de um modelo de abstração teórica, denominado *framework,* para identificação do perfil motivacional dos pesquisadores em relação às atividades de patenteamento, conforme etapas descritas a seguir. A figura 7 apresenta o esquema do desenvolvimento do *framework*.

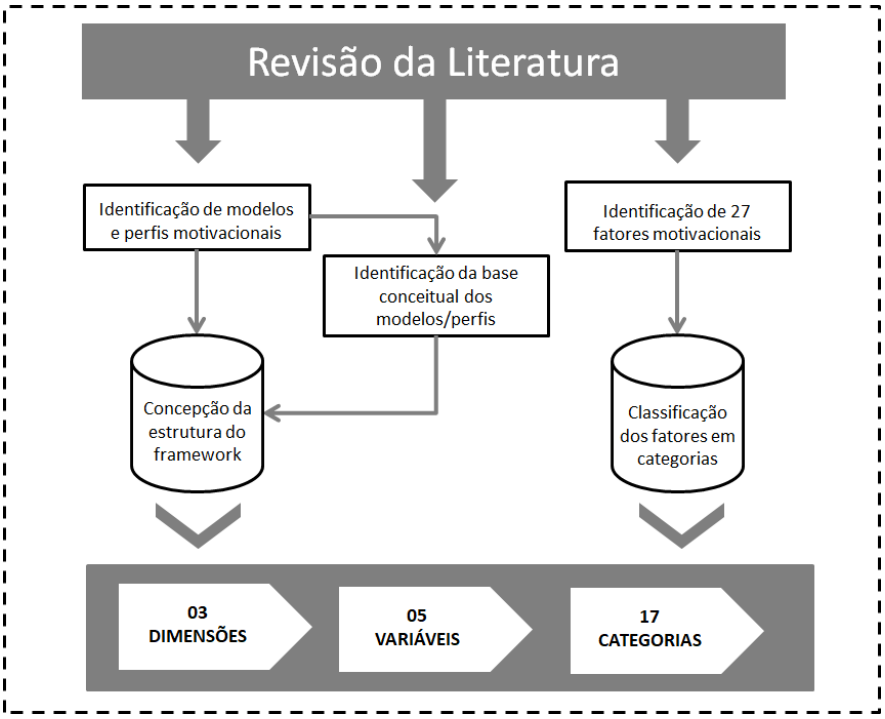


Figura 7 - Esquema de desenvolvimento do *framework*
Fonte: Elaborado pelo autor.

II.3.1. Definição de dimensões, variáveis e categorias

Conforme mencionado anteriormente, o *framework* foi estruturado com base nas informações coletadas e considerações contidas na literatura revisada, com especial destaque para os trabalhos desenvolvidos por Göktepe (2008); Lam (2011) e Blind *et al*. (2018), para, a partir do recurso da abstração, organizar, relacionar e representar as principais teorias e conceitos colecionados em um modelo analítico conceitual.

Os trabalhos de Lam (2011) e Blind *et al. (*2018) se destacam ao consolidar um modelo que congrega teorias e permite a aferição e classificação dos aspectos motivacionais. O trabalho de Göktepe (2008), por sua vez, se apresenta ao fornecimento de parâmetros para a categorização dos pesquisadores e fornecimento de expressivo número de fatores internos e externos que servem à padronização daqueles obtidos dispersos em outros trabalhos, como se verificou no quadro 2.

Ao fim, os três estudos mencionados, à luz dos demais trabalhos considerados na revisão, dão conta de um modelo que pode servir à mensuração dos fatores motivacionais ligados a pesquisadores que desenvolvem atividades patentárias, enquanto os demais estudos oferecem diferentes e complementares aspectos da

motivação que servem como indicadores e fatores – dados de entrada para o emprego do modelo analítico desenvolvido.

Considerando a base conceitual adotada e as diversas classificações e interações dos estudos analisados, foram definidas cinco variáveis: (1) natureza de motivação; (2) resultados esperados; (3) orientação profissional; (4) frequência e parceria; e (5) relação entre padrões/valores pessoais. Essas variáveis foram agregadas em três dimensões: pessoal; comportamental; e ambiental. A dimensão pessoal compreende as variáveis 1 e 2. A dimensão comportamental é formada pelas variáveis 3 e 4. Por fim, a dimensão ambiental está associada à variável 5, refletindo assim as condições que podem favorecer ou dificultar a motivação para patentear pelos pesquisadores. No total, foram consideradas 17 categorias, distribuídas entre as cinco variáveis do *framework*.

Esse modelo foi estruturado em três dimensões (pessoal; comportamental; ambiental) abrangendo cinco variáveis (natureza motivacional; resultados esperados; orientação profissional; frequência e parcerias; relação com normas e valores), cada qual com suas respectivas categorizações, conforme mostrado no Quadro 4.

Dimensões	Variáveis	Fontes Bibliográficas	Categorias
Pessoal	Natureza da motivação	Stephan e Everhart (1998); Deci e Ryan (2000); Ryan e Deci, (2000a; 2000b); Gagne e Deci (2005); Goktepe (2008); Lam (2011); Bolli e Schlapfer (2013); Blind *et al.*(2018)	Extrínseco, Intrínseco, Misto.
	Resultados esperados	Stephan e Everhart (1998); Goktepe (2008); Blind e Gauch (2009); Lam (2011); blind *et al.*(2018)	*Econômico, Desafio, Reputação.*
Comportamental	Orientação profissional	Blind e Gauch (2009); Lam (2011)	Tradicional Puro, Tradicional pragmático, híbrido, empreendedor.
	Frequência e parceria	Goktepe (2008)	Ocasional Ativo, Serial Ativo, Ocasional Passivo e Serial Passivo.

Ambiental	Relação entre padrões /valores pessoais	Lam (2011)	Introjeção, Identificação, Integração.

Quadro 4 - Dimensões estruturantes, variáveis e categorias do *framework*
Fonte: Elaborados pelos autores

Por sua vez, o quadro 5 descreve brevemente cada categoria, trazendo descrições sucintas oriundas da referência bibliográfica já citada.

Variáveis	**Categorias**	**Descrição sucinta**
Natureza da motivação	Extrínseca	A motivação deriva do ato de fazer, como uma espécie de recompensa ou resultado pela realização da tarefa.
	Intrínseca	A motivação se manifesta no ato de fazer e a satisfação na sua realização.
	Mista	A motivação agrega componentes extrínseco e intrínseco no fazer e o que vem da conclusão da tarefa.
Resultados esperados	Econômico	Busca por recompensa financeira ou benefício econômico-financeiro, incluindo insumos para si mesmo e/ou para o projeto que integra.
	Reputação	Busca por prestígio e notoriedade, incluindo autoreconhecimento e, principalmente, por parte dos pares.
	Desafio	Superar e resolver problemas, fornecendo soluções criativas para questões e desafios intelectuais propostos.
Orientação Profissional	Puro Tradicional	Faz total distinção entre universidade e indústria, atuando e avançando profissionalmente apenas no campo acadêmico.
	Pragmático tradicional	Distingue a universidade da indústria, desenvolvendo atividades relacionadas à indústria principalmente para adquirir recursos para apoiar a pesquisa acadêmica.
	Híbrido	Valoriza a colaboração acadêmia-indústria e busca desenvolver atividades ligadas à indústria com vistas ao avanço científico.
	Empreendedor	Acredita na colaboração da academia-indústria e busca desenvolver atividades ligadas ao setor privado com vistas ao uso comercial.
Frequência e Parceria	Ocasional Ativo	Têm menos de cinco pedidos de patentes cujo desenvolvimento resulta de ação entre parceiros de atividades, interação com a indústria e até *spin-offs*.
	Serial Ativo	Têm pelo menos cinco pedidos de patentes cujo desenvolvimento resulta de ação entre pares e *spin-offs*. Empreendedor acadêmico filiado à instituição acadêmica que atua visando à comercialização e com ampla relação com a indústria.
	Ocasional Passivo	Têm menos de cinco pedidos de patentes que resultaram de parcerias com pares mais experientes, especialmente inventores em série. Vale-se do apoio de escritórios de transferência de tecnologia para registrar suas invenções.

	Serial Passivo	Têm cinco ou mais pedidos de patentes desenvolvidos predominantemente em ambientes acadêmicos e com apoio de pares. São considerados acadêmicos que apreciam os valores do empreendedorismo que podem se potencializar pela aproximação com o meio privado.
Relação entre padrões /valores pessoais	Introjeção	Comportamento impulsionado pelo cumprimento das regras, uma vez que não há congruência espontânea dessas com valores pessoais internos, resultando em uma atividade parcialmente controlada e motivada externamente a partir das normas.
	Identificação	Há a intenção do indivíduo de agir adequadamente, harmonizando valores pessoais e os esperados pelo meio, o que significa uma atividade com maior autonomia e, portanto, mais intrinsecamente motivada, mas com a presença de ingredientes extrínsecos.
	Integração	Há maior identificação pessoal e assimilação de regulamentos e valores sociais, possibilitando identificar a motivação intrínseca do indivíduo sem a necessidade de regulação externa.

Quadro 5 - Descrição das categorias do *framework*
Fonte: Elaborado pelo autor.

A Figura 8 enfatiza a dinâmica e interação das dimensões e variáveis que compõem o pano de fundo na identificação do perfil comportamental/ motivacional.

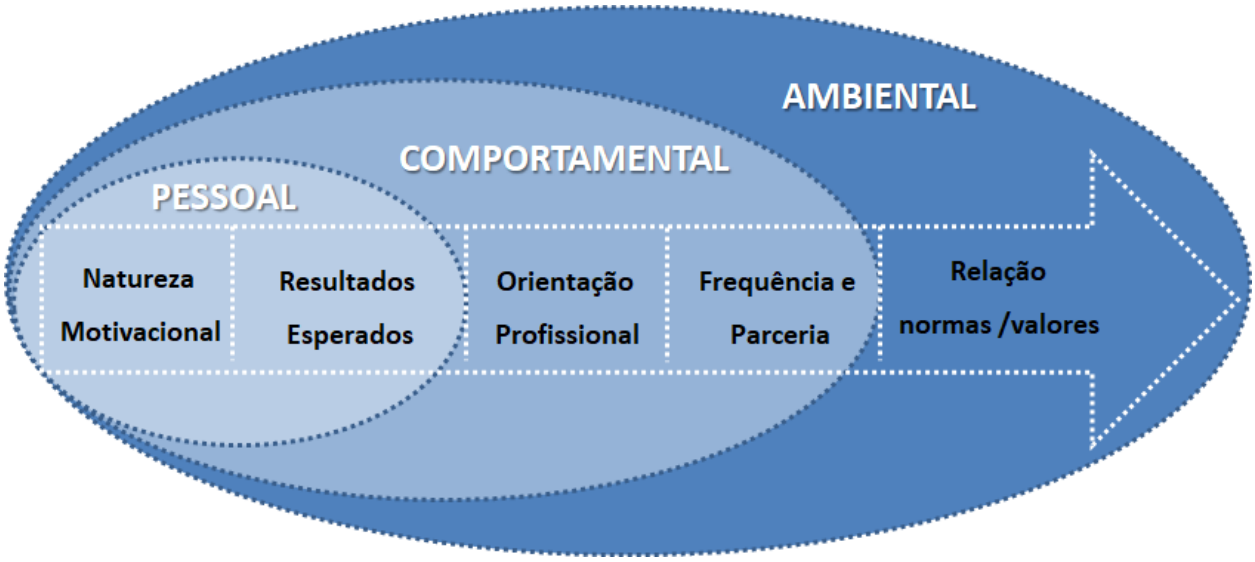


Figura 8 - Ênfase nas dimensões na aferição do perfil do pesquisador
Fonte: Elaborados pelos autores

As variáveis "natureza da motivação" e “resultados esperados" têm relação direta e, segundo Blind *et al. (*2018), combinada. Tais variáveis foram associadas à dimensão pessoal do pesquisador uma vez que são elas que estão na origem da motivação que direcionam os pesquisadores para um determinado comportamento. A pretensão

econômica do pesquisador pode ter origem material, importando em aumento de renda e outros fatores financeiros relacionados, configurando, assim, a natureza extrínseca como apontada pela Teoria da Autodeterminação, já que se orienta a partir do que decorre da atividade desenvolvida. Quanto ao resultado afetivo que se dá a partir do reconhecimento, principalmente dos pares e também pelo próprio inventor a partir do resultado obtido, se tem uma recompensa de natureza mista, considerando os aspectos intrínsecos e extrínsecos que se apresentam no decorrer da atividade. O resultado que se orienta pela solução de desafios tem um destacado caráter social, envolvendo aspectos verificados principalmente no desenvolvimento da atividade, em uma relação de ‘ciência pela ciência’ e ‘ciência pela sociedade’.

A dimensão comportamental, por sua vez, abrange as variáveis "orientação profissional" e "frequência e parcerias" uma vez que refletem a forma de atuação dos pesquisadores.

Quanto à orientação profissional, segundo Lam (2011), se tem pelo menos quatro perfis possíveis, aferindo a relação entre o ambiente acadêmico e o financiamento e absorção pela indústria quantos aos resultados da pesquisa. O *tradicional puro* faz plena distinção entre universidade, atuando e avançando profissionalmente apenas no campo acadêmico. O *tradicional pragmático* distingue a universidade da indústria, mas busca desenvolver atividades ligadas à indústria principalmente para adquirir recursos para apoiar a pesquisa acadêmica. O *híbrido* valoriza a colaboração entre academia-indústria e busca desenvolver atividades vinculadas a indústria visando ao avanço científico. O *empreendedor* acredita na colaboração da academia-indústria e busca desenvolver atividades vinculadas à indústria visando ao aproveitamento comercial.

Por sua vez, a tipologia proposta por Göktepe (2008), já apresentada graficamente na figura 3, classifica quanto à frequência de patenteamento e orientação para o mercado em perfil ocasional ativo ou passivo e serial ativo ou passivo.

No que se refere à dimensão ambiental, e estruturação do *framework* se baseou no trabalho de Lam (2011), ao analisar a relação do pesquisador, admitindo os motivos orientadores dos comportamentos que variam a partir da congruência entre os valores dos indivíduos e as normas ligadas à atividade. Assim, o comportamento se dá, pelo menos, de três formas: *introjeção*, quando o comportamento é movido pela necessidade de se encaixar com o meio e sentir-se digno na medida em que há

aceitação parcial de regulações externas, necessárias já que não há congruência com valores pessoais internos, importando em uma atividade parcialmente controlada e motivada externamente a partir das regulações; a *identificação* se dá pela intenção do indivíduo em agir de maneira apropriada, havendo a congruência entre as metas pessoais e os valores esperados pelo meio, importando em uma atividade com maior autonomia e, assim, motivada mais intrinsecamente; e, por fim, na *integração* há identificação e assimilação pessoal completa com os regulamentos e valores sociais, sendo possível a identificação da motivação intrínseca do indivíduo sem a necessidade de regulação externa.

III.3.2. Relação entre fatores e variáveis

Uma vez definidas as dimensões, variáveis e respectivas categorizações, buscou-se identificar os fatores motivacionais e sua relação com as variáveis: natureza da motivação e os resultados esperados.

Foram identificados 27 fatores motivacionais, dentre aqueles listados no Quadro 3, que emergiram da revisão sistemática. Buscou-se então classificar esses fatores em relação à dimensão pessoal, considerando as categorias relacionadas às variáveis "natureza da motivação" e "resultados esperados".

O Quadro 6 mostra a representação dos diferentes fatores coletados e que servem à redação de sentenças que condensam os fatores motivacionais, com as respectivas referências da literatura, relacionando-os com os resultados esperados e a natureza da motivação.

Fatores motivadores de patenteamento	**Fontes Bibliográficas**	**Natureza da motivação**	**Resultados esperados**
Acesso a meios e insumos, incluindo tecnológicos, para o desenvolvimento da pesquisa	Blind e Gauch (2009)	Extrínseco	*Gold*
Apoio administrativo institucional no processo de patenteamento, especialmente dos Escritórios de Transferência de Tecnologia (TTOs) da Universidade	Goektep (2008)	Misto	*Ribbon*
Cumprimento de indicadores de produção e possibilidades de publicação de pesquisas cujos resultados são patenteáveis	Burhan *et al.*(2017)	Misto	*Ribbon*

Aumento da renda pessoal e/ou benefícios econômicos resultantes da atividade de pesquisa	Ackers (2005); Goekte (2008); Chang *et al.*(2017); Halilem *et al.*(2017); Lam (2011); Moutinho *et al.*(2007); Oliveira (2011); Owen-smith e Powell (2001); Sauerman e Cohen (2010); Ouellette e Tutt (2020); Daniel e Alves (2020).	Extrínseco	*Gold*
Maior prestígio e visibilidade acadêmica com reconhecimento e reputação por pares	Goektep (2008); Blind e Gauch (2009); Blind *et al.*(2018); Burhan *et al.*(2017); Oliveira (2011); Daniel e Alves (2020).	Misto	*Ribbon*
Baixa mobilidade: apego e permanência a um local fixo e específico para o desenvolvimento da pesquisa.	Bolli e Schlapfer (2015)	Misto	*Ribbon*
Posição profissional e nível de responsabilidade que estimule o desenvolvimento da pesquisa	Sauerman e Cohen (2010); Daniel e Alves (2020).	Extrínseco	*Gold*
Concorrência e colaboração entre pares no desenvolvimento de redes pessoais e profissionais	Chang *et al.*(2017); Lam (2011); D'este e Perkmann (2011); Daniel e Alves (2020).	Misto	*Ribbon*
Contribuição de promoção e/ou desenvolvimento de carreira com impacto na remuneração	Owen-smith e Powell (2001); Ackers (2005); Goektep (2008); Baldini *et al.*(2007); Sauerman e Cohen (2010); Ouellette eTutt (2020)	Extrínseco	*Gold*
Contribuição dos resultados da pesquisa para o desenvolvimento social e econômico da sociedade	Goektep (2008); Moutinho *et al.*(2007); Sauerman e Cohen (2010)	Intrínseco	*Puzzle*
Disponibilidade de materiais, dados e recursos da indústria para o desenvolvimento de atividades de pesquisa	Baldini *et al.*(2007); Goektep (2008); D'este e Perkmann (2011)	Misto	*Ribbon*
Experiência profissional prévia como acadêmico e pesquisador	Goektep (2008)	Intrínseco	*Puzzle*
Atração por desafios, resolução de problemas de pesquisa e curiosidade intelectual	Owen-smith e Powell (2001); Goektep (2008); Blind e Gauch (2009) Sauerman e Cohen (2010); Oliveira (2011); Lam (2011); Ouellette e Tutt (2020)	Intrínseco	*Puzzle*
Capacidade de financiar e conduzir negócios	Goektep (2008)	Extrínseco	*Gold*
Independência e autonomia no desenvolvimento da pesquisa	Sauerman e Cohen (2010); Schmid e Wang (2017)	Intrínseco	*Puzzle*
Influência de colaboradores e parceiros estrangeiros	Goektep (2008)	Misto	*Ribbon*
Influência dos membros da pesquisa a partir da troca de experiências	Baldini *et al.*(2007); Goektep (2008); Lam (2011)	Intrínseco	*Puzzle*

Instalações e condições de infraestrutura para o desenvolvimento de pesquisas disponíveis na instituição	Ackers (2005); D'este e Perkmann (2011); Ouellette e Tutt (2020)	Intrínseco	*Puzzle*
Licenciamento e Comercialização de resultados de pesquisa	Blind e Gauch (2009); Blind *et al.*(2018); Burhan *et al.*(2017) D'este e Perkmann (2011); Ouellette e Tutt (2020); Daniel e Alves (2020).	Extrínseco	*Gold*
Momento profissional da carreira acadêmica ou pesquisadora	Bolli e Schlapfer (2015); Sauerman e Cohen (2010); Daniel e Alves (2020).	Extrínseco	*Gold*
Obtenção de recursos e recursos para pesquisa	Blind e Gauch (2009); Lam (2011); Moutinho *et al.*(2007); Ouellette e Tutt (2020)	Extrínseco	*Gold*
Influência de viagens e/ou experiências em outros lugares/culturas para o desenvolvimento de pesquisas	Ackers (2005)	Misto	*Ribbon*
Potencial para a aplicação de resultados de pesquisa	Moutinho *et al.*(2007); D'este e Perkmann (2011); Lam (2011); Oliveira (2011); Daniel e Alves (2020).	Intrínseco	*Puzzle*
Proteção e apropriação intelectual de resultados de pesquisa	Owen-smith e Powell (2001); Moutinho *et al.*(2007); Burhan *et al.*(2017); Goektep (2008); blind *et al.*(2018)	Intrínseco	*Puzzle*
Qualificação e competência pessoal para o desenvolvimento da pesquisa	Owen-smith e Powell (2001); Daniel e Alves (2020).	Extrínseco	*Gold*
Relações de mercado e cooperação com a indústria, incluindo a atração de recursos financeiros	(Goektep , 2008); Blind *et al.*(2018); Burhan *et al.*(2017); Schmid e Wang (2017); Moutinho *et al.*(2007)	Misto	*Ribbon*
Satisfação pessoal e realização profissional na realização da pesquisa	Goektep (2008)	Intrínseco	*Puzzle*

Quadro 6 - Indicadores para aferição do perfil do pesquisador
Fonte: Elaborados pelos autores

Ressalta-se que a decisão por atribuir e relacionar as referidas sentenças às variáveis é resultado de decisão metodológica e reflexão conceitual a partir do exame dos trabalhos de referências, o que não exclui que outras relações, mais ou menos abrangentes ou mesmo inexistentes, sejam possíveis ou admitidas.

O objetivo de estabelecer tais sentenças - relacionadas a cada um dos fatores motivacionais - é fornecer referências e dados de entrada para mensurar o impacto de cada fator observado a partir do *framework* a ser proposto e, ainda, servir de base na construção de um questionário *on-line* a que os pesquisadores são submetidos e cujo modelo está disponível no Apêndice A. De fato, a integração das referidas sentenças

com os indicadores e suas variantes dispostos no quadro 4 e 5 representa um cenário mais abrangente e qualificado para a definição do perfil do pesquisador. A Figura 9, por sua vez, é o *framework* em si que proporciona a adaptação do marco conceitual já apresentado e conjuga ambas as teorias motivacionais (Autodeterminação e *Ribbon, Gold, Puzzle).*

O modelo analítico representado na Figura 9 condensa um arcabouço intelectual a fim de subsidiar e fornecer parâmetros para a aplicação de um instrumento de apuração (questionário *on-line*) do perfil do pesquisador, individual ou em grupo, engajado na atividade patentária e integrando os indicadores relacionados nos quadros 4 e 5.

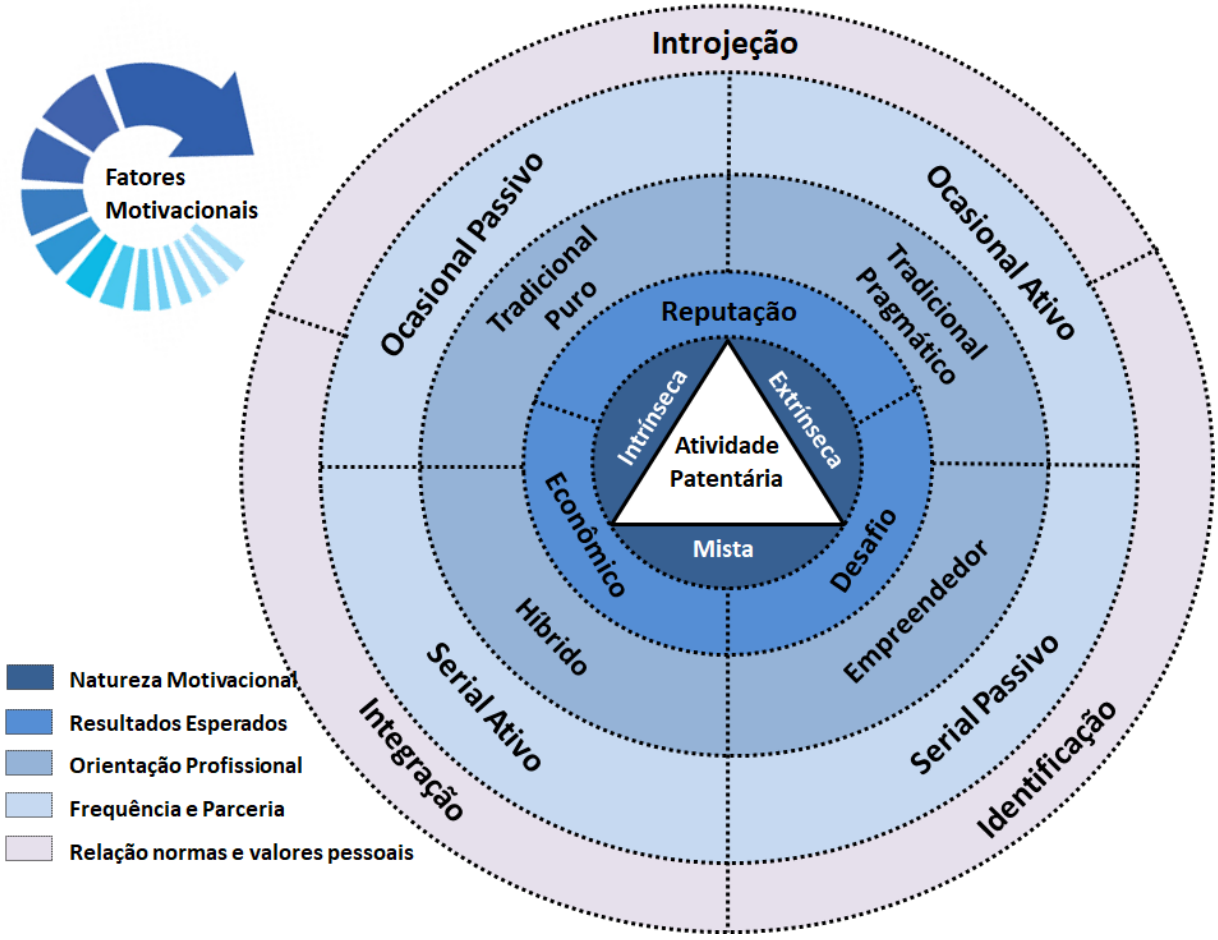


Figura 9 - *Framework* de aferição das motivações na atividade patentária.
Fonte: Elaborados pelos autores

O emprego do *framework* é instrumentalizado pelos diferentes indicadores dispostos em sentenças e conceitos oriundos do agrupamento de diferentes fatores motivacionais coletados na revisão. Trata-se de um sistema dinâmico, cujo movimento das camadas

se dá a partir da assunção das variáveis, acolhendo, portanto, uma rotação que admite a combinação de diferentes variáveis em estágios distintos, mas complementares, a fim de aferir o perfil motivacional do indivíduo ou um grupo engajado na atividade patentária. Por conjugar variáveis objetivas e subjetivas, diferentes conceitos e abordagens, o modelo revela seu potencial inovador no mapeamento do comportamento do pesquisador.

O modelo de abstração conceitual representado no *framework* aglutina os principais parâmetros e referências que servem à elaboração da metodologia de tratamento de dados e emprego de algoritmo de classificação que lhe confere automatização cujo desenvolvimento é demonstrado no capítulo V

No nível individual, cada pesquisador é classificado em uma categoria considerando cada variável que compõe o *framework*. Por exemplo, um pesquisador pode ser caracterizado como "extrínseco, econômico, empreendedor, serial-ativo e integração", enquanto outro pesquisador pode ser classificado como "intrínseco, desafio, híbrido, ocasionalmente ativo e identificação". É esse conjunto de cinco categorias, cada uma associada a uma variável, que define o perfil do pesquisador.

No Capítulo V os conceitos e parâmetros aqui explicitados são retomados a fim de orientar o processamento de dados. Serve, assim, no geral, à etapa computacional como arcabouço teórico, e, no específico cada camada e elemento do modelo como variável de saída.

Partindo do centro da representação, a primeira camada, natureza motivacional, segundo a teoria da autodeterminação, tem três naturezas ou categorias possíveis: intrínseca, extrínseca e mista.

Na segunda camada, resultados esperados, são três as categorias: desafio (*puzzle*), econômico (*gold*), reputação (*ribbon*).

As duas camadas têm relação direta, importando que a categorização de uma corresponda à outra, na relação:

Extrínseca – Econômico (*gold*)
Intrínseca – Desafio (*puzzle*)
Mista – Reputação (*ribbon*)

As 27 questões do questionário *on-line* seguem a relação apresentada, sendo dispostas por ordem alfabética em 3 grupos de 9 questões que apuram tanto a natureza das motivações quanto os resultados esperados. Esta relação foi apresentada no quadro 06.

Foi apurada a atribuição de importância em cinco níveis:

1 Nem um pouco importante
2 Pouco importante
3 Moderadamente importante
4 Muito importante
5 Extremamente importante

A terceira camada destaca a orientação profissional em quatro categorias possíveis, aferindo a relação entre o ambiente acadêmico e o financiamento e a absorção pela indústria quantos aos resultados da pesquisa – tradicional puro, tradicional pragmático, híbrido e empreendedor.

Na quarta camada estão evidenciadas a frequência e parceria no processo de pesquisar/patentear. Considerando que a metodologia admitiu na amostra pesquisadores com no mínimo cinco patentes, a amostra se divide em seriais ativos ou passivos e ocasionais ativos ou passivos.

A quinta e última camada verifica a relação entre as normas do meio e os valores pessoais do pesquisador. Baseia-se nos pressupostos do esquema trazido por Lam (2011) expostos na figura 6. Como sua definição é mais complexa seu marcador considera informações das camadas anteriores.

Marcadores para o perfil de introjeção: natureza motivacional: mista, resultado esperado: reputação. Frequência e parceria: Serial passivo. Orientação profissional: tradicional ou tradicional pragmático (a depender da resposta à pergunta própria).

Marcadores para o perfil de identificação: natureza motivacional: intrínseca, resultado esperado: desafio. Frequência e parceria: Serial ativo. Orientação profissional: empreendedor.

Marcadores para o perfil de identificação: natureza motivacional: intrínseca, resultado esperado: desafio. Frequência e parceria: Serial ativo. Orientação profissional: empreendedor.

É importante mencionar que essa combinação de categorias e variáveis permite identificar uma grande variedade de perfis de pesquisadores. São múltiplas possibilidades de perfis, que refletem a complexidade da natureza humana e das especificidades locais.

Embora o número de perfis diferentes seja grande, é possível identificar e agrupar pesquisadores com características, interesses e motivações semelhantes. Como o perfil de cada pesquisador é dado por um conjunto de cinco categorias, a semelhança ou diferença entre os pesquisadores pode variar de zero (quando dois pesquisadores apresentam o mesmo perfil) a cinco categorias de classificação (quando dois pesquisadores figuram em todas as categorias distintamente).

Outro aspecto a ser destacado é que a estrutura também pode ser usada para representar visualmente o perfil de um pesquisador, um grupo de pesquisadores ou uma instituição.

Como todas as categorias estão representadas na imagem do *framework*, é possível representar os perfis através de cores diferentes. No caso de uma universidade, por exemplo, diferentes cores podem ser usadas para representar as categorias que compõem o perfil de determinados departamentos ou áreas de conhecimento ou mesmo de um grupo de pesquisadores ligados à instituição. Também é possível utilizar uma escala de cor para ilustrar a frequência de cada categoria do *framework*, permitindo uma visualização rápida e fácil do perfil geral dos pesquisadores daquela instituição. As figuras 10a e 10b apresentam duas versões gráficas para a exposição de uma classificação, dentre outras possíveis, como exemplos de representação visual da classificação aplicáveis à apresentação de perfil individual e de grupo de pesquisadores de uma eventual universidade.

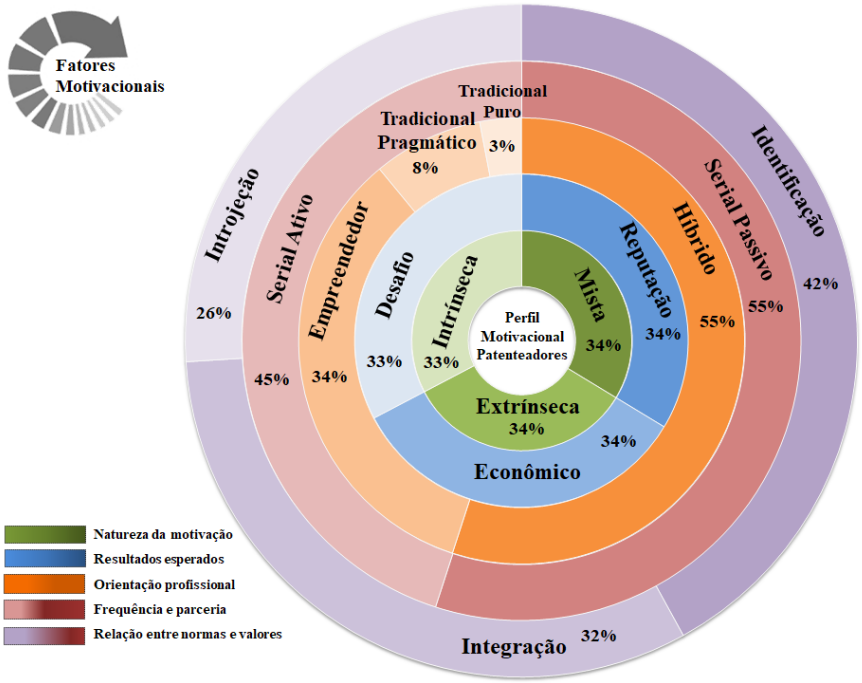


Figura 10a - *Framework* no perfil individual

Fonte: Elaborados pelos autores

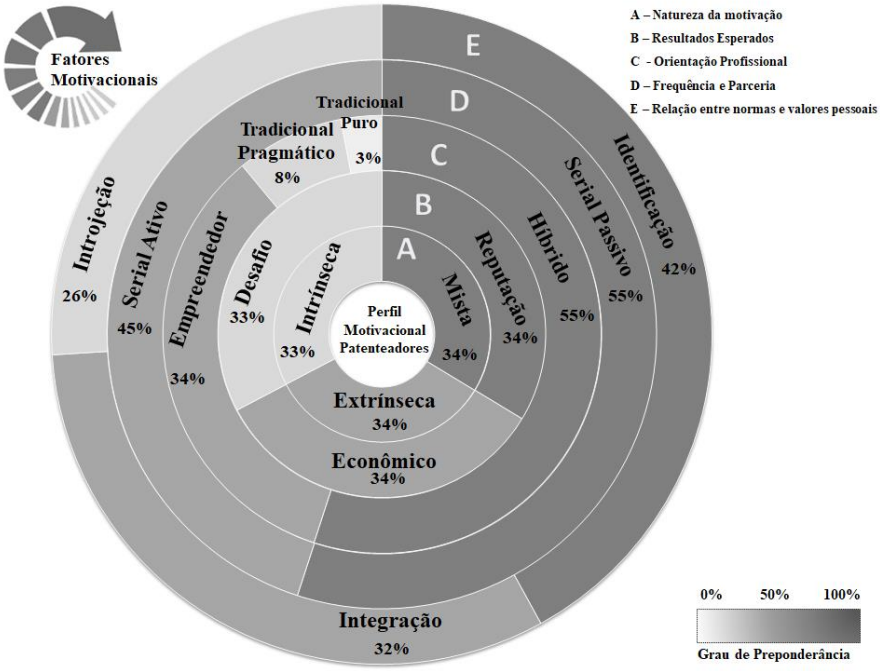


Figura 10b - *Framework* no perfil de grupo

Fonte: Elaborados pelos autores

Como se pode observar, tais modelos são capazes de mostrar o perfil dos pesquisadores ou grupos de pesquisadores de uma instituição - o que é algo complexo - de forma simples, direta e objetiva. Esse é um aspecto importante, pois essa representação visual pode ajudar a facilitar a análise e a tomada de decisões pelos gestores institucionais.

O *Framework*, é importante ressaltar, não tem o objetivo de atribuir uma classificação final e definitiva aos titulares de patentes. Os estudos consultados, como Ryan (2014), Ryan e Berbegal-Miraben (2016) e Tijssen (2018), por exemplo, sugerem que a identificação de perfis é uma questão complexa e orientada por diferentes valores e contextos que se refletem em comportamentos. Por esse motivo, adota-se uma escala de preponderância para representar o resultado, indicando mais uma tendência e não necessariamente um diagnóstico conclusivo.

Com base na pesquisa e no relato de um conjunto de fatores motivacionais que se traduzem em variáveis de medição, a presente pesquisa é pioneira no desenvolvimento de uma estrutura que delineia e representa as principais motivações, ao mesmo tempo em que sintetiza e fornece um modelo para identificar perfis.

Conforme aplicado por Göktepe (2008) e Lam (2011), as representações gráficas servem para ilustrar a capacidade do modelo de representar o perfil individual ou de um grupo de pesquisadores. O modelo aplica e ilustra suas dimensões e elementos estruturantes de forma clara e perceptível, permitindo tanto a compreensão teórica quanto a finalidade pretendida na criação de perfis. Esse é um aspecto importante, pois a interface visual é o resultado de um processo, cujo objetivo é fornecer dados e parâmetros para apoiar a análise e a tomada de decisão dos gestores institucionais.

II.4. Considerações finais do capítulo

O desenvolvimento da revisão sistemática se revelou eficiente em colecionar estudos com os principais e mais recentes dados científicos, permitindo a sistematização dos temas, a comparação de métodos, obtenção de contribuições e a observação de eventuais lacunas.

As contribuições teóricas da presente são evidenciadas de diferentes formas. Em primeiro lugar, há uma contribuição para o avanço do conhecimento científico na

medida em que a revisão da literatura contribui para o conhecimento desta temática complexa e fragmentada.

Em segundo lugar, embora algumas pesquisas tenham abordado temas relacionados à inovação, motivação e empreendedorismo acadêmico, este trabalho enriquece a literatura acadêmica ao reunir e sintetizar teorias, conceitos, modelos e metodologias, dando centralidade ao tema da medição dos perfis motivacionais de pesquisadores ligados às atividades de pesquisa e patenteamento.

Os achados do presente estudo sugerem que a identificação de perfis não é uma questão trivial, pois valoriza diferentes comportamentos, valores e, sobretudo, motivações. Nesse sentido, o trabalho é pioneiro no mapeamento e relato de um conjunto de 27 fatores motivacionais e três grupos de elementos que possibilitaram o desenvolvimento de uma estrutura que modula visualmente e representa ao mesmo tempo em que sintetiza e disponibiliza um modelo estruturado para avaliação de perfil.

Visando a revisão de artigos científicos é estabelecida a estratégia de busca e a aplicação dos critérios de inclusão e exclusão de artigos, foi apresentado o levantamento de 41 trabalhos científicos (n=41). Esse levantamento foi posteriormente complementado e atualizado resultando no detalhamento de publicações em cujo escopo se observou o tratamento da questão da motivação relacionada a pesquisadores, sendo possível colecionar fatores considerados motivacionais.

O tema da cooperação ciência-indústria dá conta da interação CT&I a fim de gerar e transmitir conhecimento científico e tecnológico. Remonta ao cenário de interação mútua dos elementos da tríplice hélice (universidade, indústria e governo) que, na esteira do conhecimento empreendedor, evoca principalmente as universidades, mas também centros de pesquisas públicos, a exercerem a denominada terceira missão – o incentivo à inovação em estruturas públicas, apoiadas por investimentos privados que, por fim, promovem o engajamento de inventores e reconhecem a propriedade da invenção desenvolvida em ambientes acadêmicos financiados por recursos públicos. No modelo de inovação não há ênfase em protagonismos e sim em uma visão integrada e sistêmica que promove a inovação, garantindo a propriedade intelectual e incentivando a atividade de patenteamento.

A atividade patentária se demonstrou um objeto de recente e também crescente análise, importando em relevante indicador das atividades de transferência tecnológica

e de efetiva proteção à propriedade intelectual. O protagonismo das patentes universitárias revela o potencial de moderação e incentivo das entidades de pesquisas públicas e escritórios de transferência tecnológica em desenvolver nos pesquisadores uma participação efetiva no sistema patentário.

As percepções, condições e motivações para engajamento na atividade de patenteamento revela a incipiência e também o potencial de investigação do tema. Dos artigos revisados foi possível colecionar fatores motivacionais que servem como indicadores e parâmetros para a investigação do perfil motivacional dos pesquisadores. A maioria dos estudos, apesar de oferecerem fatores motivacionais, não se basearam em uma teoria específica para este fim, se valendo de revisão da literatura para tal. Alguns trabalhos fundamentaram seu desenvolvimento na Teoria da Autodeterminação, sobretudo quanto à classificação de pesquisadores e natureza das motivações.

Para além da apuração de fatores motivacionais o estudo permitiu coletar indicadores, variáveis e sentenças que contribuiu na elaboração de questionário de consulta e aferição do perfil motivacional. Proporcionou ainda variáveis e categorias como os resultados esperados, frequência e estabelecimento de parceria no processo de patenteamento, natureza da motivação, orientação profissional e a relação dos valores pessoais do pesquisador com as normas vigentes no meio. De tal levantamento foi desenvolvido e apresentado um *framework* que a partir de sentenças preestabelecidas remontou a um arcabouço conceitual, oferecendo um modelo analítico e dinâmico que afere as motivações e comportamento de pesquisadores na atividade patentária.

O *framework* proposto considera três dimensões (pessoal, comportamental e ambiental) em cujo escopo cinco conjuntos de variáveis foram progressivamente organizados em 17 categorias, oferecendo diversas combinações possíveis de perfis compostos a partir da conjugação e interação dos diferentes fatores motivacionais.

Finalmente, o *framework* proposto consolida uma contribuição efetiva para este trabalho na medida em que se aplica e representa visualmente de forma clara e facilmente perceptível seus elementos estruturantes, permitindo tanto sua compreensão teórica quanto sua constituição na pronta aferição de perfis.

O modelo estrutural proposto tem sua principal aplicação na identificação do perfil motivacional dos pesquisadores, atendendo principalmente gestores e tomadores de decisão na geração de políticas de promoção e patenteamento de pesquisas. A partir

da identificação da forma como um determinado pesquisador ou mesmo um grupo de pesquisadores, setores ou instituições, são motivados, o gestor pode desenvolver incentivos para o desenvolvimento da pesquisa de forma mais personalizada e estabelecer políticas e ações para promover a inovação, o empreendedorismo acadêmico e o patenteamento de forma mais assertiva e de acordo com cada perfil identificado.

Além da produção de políticas mais assertivas, a identificação de perfis é uma grande aliada na prevenção e mediação de conflitos de papel na vida acadêmica, bem como na promoção de uma alocação melhor e mais satisfatória de recursos humanos a partir de suas motivações.

Embora o *framework* tenha como principal foco pesquisadores afiliados às universidades é possível adaptá-lo para aferir perfis de outros grupos e organizações envolvidas em outras etapas do processo de inovação e atividades de pesquisa.

Nesse sentido, o *framework* tem várias aplicações. Serve para diagnosticar as motivações que impulsionam os pesquisadores vinculados a instituições de pesquisa em seu interesse em patentear. Uma vez conhecidos os fatores de menor ou maior relevância atribuídos por um grupo de pesquisadores, é possível desenvolver políticas de apoio institucional mais eficazes, bem como o estabelecimento de políticas públicas para aprimorar os processos de inovação e patenteamento, inclusive na relação com a indústria. Também se aplica ao desenvolvimento de incentivos, com base em tendências estabelecidas, visando à atração e à retenção de recursos humanos. Essas e outras possibilidades têm o potencial de estabelecer perfis por áreas de pesquisa, grupos de interesse, além de comparações entre entidades públicas e privadas, regiões de interesse econômico e acadêmico e, por fim, a identificação de perfil ou conjunto de perfis, individuais e institucionais, comparando medições em cenários nacionais e internacionais.

CAPÍTULO III – METODOLOGIA PARA CONSTRUÇÃO DO *DATASET*

O presente capítulo busca identificar o universo de análise e definir as amostras de pesquisa que culminaram na construção de um *dataset*. Esse *dataset* foi utilizado para traçar o perfil sociodemográfico dos patenteadores das universidades brasileiras e como base para a estruturação do modelo de classificação do perfil motivacional desses pesquisadores.

Inicialmente é apresentada, de forma sucinta, conceituações relacionadas às técnicas computacionais da fase de coleta de dados como o *web Scraping* e linguagem *Python* empregadas na extração, organização e anonimização dos dados.

Na sequência, é feita a descrição de como foi realizada a coleta de dados em diversas etapas: (a) identificação dos pesquisadores, a partir de um levantamento realizado na base de patentes da *Derwent Innovation Index*, (b) definição do tamanho da amostra do estudo; (c) levantamento de informações sobre os pesquisadores na base de currículos *Lattes*; e (d) aplicação de questionários.

III.1 Conceituações básicas

Dados são fundamentais à pesquisa científica como insumo elementar ao desenvolvimento de estudos nos variados campos acadêmicos, além de servir como fator de qualidade, aplicabilidade e rigor metodológico. Por sua vez, a fonte e o método de coleta de dados são aliados do que Gil (2011) denomina como conhecimento científico cujas principais características são a verificabilidade e a confiabilidade.

Um desafio para todo pesquisador é a fase de coleta dos dados, crucial etapa de pesquisa, considerando, como salientam Castro e Ferrari (2016), o volume, variedade e velocidade da massa de dados crescentemente disponível. Como salientam Marconi e Lakatos (2010), o aproveitamento e adequação dos dados coletados variam conforme a habilidade, experiência e capacidade do pesquisador.

As técnicas para coleta de dados variam em função da natureza, abordagem e da estratégia de pesquisa, importando em um delineamento da investigação que culmina no tratamento dos dados e caracterização dos resultados. As múltiplas técnicas e a apreciação multidisciplinar conferem validade e colaboram para o rigor dos meios e fins da pesquisa (SEVERINO, 2007; MARCONI e LAKATOS, 2010; GIL, 2011).

Para vários campos de investigação a *Internet* é o ambiente que concentra uma gama de dados e informações, permitindo acesso instantâneo, em diferentes formatos, interfaces e endereços, mas tão pulverizados quanto disponíveis ao pesquisador que se proponha a buscá-los. A extração de informação, a indexação ou o processamento semântico dos dados não são tarefas triviais e, se realizadas manualmente, sem o devido equilíbrio entre rigor e flexibilidade de coleta e tratamento, estão sujeitas a incorreções e incompletudes.

A fim de acessar e coletar o vasto conteúdo deste universo digital a técnica denominada *Web Scraping* emerge como uma solução baseada no desenvolvimento de um algoritmo na linguagem de programação *Python,* instrumentalizando e automatizando a extração e estruturação de dados (SAURKAR *et al*., 2018; DIOUF *et al*., 2019).

A conjugação de diferentes naturezas, abordagens e técnicas no desenvolvimento da pesquisa acadêmica visa atender ao que Yin (2010) denominou como organização de pesquisa multidisciplinar, conjugando habilidades, experiências e competências distintas, porém complementares como recomendam Marconi e Lakatos (2010).

Na presente pesquisa a competência multidisciplinar se faz necessária, principalmente aquelas afeitas às áreas da engenharia de produção, ciência da informação e mineração de dados, oportunizando a concertação de conhecimentos e técnicas para desenvolver uma solução, principalmente com ferramentas computacionais, para o acesso e coleta de informações em bancos de dados disponíveis na *web*.

III.1.1 A técnica *Web Scraping*

Informações de diferentes tipos e origens podem ser obtidas através da *internet*, algumas de cunho financeiro, social, segurança, acadêmico, dentre outras. As informações estão disponíveis na rede mundial em diferentes formatos e através de interfaces distintas de acesso. Diferentes *sites* que pertencem a categorias específicas exibem informações nos mais variados formatos, mesmo em um único endereço, elevando à impossibilidade a tarefa de acessar e coletar todos os dados de uma só vez, pois podem estar distribuídos em várias páginas díspares - tarefa complexa e nada trivial.

Web Scraping é a denominação de uma técnica que busca solucionar tal problema. É a extração ou raspagem de dados da *web*, uma técnica usada para transformar dados não estruturados em dados estruturados (DIOUF *et al*., 2019), facilitando sua visualização e análise posterior (SAURKAR *et al*., 2018). Existem diversas formas de *Web Scraping* desde o elementar copiar e colar, baseadas em correspondência com expressões regulares, programação, análise de código *HTML* e computação visual (SIRISURIYA, 2015). Há ainda diversos *softwares* para *Web Scraping* disponíveis, em boa parte baseados nas linguagens *Java, Python e Ruby*. Dentre os *softwares* disponíveis há opções gratuitas e comerciais, em código aberto ou fechado.

A técnica tem como objetivo extrair informações de *sites* e estruturá-las, tornando-as mais compreensíveis e, portanto, facilitando o acesso e análise, como em planilhas, banco de dados, arquivos .CSV. A Figura 11 ilustra o processo de extração de dados e a arquitetura básica da técnica.



Figura 11 - Arquitetura básica de *Web Scraping*
Fonte: Elaborados pelos autores

Cumpre destacar que a aplicação da referida técnica na presente pesquisa busca desenvolver uma investigação com uma equipe multidisciplinar cujo legado principal seja permitir que pesquisas desenvolvidas na área de estudo de engenharia de produção, por exemplo, possam se valer de ferramentas das áreas da ciência da informação a fim de serem replicadas. No escopo do estudo ora desenvolvido, por exemplo, isso se dá a partir do desenvolvimento de um algoritmo na linguagem *Python* a fim de aplicar a citada técnica na extração de dados disponíveis em bancos de patentes e pesquisadores.

Existem diferentes formas de *Web Scraping* (SAURKAR, A. V. *et al*., 2018). Neste trabalho, o *Web Scraping* desenvolvido é baseado na técnica *Hyper Text Markup Language (HTML) Parsin*g. Tal técnica utiliza-se de linguagens de consulta de dados semiestruturados, para analisar páginas *HTML* com o objetivo de recuperar e transformar o conteúdo das páginas da *web* analisadas.

III.1.2 A linguagem *Python*

Para múltiplas aplicações em variados campos do conhecimento acadêmico e profissional é requisito o conhecimento de uma linguagem de programação. A tarefa de coleta e análise de dados pode ser árdua, onerosa e mais sujeita a imprecisões na obtenção de dados de forma estruturada e segura sem os recursos de programação como a linguagem *Python,* gerida pela *Python Software Foundation (PSF)*, bastante difundida.

Para o objetivo a que o presente trabalho se propõe, a referida linguagem é aplicada considerando sua simplicidade e robustez, enquanto oferece bom desempenho e versatilidade na integração com outras linguagens. Além de ser empregada no acesso automatizado a bancos de dados acadêmicos, como realizado por Rose e Kitchin (2019) e Jesus *et al.* (2020) ao, respectivamente, desenvolverem um recurso aplicado em consulta na base *Scopus* e em anais de congressos, proporcionando menor taxa de erros na coleta de dados e permitindo maior reprodutibilidade metodológica (GENTZKOW e SHAPIRO, 2014).

A linguagem *Python* é reconhecida pelo seu formato intuitivo, sensível e estruturado, suportando modularidade, já que cada arquivo gerado é considerado um módulo, sendo possível a quebra do código em partes menores e, portanto, mais gerenciáveis (Da SILVA e SILVA, 2019).

O algoritmo desenvolvido em linguagem *Python* requer algumas de suas bibliotecas básicas, como *'os'* e *'operator'* por exemplo, bem como três bibliotecas específicas: (i) *xlsxwriter*: para criar e gerenciar um documento *Microsoft Excel*; (ii) *re*: para tratar expressões regulares; e (iii) *AdvancedHTMLParser*: utilizada na extração dos dados contidos no *HTML*. A Figura 12 apresenta um exemplo de algoritmo desenvolvido para *Web Scraping* em *Python*.

Os dados coletados via *Web Scraping*, após a importação das bibliotecas necessárias, são geralmente disponibilizados em uma planilha eletrônica, em formato *MS Excel,* por exemplo, cujas ferramentas permitem a correlação e organização dos dados, inclusive na geração de representações gráficas.

Pseudocódigo do Web Scraping em Python

Requer: importar bibliotecas os, $xlsxwriter$, $operator$, $shutil$, $urllib$, re, $AdvancedHTMLParser$.

```
criaDocumento()
termoBusca ← "patent"
codificacao ← "ISO-8859-1"
ano ← 2006
enquanto (ano ≤ 2019) faça
    link ← criaLink(ano, termoBusca)
    pagina ← baixaPaginaHTML(link, codificacao)
    html ← criaAnalisadorHTML(pagina)
    totalArtigosPorArea ← obtemTotalArtigosPorArea(html)
    totalArtigosPorSubarea ← obtemTotalArtigosPorSubarea(html)
    artigos ← obtemArtigos(html)
    qtdArtigos ← obtemQtdArtigos(artigos)
    para cada (artigo ∈ artigos) faça
        titulo ← obtemTitulo(artigo)
        area ← obtemArea(artigo)
        subarea ← obtemSubarea(artigo)
        palavrasChave ← obtemPalavrasChave(artigo)
        ocorrenciasTermoBusca ← obtemNumeroOcorrencias(artigo, termoBusca)
        qtdArtigosPorArea ← obtemQtdArtigosPorArea(html, area)
        qtdArtigosPorSubarea ← obtemQtdArtigosPorSubarea(html, subarea)
        escreveDocumento()
    fim para cada
    ano ← ano + 1
fim enquanto
```

Figura 12 - Algoritmo *Web Scraping* em *Python*

Fonte: Elaborado pelo autor

III.2 Construção do *dataset*

Essa seção apresenta o detalhamento dos procedimentos adotados para construção do *dataset*, abrangendo a definição da amostra, coleta e tratamento dos dados.

II.2.1 Definição da amostra

A definição da amostra é fase crucial para o desempenho da pesquisa, importando na indicação de uma parte que seja representativa do universo a ser representado. Para tanto o perfil da amostragem precisa guardar características da população a ser observada.

Na pesquisa em tela o universo observacional inclui pesquisadores brasileiros afiliados a universidades públicas nacionais e que desenvolvam atividades de patenteamento,

ou seja, tenham solicitação de depósitos de patentes de inventos derivados de sua atividade acadêmica desenvolvida em universidade pública nacional.

Como instituições historicamente voltadas para o ensino e a pesquisa, as universidades têm sido estimuladas a agregar uma terceira missão, vinculada a uma contribuição mais efetiva à sociedade, principalmente no estabelecimento de parcerias denominadas não acadêmicas (DEGL'INNOCENTI *et al*., 2019; COMPAGNUCCI e SPIGARELLI, 2020;). O presente trabalho, especialmente em termos de escopo conceitual e de abordagem metodológica, aborda uma amostra de profícuos patenteadores brasileiros como importantes protagonistas do denominado empreendedorismo universitário a fim de conhecer suas motivações no processo de patenteamento.

A fase de definição da amostra seguiu metodologia própria, se dando em etapas distintas e complementares, conforme Figura 13, a começar pela eleição da base de consulta de patentes registradas em âmbito nacional e respectivos resultados expostos e comentados a seguir.

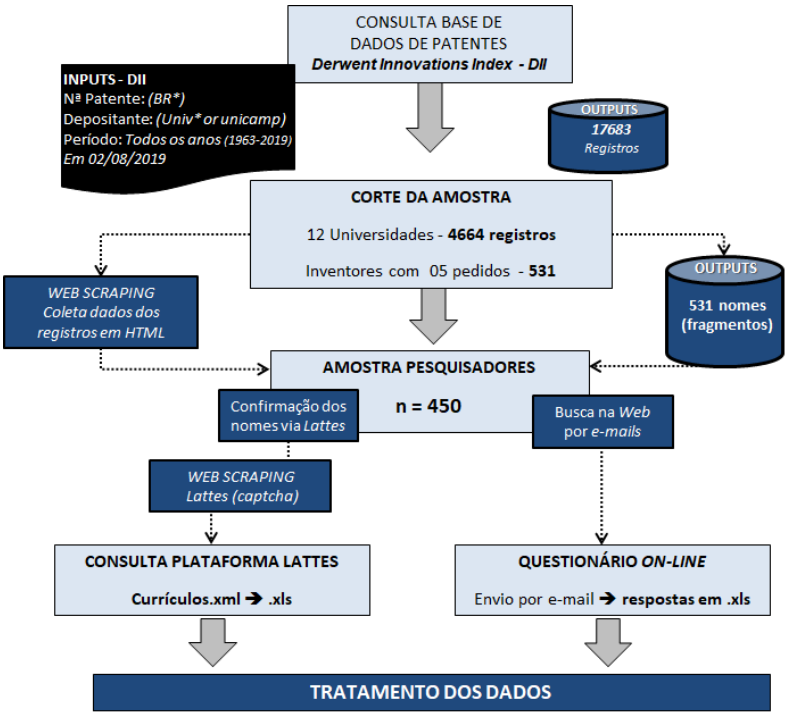


Figura 13 - Etapas de coletas dos dados

Fonte: Elaborados pelos autores

O esforço de pesquisa para definir a amostra se concentrou na seleção das instituições, dos pesquisadores e, por fim, dos registros de patentes para extração dos

dados dos titulares de patentes. A tarefa demandou a escolha de uma base de dados cujo conteúdo pudesse ser selecionado por critérios definidos de acordo com os objetivos da pesquisa.

A *Derwent Innovations Index - DII vinculada à Web of Science/ Clarivate Analytics* é uma base de dados de patentes da *Thomson Reuters*, integrada à plataforma *Web of Knowledge*, aqui eleita em função da reconhecida cobertura geográfica mundial, pela ampla cobertura cronológica desde 1963 e por indexar, no momento da consulta, mais de 33 milhões de referências, abranger vários organismos oficiais e, no que mais afeta a presente pesquisa, permitir acesso integral ao documento da patente, oferecendo inclusive análises estatísticas de resultados por inventor, classificação e tempo.

O acesso à *DII* se deu em 02/08/2019 via portal de periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) com busca realizada via inserção dos seguintes termos:

- Número da Patente: *(BR*)* – busca por patentes depositadas no INPI.
- Nome e código do depositante: *(Univ* or unicamp)* – busca por patentes que tenham universidades como depositante.
- Tempo estipulado: *(Todos os anos)* – desde 1963 até 2019, ano da busca.

O primeiro passo foi definir uma amostra de universidades onde as invenções foram desenvolvidas e que provocaram os registros de patentes. A consulta resultou em 17683 registros de pedidos de patentes, volume de dados bruto encontrado antes da aplicação de filtros de refinamentos.

O primeiro corte da amostra por decisão metodológica se deu a partir do ranqueamento e tomada das 12 instituições com maior número de pedidos de registros de patentes, resultado da consulta à base.

Seguiu-se o detalhamento por universidades públicas federais e estaduais, resultando no seguinte *ranking* por instituição por ordem decrescente de pedidos de patentes, apresentado na tabela 3:

Instituição	Qte. de pedidos de registros de patentes
UNICAMP UNIV ESTADUAL CAMPINAS	1062
UNIV SAO PAULO USP	845
UNIV FEDERAL MINAS GERAIS	642
UNIV FEDERAL DO RIO GRANDE DO SUL	503
UNIV FEDERAL DO PARANA	387
UNIV FEDERAL MINAS GERAIS UFMG	193
UNIV DO RIO DE JANEIRO	185
UNIV FEDERAL DE VICOSA	156
UNIV FEDERAL DO CEARÁ	150
UNIV FEDERAL DA PARAIBA	150
UNIV FEDERAL SERGIPE	141
UNIV FEDERAL RIO DE JANEIRO UFRJ	133
UNIV FEDERAL SANTA CATARINA	126
UNIV FEDERAL PERNAMBUCO	125

Tabela 3 - *Ranking* de universidades públicas por pedidos de registros de patentes.

Fonte: Elaborado pelo autor.

Nesta etapa se verificou problemas, por exemplo, na duplicidade de nomenclatura das entidades depositantes, principalmente dos estados de MG e RJ, gerando a inicialmente indevida, mas superada, sobreposição de dados com potencial efeito negativo no rigor da consulta. A solução se encaminhou pelo agrupamento dos dados encontrados em ambos os casos, importando na eliminação dos registros duplicados. Após refinamento e descontadas as duplicidades, restaram 4664 registros de patentes para a amostra de universidades públicas, cujos arquivos no formado *HTML* foram gerados e arquivados.

Cumpre salientar que a regra para obtenção da amostra considerou as 12 universidades públicas nacionais mais bem ranqueadas conforme registros da base de dados *DII*, restrição em função do caráter amostral da pesquisa e da limitação de recursos se considerada amostra superior. A relação de universidades e a hierarquização, aqui adotada, por número de pedidos de patentes guarda similaridade com a realidade expressa na divulgação oficial, por exemplo, em 2018 pelo INPI e ainda com a média anual (2013-2016) divulgada pela revista Fapesp (2017).

Cumpre salientar que o *ranking* das universidades com maior número de depósitos de patentes, bem como do quadro de pesquisadores considerado, é dinâmico e pode estar sujeito a alterações ao longo do tempo, afetando os resultados obtidos, principalmente nos aspectos subjetivos quanto à percepção dos fatores motivacionais. No entanto, essa amostra pode ser considerada representativa, pois 8 dessas instituições foram listadas no estudo de Fisher *et al*. (2019) na temática que também envolveu 12 universidades com maior número de depósitos de patentes nos anos de 1994/2004/2014. Estatísticas divulgadas pelo INPI (2021) também mostram que as 12 universidades da amostra aparecem no *ranking* de maiores depositantes residentes no ano de 2020.

Segue relação de instituições, doravante tomada como definitiva, processados os ajustes e refinamentos para efeito da presente pesquisa, organizada por ordem alfabética e compondo a amostra da pesquisa.

UNICAMP	Universidade Estadual Campinas
UFC	Universidade Federal do Ceará
UFPB	Universidade Federal da Paraíba
UFPR	Universidade Federal do Paraná
UFRGS	Universidade Federal do Rio Grande do Sul
UFMG	Universidade Federal de Minas Gerais
UFPE	Universidade Federal de Pernambuco
UFRJ	Universidade Federal do Rio de Janeiro
UFSC	Universidade Federal de Santa Catarina
UFS	Universidade Federal de Sergipe
UFV	Universidade Federal de Viçosa
UFRJ	Universidade do Rio de Janeiro
USP	Universidade de São Paulo

Definida a relação de universidades (n=12), procedeu ao refinamento a fim de se obter a amostra dos indivíduos inventores afiliados às instituições. Considerando a abrangência e recursos da pesquisa e, sobretudo, uma amostra das entidades e, por fim, dos pesquisadores mais profícuos, a decisão metodológica visou admitir pesquisadores com pelo menos cinco pedidos de patente na base de dados, reduzindo o número de registros de patentes para 531.

O critério de admissão adotado – o pesquisador contar com, no mínimo, cinco pedidos de patentes – se demonstra plausível já que, considerando a amostra de universidades admitida, a Unicamp (com maior número de pedidos) conta com seu mais proeminente

pesquisador com 39 pedidos de patentes e a UFPE (com menor número de pedidos) tem seu mais proeminente pesquisador com 10 pedidos. Outrossim, para efeito de apuração de frequência um dos trabalhos tomados nesta pesquisa como referência, Goektepe (2008), adota o corte de três patentes por patenteador.

Em suma, considerados o arrolamento de 12 universidades e o filtro por inventores com pelo menos cinco registros de patentes se obteve a relação com as informações coletadas dos registros de patentes, organizados no formato de planilha eletrônica, sendo possível a apuração de uma amostra provisória com 531 registros.

Desconsiderados registros do mesmo pedido em nome de autores diferentes e duplicidades, foi obtida a amostra final de 450 pesquisadores/inventores, assim considerados os que figuraram no registro de patentes como nome principal do pedido de patente registrado.

Destaque-se que os nomes dos pesquisadores informados nos registros se apresentaram fragmentados, em maioria compostos a partir do sobrenome agregado a fragmentos do nome, sendo necessária a confirmação e apuração dos nomes completos. A solução adotada considerou, a partir dos fragmentos extraídos dos registros, empreender busca/consulta manual no sítio de currículos acadêmicos da plataforma *Lattes* (https://www.lattes.cnpq.br/) inserindo no campo de pesquisa o nome (ou parte dele) de pesquisadores, visando à identificação completa e verificação da relação do nome com o respectivo currículo *Lattes*, em cuja plataforma se extraiu os dados relacionados à vida acadêmica dos integrantes da amostra.

O recorte da amostra de instituições e registros de patentes foi determinante na definição da amostra de titulares de patentes. Dos 531 registros colhidos, foram admitidos 450 pesquisadores como amostra final.

III.2.2 Coleta dos dados

Definida a amostra de pesquisadores se procedeu a coleta dos dados, cujo método e processo estão representados na figura 14. A partir de um esforço desenvolvido por uma equipe multidisciplinar que incluiu especialistas em computação, um algoritmo na linguagem *Python* foi desenvolvido a fim de aplicar a técnica *web Scraping*, permitindo que os dados principais de cada registro pudessem ser extraídos com precisão dos arquivos dos registros de patentes baixados no formato *HTML*. A técnica permite

extrair dados de arquivos a fim de estruturá-los, facultando sua visualização e análise (SAURKAR *et al*., 2018; DIOUF *et al*., 2019).

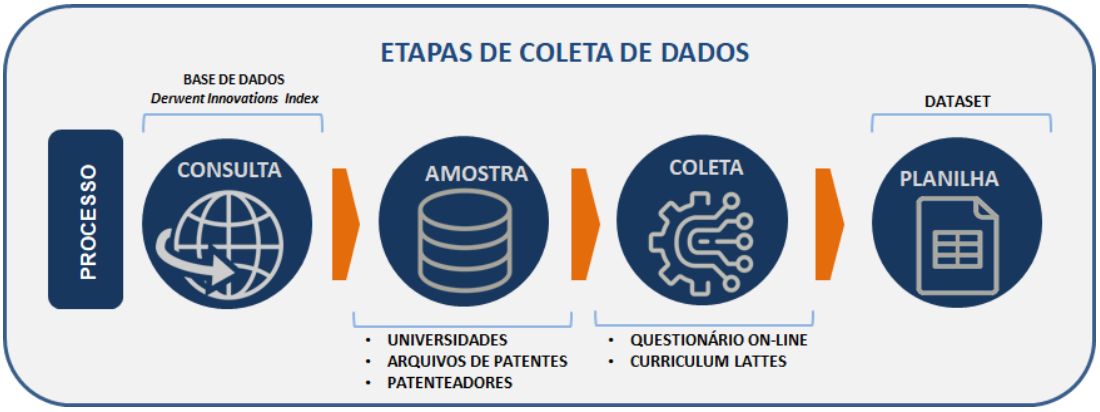


Figura 14 - Processo de coleta dos dados

Fonte: Elaborados pelos autores

A extração dos dados se deu a partir de tais parâmetros e observações:

- INVENTOR (AU) – geralmente há mais de um inventor sendo importante correlacionar com os nomes dos inventores com o *Lattes* (considerar as diferentes formas dos nomes e respectivas citações)
- TITULAR DA PATENTE (AE) – Universidade (verificar parcerias com outras entidades públicas ou privadas)
- PATENT DETAILS (PD)
    - Main IPC* (identificar a classificação internacional de patentes observando a que área de pesquisa está vinculada a patente)
    - Publ. Date**

Patent Number	Publ. Date**	Main IPC*	Week	Page Count	Language
BR102017027480-A2	09 Jul 2019	C12P-041/00	201956	Pages: 13	English

- DETALHES DO DEPÓSITO E DATAS (AD) – data de pedido de registro da patente (importante para gerar histórico de pedidos/ano/titular)
- Gerar número de registro de patentes Total/por universidade x Ano (histórico)

BR102017027480-A2	BR10027480	19 Dec 2017

III.2.2.1 Coleta através da Plataforma *Lattes*

Com a confirmação do quantitativo e dos respectivos nomes dos pesquisadores admitidos na amostra, duas foram as fontes de dados que se consolidaram, sendo acessadas para coleta de dados.

A primeira fonte de dados se dá pelo acesso aos dados públicos dos currículos acadêmicos dos 450 pesquisadores na plataforma governamental *online* de currículo acadêmico, a plataforma *Lattes*.

O objetivo do emprego do método adotado consiste em acessar o currículo acadêmico de cada pesquisador admitido na amostra a fim de coletar os dados pertinentes à pesquisa, como nome – principal dado para cruzamento com outras fontes –, filiação acadêmica, áreas de pesquisa, quantidades de patentes, produção científica, etc.

Nesta etapa a modalidade de consulta manual foi inicialmente considerada, mas, além de trabalhosa estaria sujeita a imprecisões. Assim, foi empregada novamente a técnica *web Scraping* a fim de extrair os dados necessários à pesquisa. Contudo, a plataforma *Lattes* emprega o *captcha* (acrônimo da expressão "*Completely Automated Public Turing test to tell Computers and Humans*"), recurso comum para se evitar acessos não humanos a sítios eletrônicos. A solução foi acessar manualmente cada currículo na plataforma *Lattes* e, a partir de um recurso disponível na plataforma, baixar cada currículo manualmente no formato *.XML* e, finalmente, aplicar a referida técnica para a extração robotizada dos dados via *web Scraping*.

A extração pretendeu a tais informações:

- Nome (para cruzar com os dados oriundos do questionário – o nome é o único dado para cruzar com os dados do *Lattes*)
- Nome em citações bibliográficas (importante para cruzar com o nome de inventores das patentes)
- Formação acadêmica/titulação (para verificar áreas e região geográfica de formação – instituição, local, ano)
- Atuação profissional
    - Último vínculo institucional (verificar atual filiação do inventor)
    - Linhas de pesquisa (comparar com pesquisa que resultou em patente)

    - Áreas de atuação (comparar com área de pesquisa que resultou em patente)
- Artigos completos publicados em periódicos (interesse na quantidade total apenas)
- Capítulos de livros publicados (interesse na quantidade total apenas)
- Trabalhos completos publicados em anais de congressos (interesse na quantidade total apenas)
- Resumos publicados em anais de congressos (interesse na quantidade total apenas)
- Patentes e registros
    - Patente (interesse na quantidade total)

Da extração resultou relação dos currículos *Lattes* no formato de arquivo *XML* zipados apurada com os dados disponíveis nos *Lattes* dos 450 pesquisadores pertencentes a amostra. O código-fonte aplicado na geração da citada relação, dentre outros, está disponível em repositório conforme informações apresentadas no Apêndice D.

III.2.2.2 Coleta através de questionário online

A segunda fonte de dados tem a finalidade de complementar as informações obtidas na plataforma *Lattes*. Uma vez definidas as universidades, selecionados os inventores e baixadas as principais informações do *Lattes*, a fase seguinte adotou a estratégia metodológica de colecionar as informações adicionais a partir da aplicação de questionário *on-line* desenvolvido via *google forms* para esta finalidade, sendo então necessária a obtenção dos endereços eletrônicos dos pesquisadores da amostra. O modelo do questionário aplicado está no Apêndice A.

Cumpre salientar que atendendo à metodologia geral considerada no presente trabalho, o questionário se desenvolveu a partir dos conceitos, variáveis, parâmetros e referencial bibliográfico obtido da revisão sistemática e já evidenciados no Capítulo II.

O questionário foi organizado em duas partes. Na primeira foram feitas perguntas para identificação do pesquisador incluindo dados pessoais e informações sobre patentes, visando saber se foram feitos depósitos internacionais e se o pesquisador teve patentes concedidas e licenciadas, bem como, patentes decorrentes da interação com a indústria.

A segunda parte do questionário teve como propósito identificar as motivações do pesquisador no tocante às atividades de patenteamento. Assim sendo, foi solicitado que o pesquisador indicasse a importância de cada um dos 27 fatores motivacionais que foram identificados na revisão sistemática, de acordo com a seguinte escala: extremamente importante; muito importante; moderadamente importante; pouco importante; e nem um pouco importante.

Ressalta-se que o referido questionário foi submetido e aprovado pelo Comitê de Ética em Pesquisa, através da Plataforma Brasil. O questionário, assim como o Termo de Consentimento Livre e Esclarecido (TCLE), encontram-se disponíveis nos Apêndices A e B, respectivamente.

Para que a amostra de 450 pesquisadores tivesse acesso ao questionário se empreendeu dois movimentos. Consulta às universidades por meio de envio de *e-mail* aos NITs, ou equivalente, em cada instituição, apresentando a intenção de pesquisa e disponibilizando o Termo Circunstanciado de Livre Esclarecimento (TCLE) – disponível no Apêndice B –, além de enviar relação anexa com os nomes dos pesquisadores da instituição consultada a fim de receber seus respectivos *e-mails* ou propondo que a instituição repassasse o questionário aos pesquisadores afiliados e pertencentes à amostra. Outro movimento foi o de coletar manualmente os *e-mails* dos pesquisadores disponíveis nos *sites* das instituições a que estavam filiados e em outras fontes digitais de divulgação de conteúdo científico, sendo este último, apesar de trabalhoso, o que trouxe melhor resultado.

Considerando o volume de 450 pesquisadores selecionados da base de dados *DII*, cujos dados foram confirmados na base de *currículos Lattes*, foram obtidos com sucesso os endereços eletrônicos de 437 pesquisadores. Para estes foram remetidos *e-mails* com a apresentação da pesquisa, detalhes e o *link* para responder ao questionário *on-line*.

Foram realizadas cinco ondas de envios para os respectivos *e-mails* obtidos visando aumentar a taxa de respondentes. As datas de envio foram 26/03/2020; 01/04/2020; 18/06/2020; 02/09/2020 e 01/11/2020. A cada envio os respondentes foram desconsiderados da próxima onda e realizados correções e ajustes de contato/*e-mail*. O envio em diferentes e adicionais ondas foi uma estratégia adotada para, além de obter mais respostas, superar o desafio de contatar os respondentes que, devido à

pandemia do *Corona Virus Disease* (COVID-19) desde os primeiros meses do ano de 2020, estavam em trabalho remoto, com aumento na carga de trabalho e/ou em teletrabalho como docente/pesquisador, podendo ainda estar licenciado ou afastado por residir/trabalhar em área com deficiência de redes de *internet*. Após os citados envios, se obteve o número de 131 respostas. Considerado o envio para 437 pesquisadores, se tem a taxa de reposta superando 30% de respondentes.

III.2.3 Estruturação do *dataset*

Os dados coletados na plataforma *Lattes* e das respostas aos questionários foram submetidos ao cruzamento de informações a partir do atributo comum – nome do pesquisador, obtido nas duas fontes de dados. Considerando eventuais diferenças nos nomes causados por usos de abreviações, omissões de partes ou erros de preenchimento, foi realizado o casamento aproximado de cadeias de caracteres dos nomes, visando obter um padrão de correspondência já que não se pode contar com coincidências perfeitas. Com o respectivo cruzamento e consolidação de informações se deu a construção de um *dataset* que aglutina e organiza os dados em planilha eletrônica.

O volume de dados é do tipo primário, sendo que as informações legalmente denominadas como sensíveis foram anonimizadas visando preservar a identidade dos participantes da amostra e atender a lei geral de proteção de dados pessoais (LGPD) Por fim, o endereço eletrônico do repositório em que o *dataset* está armazenado é disponibilizado no Apêndice C.

O conjunto de dados quantitativos e qualitativos no *dataset* é abrangente e diversificado: contém aspectos objetivos e sociodemográficos como ano de nascimento, gênero, histórico geográfico, dados acadêmicos como produção científica e áreas de pesquisa e profissionais como afiliação a instituições de pesquisa, além de quantidades de patentes e, por fim, de atributos subjetivos, como características comportamentais e motivacionais notadas e inferidas a partir das respostas às diversas questões do questionário. Esse conjunto de dados, além de servir de insumo para a presente investigação, fornece subsídios para o estabelecimento de políticas mais assertivas de incentivo à inovação e patenteamento nas universidades.

CAPÍTULO IV – PERFIL SOCIODEMOGRÁFICO, PATENTEAMENTO E FATORES MOTIVACIONAIS DOS PESQUISADORES BRASILEIROS

Esse capítulo apresenta o perfil sociodemográfico dos pesquisadores brasileiros e o posicionamento deles quanto ao patenteamento e diversos fatores motivacionais identificados nesse estudo.

Os resultados tiveram como base o *dataset* elaborado conforme descrito no Capítulo III, que reúne atributos sociodemográficos, fornecendo também informações sobre áreas de pesquisa e produção científica, além de características relacionadas a aspectos comportamentais e motivacionais da atividade de patenteamento.

É importante ressaltar que o perfil sociodemográfico e o posicionamento em relação ao patenteamento e aos fatores motivacionais foram identificados a partir dos dados públicos, coletados a partir dos *curricula vitae* na Plataforma Lattes, abrangendo um total de 450 pesquisadores (n=450) e, a partir dos dados institucionais, coletados, por sua vez, das respostas de questionário, abrangendo uma subamostra de 131 pesquisadores (n=131).

IV.1 Perfil sociodemográfico

Quanto aos dados sociodemográficos, são apresentados, dentre outros, intervalo de nascimento, gênero e local de nascimento. O modo de apresentação do intervalo de nascimento visa garantir a anonimização dos entrevistados. Para o local de nascimento dos pesquisadores brasileiros, foi adotado o formato da sigla da unidade da federação (estados) que representa uma federação do país. No caso de estrangeiros, foi indicado o nome do país de nascimento.

Do agrupamento e da correlação estatística dos dados é possível conceber um perfil sociodemográfico e também acadêmico da amostra. A tabela 4 toma o atributo intervalo de nascimento como base para ilustrar a composição da amostra dos respondentes ao questionário quanto ao gênero e local de nascimento referente aos dados institucionais dos respondentes ao questionário (n=131).

Intervalo de nascimento	Total	Mulher	Homem	Brasileiros	Estrangeiros
1940-1944	1	0 (0%)	1 (100%)	0 (0%)	1 (100%)
1945-1949	11	1 (9%)	10 (91%)	11 (100%)	0 (0%)
1950-1954	8	2 (25%)	6 (75%)	8 (100%)	0 (0%)
1955-1959	14	3 (21%)	11 (79%)	14 (100%)	0 (0%)
1960-1964	25	11 (44%)	14 (56%)	22 (88%)	3 (12%)
1965-1969	18	7 (39%)	11 (61%)	18 (100%)	0 (0%)
1970-1974	19	7 (37%)	12 (63%)	18 (95%)	1 (5%)
1975-1979	13	4 (31%)	9 (69%)	13 (100%)	0 (0%)
1980-1984	10	5 (50%)	5 (50%)	9 (90%)	1 (10%)
1985-1989	10	5 (50%)	5 (50%)	10 (100%)	0 (0%)
1990-1994	2	2 (100%)	0 (0%)	2 (100%)	0 (0%)
Total	**131**	**47 (36%)**	**84 (64%)**	**125 (95%)**	**6 (05%)**

Tabela 4 - Amostra por intervalo de nascimento, gênero e local de nascimento.
Fonte: Elaborado pelo autor

Tem-se que aproximadamente 75% dos respondentes nasceram nos anos 50 e 70 e a média da amostra é de 55 anos. Apesar da predominância do gênero masculino (64%) se observa significativa redução nesta diferença no decorrer do tempo até a representação equitativa de gêneros a partir de 1980. Observa-se ainda que 5% da amostra é de estrangeiros que constam como titulares de pedidos de patentes.

Não se empreendeu a comparação de dados da idade média da amostra do estudo (55 anos) com a composição geral das 12 universidades aqui consideradas. Contudo, os resultados do censo escolar 2021, realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), uma autarquia vinculada ao Ministério da Educação, revelou que mais de 300 mil professores atuam no ensino superior brasileiro. Os homens são maioria e a idade média docente dos que trabalham em instituições públicas é de 39 anos (INEP, 2021). Comparado com a média nacional e considerando o escopo desta pesquisa, o trabalho foi bem sucedido em reunir as

percepções dos pesquisadores com maior experiência e resultados no processo de pedido de patentes.

Também é apontada a localização das instituições em que ocorreu a formação acadêmica dos pesquisadores nos cursos de bacharelado, mestrado e doutorado: para instituições brasileiras, a unidade de federação de localização é indicada; para instituições estrangeiras, a denominação dos países. O gráfico 1 demonstra o panorama da internacionalização educacional. Da amostra com 450 pesquisadores, com base nos dados públicos, têm-se o predomínio da formação em território nacional. No doutorado se percebe elevação da internacionalização em quase 20% da amostra. O código-fonte da geração deste e de outros gráficos pode ser acessado em repositório conforme informações apresentadas no Apêndice D.

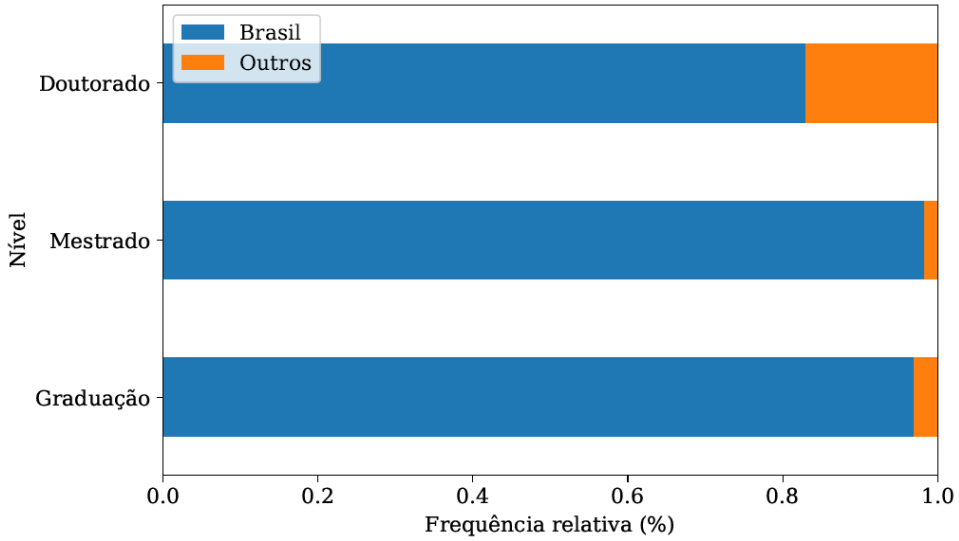


Gráfico 1 - Internacionalização educacional da amostra
Fonte: Elaborados pelo autor

No que tange às áreas de investigação e de produção científica, o gráfico 2 apresenta a preponderância nas áreas de pesquisa em saúde, natureza, engenharia e tecnologia, pela alta geração de soluções absorvidas pela sociedade, com maiores incentivos governamentais e pelo interesse no licenciamento e comercialização pela indústria. Nestas áreas o pesquisador desenvolve os seus projetos de investigação e destas derivam as invenções patenteáveis.

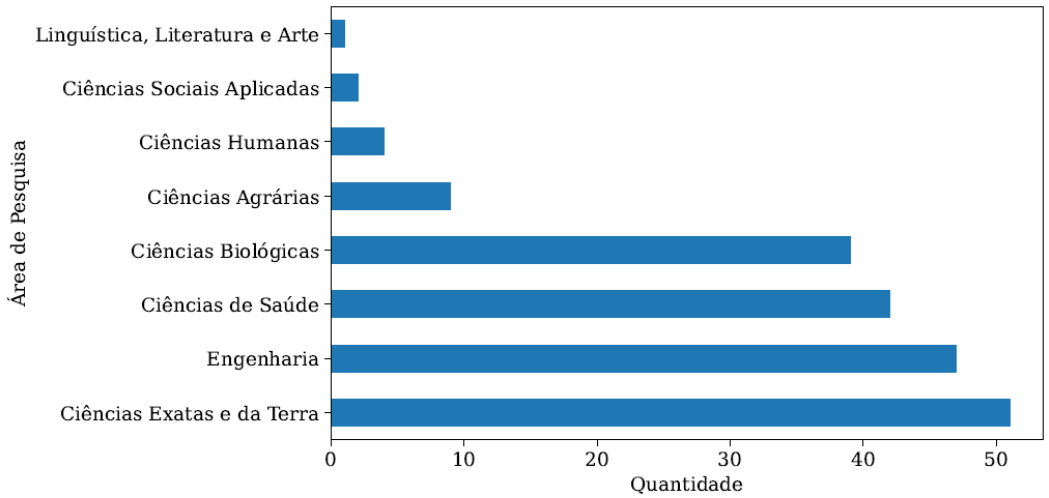


Gráfico 2 - Distribuição por áreas de pesquisa

Fonte: Elaborados pelo autor

Quanto à carreira acadêmica foram produzidas outras representações gráficas que apresentam a correlação dos atributos afeitos à produção acadêmica, como publicação de artigos em periódicos e congressos, de capítulos em livros e resumos expandidos. O atributo relacionado ao número de pedidos de patentes é agregado a fim de compará-lo aos demais indicadores de produção científica. Neste sentido o gráfico 3 traz um histograma que demonstra o resultado acumulado das quantidades de registros de patentes por pesquisador. Este e alguns dos próximos gráficos estão gerados na base logarítmica que reduz a escala gráfica e facilita a representação, oferecendo melhor visualização da proporção dos valores.

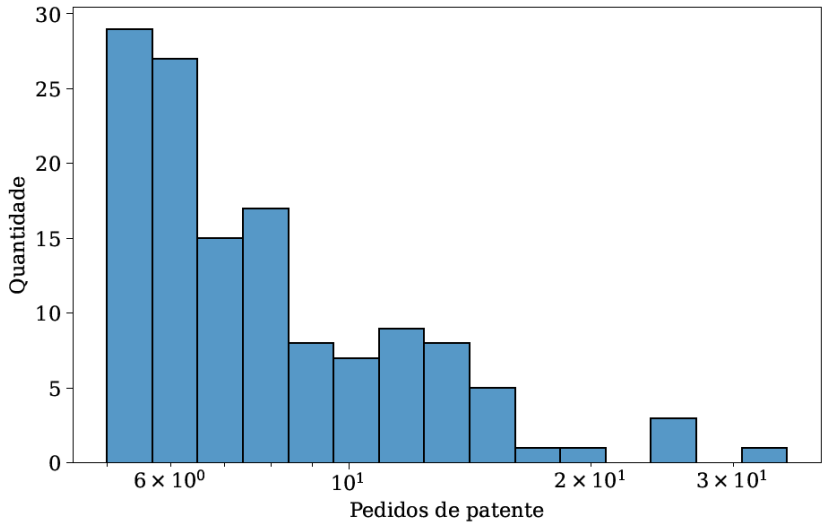


Gráfico 3 - Números de pedidos de patentes.

Fonte: Elaborados pelo autor

O gráfico 4 é um modelo de dispersão ou *scatter plot* em que são correlacionados quatro atributos afeitos à produção científica da amostra de 450 patenteadores. O gráfico de dispersão mais comum é o bivariado com duas variáveis contínuas. Quando há uma terceira ou quarta variável se tem um gráfico de bolhas, versão aqui adotada.

Cada bolha representa um patenteador. Os atributos quanto a publicação de artigos em periódicos e congressos (*papers in journals e conferece proceeedings)*, ambos em escala logarítmica, estão dispostos nos eixos *y* e *x*, respectivamente. A maior quantidade de resumos publicados em congressos (*abstracts in conference proceeding)* está simbolizada nas bolhas em tons mais escuros. O volume de capítulos de livros (*book Chapters)* se retrata na dimensão das bolhas. Comparando as múltiplas variáveis no gráfico e desconsiderando eventuais pontos isolados (*outliers)* se percebe uma concentração da amostra entre pesquisadores com mais de 50 artigos em periódicos e congressos e a partir de 10 trabalhos em congressos/ conferências. Quanto a resumos em conferências predomina números abaixo de 500, enquanto a capítulos de livros não supera 60.

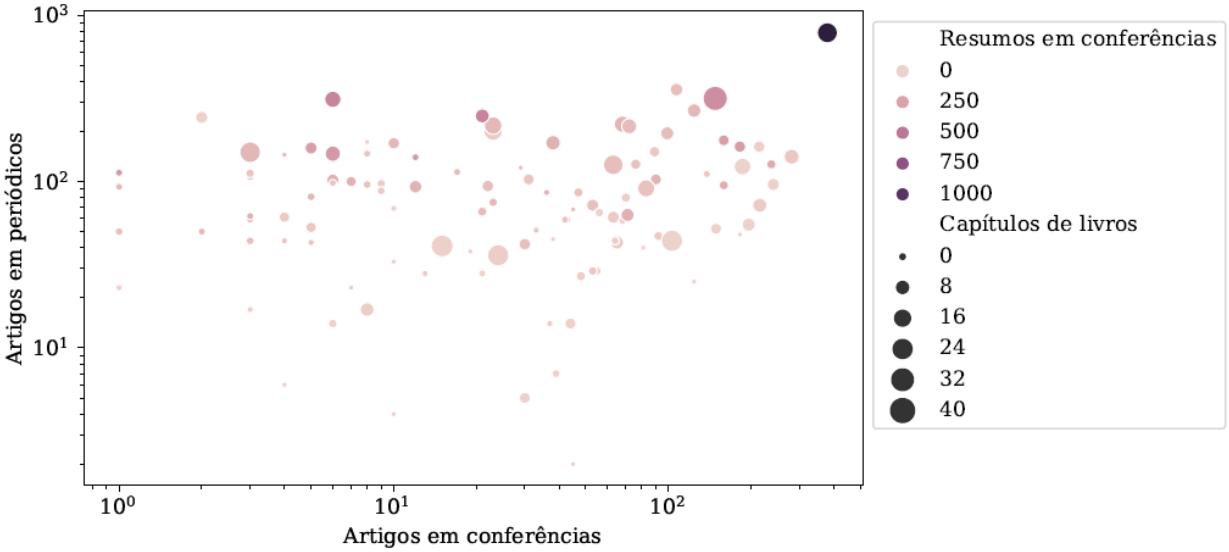


Gráfico 4 - Gráfico de dispersão da produção acadêmica

Fonte: Elaborados pelo autor

Os mesmos atributos são espelhados no gráfico 5 que apresenta o gráfico de função de distribuição cumulativa empírica (*Empirical Cumulative Distribution Function*) da amostra de 450 patenteadores, sendo adicionado um quinto atributo referente ao número de pedidos de registros de patentes. Tal modelo permite uma análise

estatística da distribuição de dados, pois resulta da prática e da observação do mundo real e, portanto, é empírico.

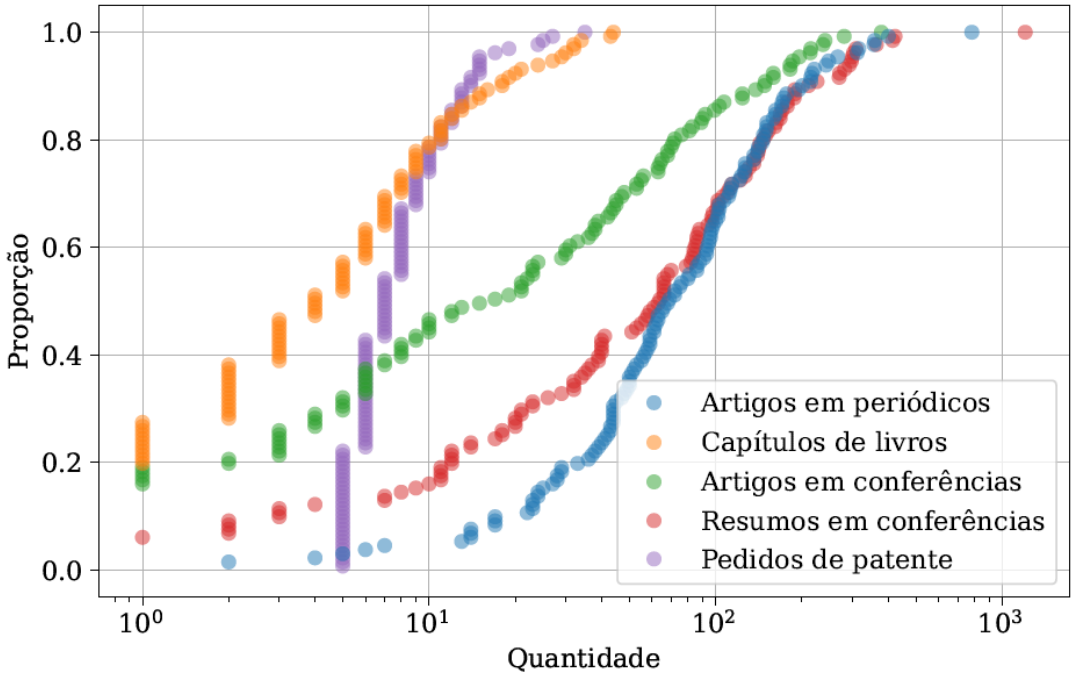


Gráfico 5 - Gráfico função de distribuição cumulativa empírica da produção acadêmica
Fonte: Elaborados pelo autor

No eixo *x* em base logarítmica os atributos se orientam em ordem crescente dos dados quantitativos, enquanto o eixo *y* se apresenta em uma escala linear, tendo o mínimo no '0' e o máximo em '1'. Da comparação com a variável referente às patentes, como se deduz de Burhan *et al*. (2017) e Blind *et al. (*2018), é possível interpretar que há alinhamento entre a produção científica e patenteamento. Fica manifesta a relação entre patenteamento e número de publicações como apontado por Stephen *et al*. (2005) e Souza *et al*. (2011). Patenteadores são também prolíferos na produção científica, ficando cada vez mais superado o dilema citado por Perucchi e Mueller (2014) quanto ao pesquisador que oscila entre o prestígio acadêmico público, geralmente decorrente da publicação científica e a garantia no anonimato do domínio da inventividade para si, patenteando seu invento.

O Gráfico 6 ilustra a comparação entre os dados do pedido de patente e da publicação científica. Em relação à curva de pedidos de patentes, cabe destacar que: seu percurso inicia-se no limite mínimo de corte permitido na amostra (cinco patentes); os dados de

patenteamento têm volume nominal inferior aos dados referentes à publicação científica; e a maioria dos titulares de patentes tem entre 5 e 10 pedidos de patente.

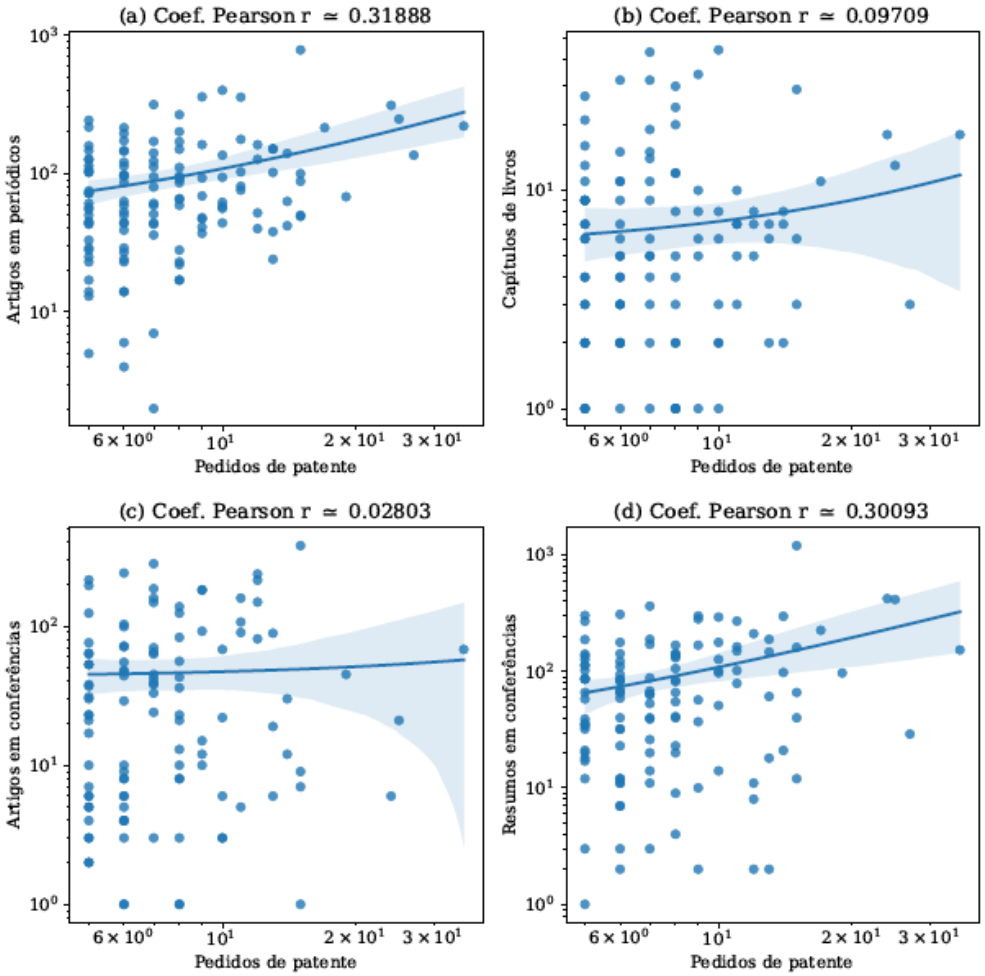


Gráfico 6 - Correlação entre indicadores de patenteamento e publicação científica
Fonte: Elaborados pelo autor

Quanto ao alinhamento entre produção científica e patenteamento, a figura com gráficos de correlação compara e ilustra cada indicador de publicação, verificando sua correlação com os pedidos de patentes. Nenhuma combinação de indicadores comparados exibe uma correlação negativa. O maior valor de r está exatamente entre os números de pedidos de patentes e registros e os artigos completos publicados em periódicos, indicando uma correlação notavelmente positiva ($p < 0{,}001$).

Uma análise paralela que deriva dos dados colecionados, sobretudo aqueles oriundos plataforma *Lattes*, se dá quanto à mobilidade acadêmica de pesquisadores que Ackers (2005) denomina de *brain drain* que em uma tradução livre seria “fuga de cérebros”. Um movimento que caracteriza um fato da vida natural, mas que pode representar um

fenômeno de observação à medida que explicita a movimentação de pesquisadores durante seu processo de formação acadêmica.

Considerando os dados dos pesquisadores que informaram a data de nascimento por ocasião do preenchimento do questionário de pesquisa, se tem as seguintes representações que evidenciam a migração acadêmica. Os gráficos 7 a 9 representam dados binários, comparando, respectivamente, o local de nascimento com o local de realização do bacharelado, o local do bacharelado com o do mestrado e, por fim, o local do mestrado com o doutorado.

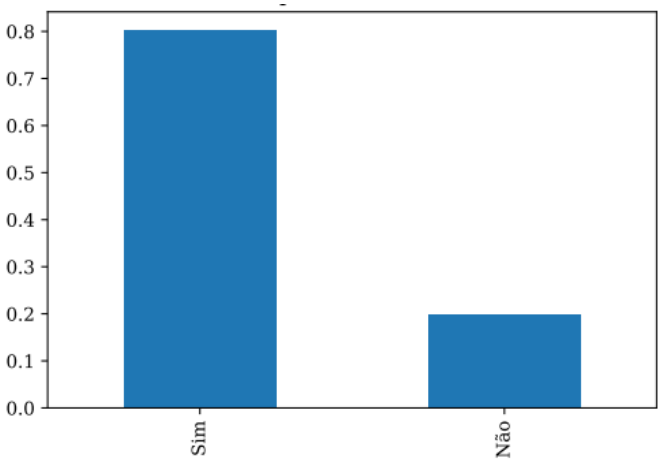


Gráfico 7 - Local de nascimento = local do bacharelado?
Fonte: Elaborados pelo autor

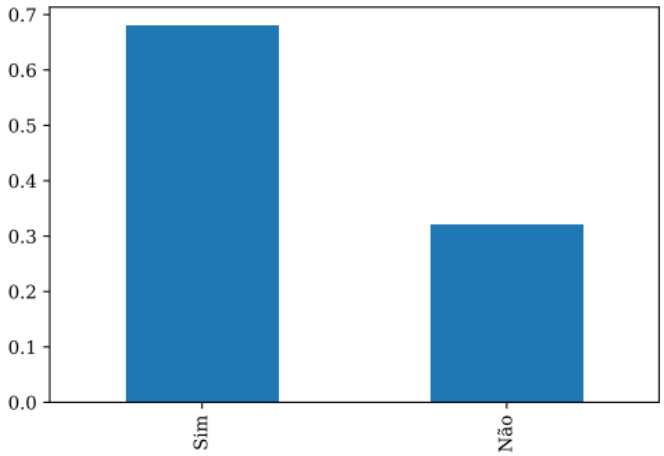


Gráfico 8 - Local do bacharelado = local do mestrado?
Fonte: Elaborados pelo autor

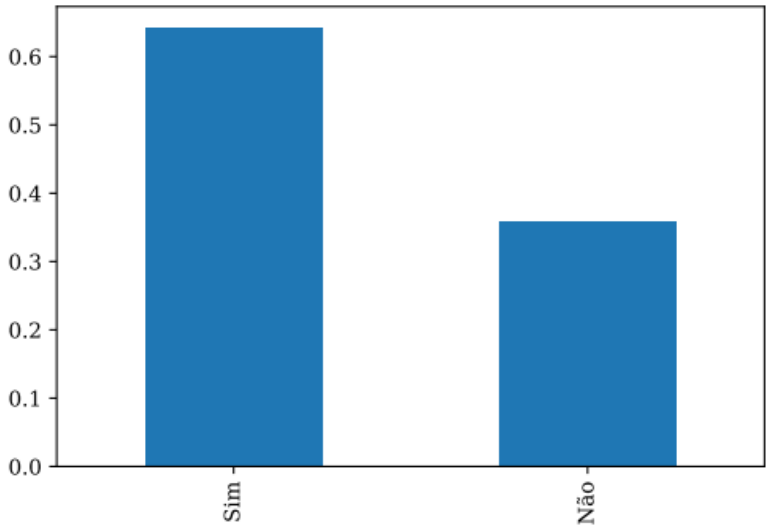


Gráfico 9 - Local do mestrado = local do doutorado?
Fonte: Elaborados pelo autor

Dos três gráficos expostos se verifica a migração ocorrendo na proporção do avanço na especialização acadêmica. Ampliando este ponto e a fim de melhor ilustrar o fluxo da migração acadêmica de pesquisadores, considerando os dados coletados e a correspondente limitação geográfica dos mesmos, segue o gráfico 10 que retrata a situação da amostra dos respondentes (n=131), considerando o local de nascimento. As colunas representam estágios a partir do local de nascimento seguido do local de formação nas fases de formação na graduação, mestrado e doutorado. As linhas horizontais indicam pela espessura a intensidade do fluxo. Os rótulos são as denominações dos estados, entes federativos brasileiros. Os números representam a quantidade do fluxo de pesquisadores. O gráfico 11, sem a informação do local do nascimento, retrata a situação da amostra ampla (n=450).

Aqui uma análise adjacente é feita sobre a mobilidade acadêmica dos pesquisadores, representada nos gráficos. Como pode ser observado, para os cursos de bacharelado, mestrado e, principalmente, doutorado, há um fluxo para a região de São Paulo e outras regiões do Sudeste do Brasil. Por fim, principalmente na fase de doutorado, destaca-se o fluxo para o exterior.

Esse é um movimento que pode impactar o processo de pesquisa e formação acadêmica. É importante considerar que o Brasil é um país de dimensões continentais e com grandes assimetrias regionais. O país é dividido em cinco regiões contendo 27 Unidades da Federação (UF). A região Sudeste (composta por São Paulo, Rio de

Janeiro, Minas Gerais e Espírito Santo) concentra mais de 50% do Produto Interno Bruto (PIB) do país. Só o estado de São Paulo responde por mais de 30% do PIB nacional. É a região com a maior parcela da população e infraestrutura industrial do Brasil.

Essa assimetria econômica também acontece em relação ao sistema de Ciência, Tecnologia e Inovação. A região Sudeste, com protagonismo do estado de São Paulo se destaca pela quantidade de programas de pós-graduação, reunindo importantes e reconhecidas universidades e institutos de pesquisa. Por outro lado, alguns estados não possuem condições adequadas para a formação de pessoal qualificado em nível de doutorado em diversas áreas do conhecimento.

Portanto, esse fluxo migratório pode evidencar a mobilidade dos pesquisadores dentro do país em busca de melhores condições de formação e desempenho acadêmico, bem como, de um ecossistema de inovação mais consolidado diante da realidade nacional. Trata-se de uma dinâmica semelhante ao que Ackers (2005) chama de fuga de cérebros (*brain drain*), mas, nesse caso, dentro das fronteiras do país.

Conforme sugerido por Bolli e Schläpfer (2015), considerando a amostra altamente qualificada dos quadros de pesquisadores universitários, os principais fatores de mobilidade podem estar associados à reputação e busca por maior autonomia, acesso a fontes de pesquisa e progressão na carreira. Para Ge *et al.* (2016) é um fenômeno de realocação de capital intelectual com efeitos na inovação e no empreendedorismo.

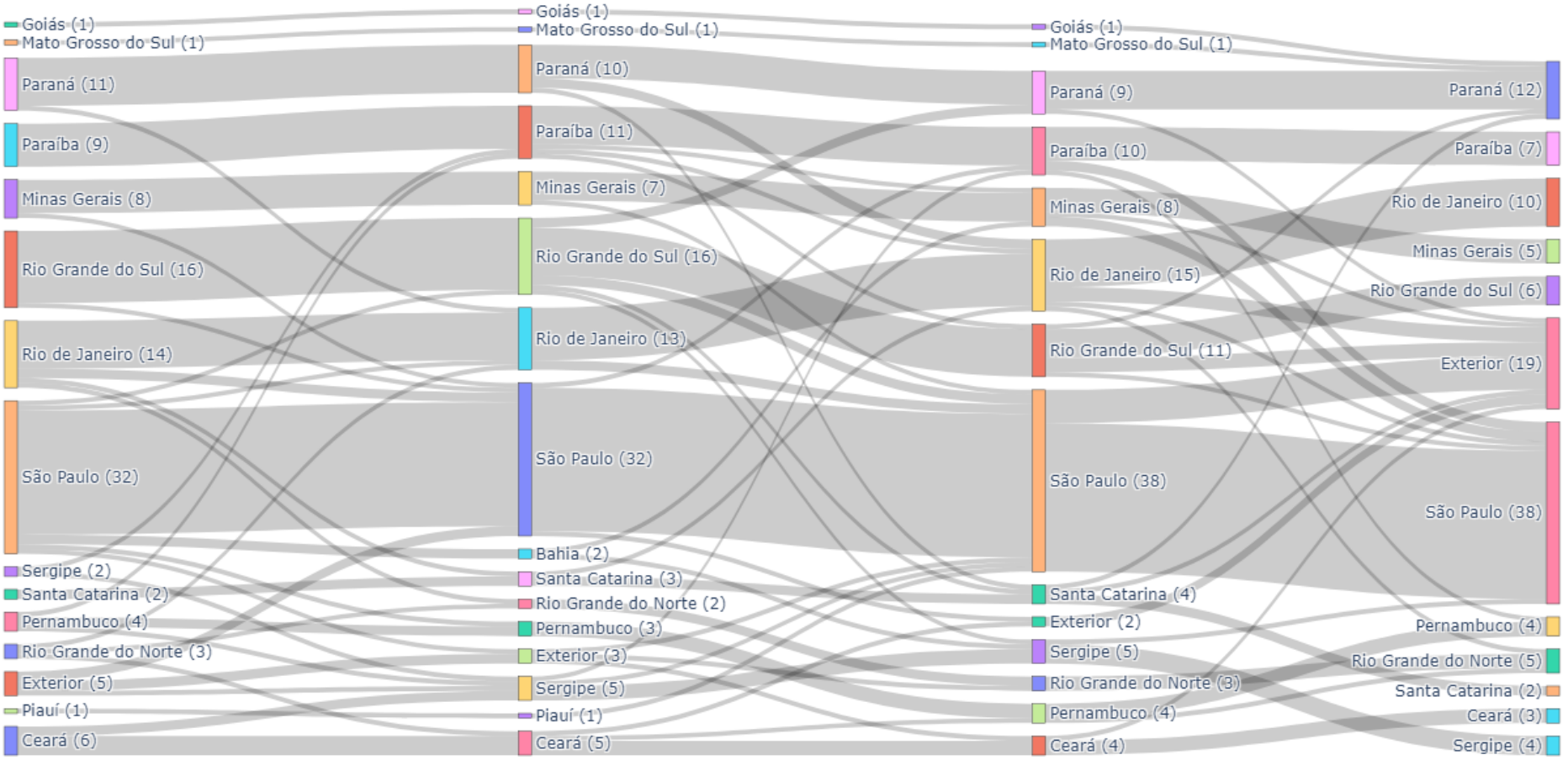


Gráfico 10 - Fluxo de migração acadêmica de pesquisadores (n=131).

Fonte: Elaborados pelo autor

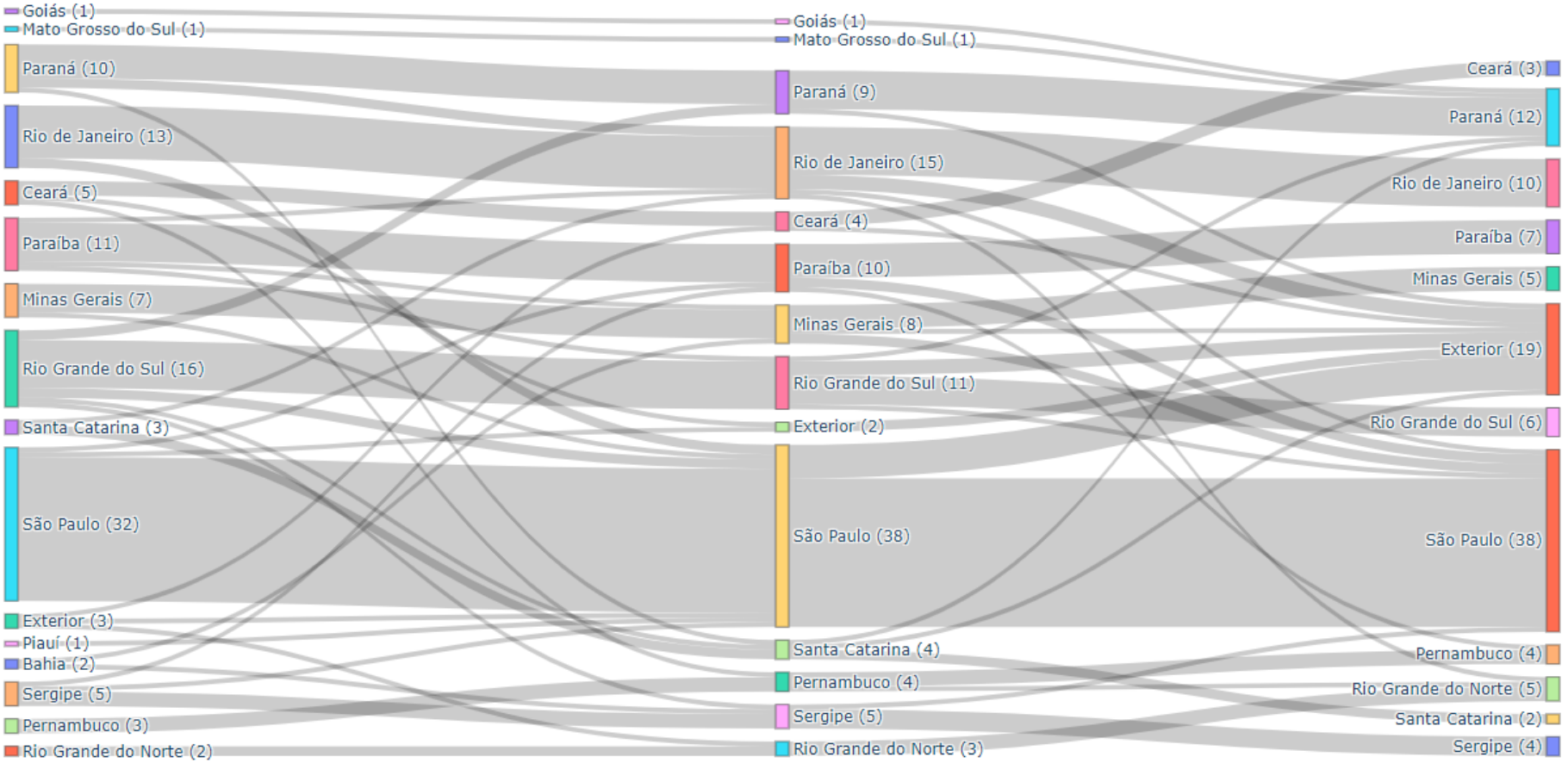


Gráfico 11 - Fluxo de migração acadêmica de pesquisadores (n=450).

Fonte: Elaborados pelo autor

IV.2 Posição quanto ao patenteamento

Considerando as informações extraídas da resposta ao questionário, foram realizadas algumas perguntas relacionadas ao processo de patenteamento cujo panorama está no gráfico 12.

Quanto à concessão de patentes quase 70% (90 respondentes) tiveram o pedido de registro analisado e concedido. Pela indisponibilidade da informação, não é possível mensurar o tempo entre o pedido e a concessão, tema abordado por Garcez e Moreira (2017) quanto ao fenômeno do *backlog*. Referente ao licenciamento de patentes mais de 65% (87 respondentes) outorgaram a terceiros o direito de uso ou exploração da invenção patenteada.

Quanto a internacionalização do processo de patentes os pesquisadores responderam à pergunta; "*Você já teve alguma patente depositada no exterior ou via Tratado de Cooperação de Patentes (PCT)*?". Ao que 60% (79 respondentes) responderam negativamente, demonstrando não terem patentes registradas nos países membros do PCT. Na apuração da relação com o meio privado mais de 50% (72 respondentes) afirmaram que a patente requerida não resultou de interação com a indústria.

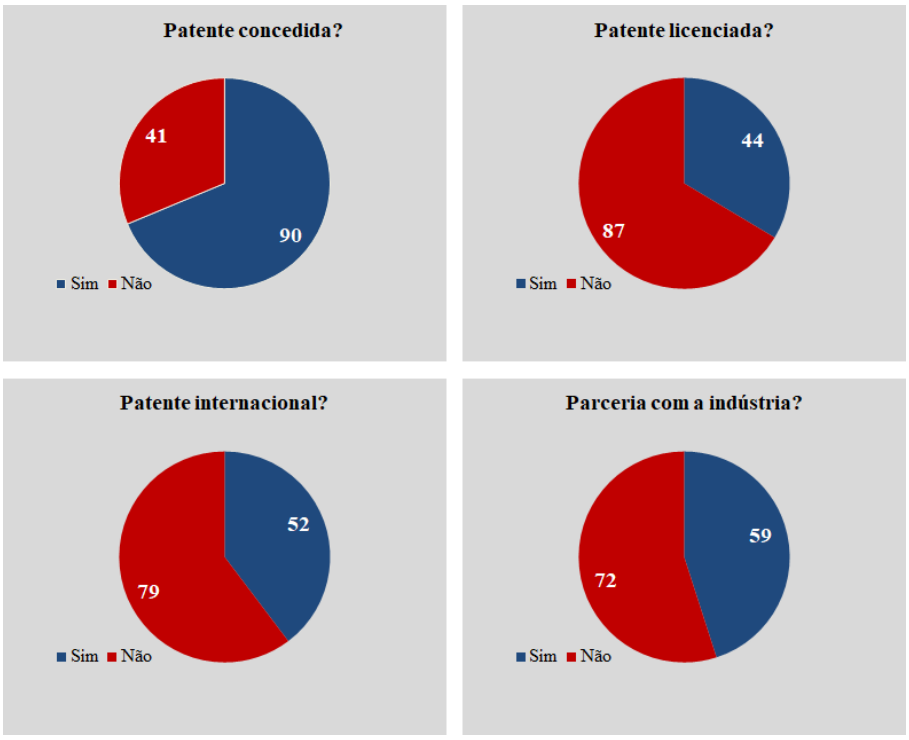


Gráfico 12 - Dados do processo de patenteamento
Fonte: Elaborados pelo autor

Tomando por referência o trabalho desenvolvido por Lam (2011), que analisou o processo de patenteamento com vistas à comercialização, a pesquisa coletou a percepção dos pesquisadores que responderam ao questionário (n=131) quanto à colaboração na relação academia-indústria, importando em quatro perfis ou orientações profissionais possíveis como descritas a seguir e ilustradas no gráfico 13.

Empreendedor- acredita na importância fundamental da colaboração acadêmico-industrial e desenvolve atividades de vínculos industriais para aplicação e exploração comercial;

Híbrido - acredita na importância fundamental da colaboração acadêmico-indústria e desenvolve atividades de vínculos industriais para o avanço científico;

Pragmático Tradicional - acredita que a colaboração acadêmico-indústria deve ser distinta, mas desenvolve atividades de vínculos industriais principalmente para adquirir recursos para apoiar a pesquisa acadêmica;

Tradicional Puro - acredita que o acadêmico e a indústria devem ser distintos e busca o sucesso estritamente na área acadêmica.

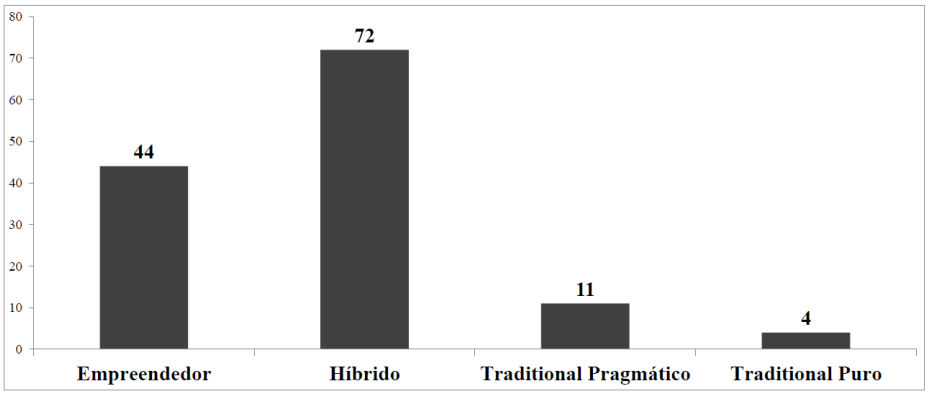


Gráfico 13 - Percepção em relação à colaboração universidade-indústria.
Fonte: Elaborados pelo autor

Tem-se que a maioria com quase 90% de respostas entende a importância da relação academia-indústria na promoção e avanço das pesquisas acadêmicas e opera no campo científico com vistas ao aproveitamento comercial das invenções. Desta parcela 53% apontam que seus pedidos de patentes resultaram de relação com a indústria, mas 60% não têm patentes licenciadas.

Da parcela que apresenta uma visão mais conservadora e reticente na relação com a indústria, 80% dos respondentes indicaram que seus pedidos de patentes não resultaram de parcerias com a indústria e 60% não licenciaram suas patentes.

Alguns depoimentos de respondentes da pesquisa, colhidos em campo de livre preenchimento no questionário, principalmente sobre a relação ciência-indústria, são trazidos a fim de compartilhar percepções adicionais da amostra e corroborar com os achados. O tema relação ciência-indústria foi objeto da maior parte de comentários livres realizados por pesquisadores, ajudando na compreensão de como percebem e reagem nesta interação e com relação a outros temas por eles trazidos.

Um pesquisador que se autoidentifica com o perfil denominado ‘tradicional puro’ e sem patentes concedidas, se demonstra preocupado com eventual conflito de interesse e seu impacto na credibilidade da pesquisa:

> *É minha preocupação que os resultados das pesquisas sejam questionados por conflito de interesse, caso tenham investimento do setor privado. Prefiro manter distância e garantir minha reputação e a credibilidade dos resultados.*

Uma pesquisadora do mesmo perfil anterior compartilha suas impressões:

> *[...] acho muito importante essa parceria, no entanto, é difícil se conseguir de fato uma colaboração. As indústrias que nos procuram em geral querem consultoria gratuita, não querem dar nada [...]*

Outro respondente que conta com patentes concedidas, licenciadas e depositadas no exterior, mas nenhuma derivada da interação com a indústria, ressalta:

> *a indústria e a academia falam diferentes idiomas e atuam em diferentes tempos, seria importante fazer com que possam convergir para um fim comum. A burocracia na Universidade para a interação com a indústria é praticamente intransponível.*

Comentando sobre a demanda da indústria em relação à academia, uma pesquisadora autoidentificada como empreendedora diz:

> *sou a favor da pesquisa aplicada. não concebo a possibilidade de um bom resultado não ser patenteado. Por outro lado o diálogo com a indústria não é fácil... eles querem o produto acabado dentro do frasco, rotulado e pronto para vender... e não é assim na pesquisa acadêmica. Nós vimemos de financiamento e a pesquisa é resultado de teses e dissertações...*

Em linha com a alegação anterior, um respondente, entende que:

> *Essa interação deve ser estimulada, deixando claro que não é papel da universidade desenvolver produtos, mas garantir o subsídio científico e*

> *tecnológico para que o desenvolvimento do produto possa ser executado pela indústria.*

Ampliando a discussão para a atuação do Governo e o efeito das regulações, um pesquisador que se reconhece como um empreendedor e tem patente concedida e licenciada, declara:

> *Algumas empresas brasileiras e em geral, as estrangeiras relutam muito em adquirir os licenciamentos (e consequentemente pagar os royalties) porque o próprio governo brasileiro promove licitações de fabricação de equipamentos e itens que incentivam essas empresas/ indústrias a não realizarem os acordos de licenciamentos e defendem essa posição, visando minimizar os custos de comercialização. Dessa forma, as empresas que adquirem a licitação de forma não-exclusiva levam uma desvantagem quando competem nessas licitações.*

Abordando os efeitos da burocratização outro respondente salienta a *"carga adicional de burocracia"* destacando que as universidades quando aumentam *"a formalização, criando órgãos específicos para cuidar do assunto (NITs, DITS, etc.), num primeiro momento, ou seja, na fase inicial, ficou muito mais complicado pensar em fazer uma patente"*

Comentando os efeitos da burocratização um pesquisador cita que

> *...no Brasil a grande concentração da pesquisa se encontra em instituições públicas, que apresenta um enorme entrave burocrático para estabelecimento de parcerias com o setor produtivo. Estamos precisando de mudanças urgentes para esta situação, sob pena de perdermos o bonde do desenvolvimento e da inovação.*

Uma respondente que se identifica com o perfil empreendedor e que conta com patente concedida, inclusive com depósito no exterior ou via Tratado de Cooperação de Patentes (PCT), comenta a diferença de prazos na concessão das patentes do cenário brasileiro comparado ao internacional:

> *A patente foi concedida pelo INPI após 11 anos de análise. A mesma patente foi concedida no exterior (Europa) após 6 anos. O tempo de análise no Brasil dificulta o usufruto de possíveis royalties.*

Outra pesquisadora, também com patente concedida e depositada no exterior, auto identificada como empreendedora que entende e busca a relação, inclusive comercial com a indústria, compartilha sua experiência:

> *...infelizmente nunca consegui licenciar uma das minhas patentes. Todo trabalho foi realizado com pequenos financiamentos do CNPq e da (nome da instituição suprimido) provenientes de projetos que visavam financiar a pesquisa nas universidades.*

Um respondente compartilha sua experiência em relação à instituição a que está vinculado e como percebe a motivação da equipe que integra, movida pelo atendimento de desafios de natureza intrínseca:

> *No meu caso os pedidos de patente visavam mais proteger a Universidade (titular das patentes), deixando documentado que a ideia nasceu ali, na Universidade, como resultado do trabalho de uma equipe tentando achar uma solução para um problema.*

Um pesquisador, se demonstrando sensível aos resultados do patenteamento para a sociedade, comenta que:

> *...patente é aquilo que movimenta e aumenta a riqueza econômica, social e intelectual do país."*

Quanto ao ***perfil sociodemográfico***, em suma, o patenteador brasileiro se revela maduro e com longa e profícua carreira acadêmica. Apesar de majoritariamente masculino, a participação de mulheres tem gradualmente se equiparado. Sua formação é desenvolvida predominante em território nacional, com tendência a internacionalização nos níveis mais altos de especialização. Relevante parcela é aberta a movimentações geográficas que ocorrem principalmente na região Sudeste do Brasil e para o exterior visando melhores recursos para formação e ascensão na carreira. Pesquisa majoritariamente nas áreas biológicas, saúde, engenharias e tecnológicas, tendo uma produção científica condizente com sua qualificação, sem concorrer com os números crescentes de envolvimento no processo patentário.

IV.3 Posição quanto aos fatores motivacionais

A exposição dos dados aqui considerados se referem aos fatores motivacionais a fim de indicar a distribuição de importância por fatores e agrupamentos de fatores que foram avaliados pelos respondentes a partir de uma escala crescente de importância/ influência no processo de patenteamento, variando de 1 a 5 na escala *Likert*, em resposta às questões.

A exposição e análise dos dados motivacionais da amostra restante (n=319) é objeto do capítulo V onde o algoritmo é aplicado a fim de inferir seu comportamento tomando por referência os dados da amostra de respondentes (n=131).

A seguir é apresentada na tabela 5 a relação de fatores motivacionais pela ordem decrescente de importância atribuída pelas maiores médias de respostas informadas pelos pesquisadores respondentes. Na sequência, nos gráficos 14 a 17 ilustram a

distribuição de importância, agrupando os fatores que foram avaliados pelos respondentes. A definição dos termos empregados foram apresentados e discutidos no Capítulo II.

A tabela 5 traz a relação de fatores motivacionais por ordem decrescente de importância atribuída. Salienta-se que a escala tipo *Likert* serve para representar ordem, sem pontos intermediários. Portanto, a moda das respostas apuradas a partir da escala *Likert* foi obtida, mas, também foram geradas as médias das atribuições informadas pelos respondentes. O emprego da média aqui não visa estabelecer comparação entre os fatores, mas melhor representar a ordem de importância.

Nº.	Fatores motivacionais	Natureza da motivação	Resultados esperados	Moda	Média
10	Contribuição dos resultados da pesquisa para o desenvolvimento social e econômico da sociedade	Intrínseca	Desafio	5	4,79
13	Atração por desafios, resolução de problemas de pesquisa e curiosidade intelectual	Intrínseca	Desafio	5	4,69
23	Potencial para a aplicação de resultados de pesquisa	Intrínseca	Desafio	5	4,68
27	Satisfação pessoal e realização profissional na realização da pesquisa	Intrínseca	Desafio	5	4,67
21	Obtenção de recursos e recursos para pesquisa	Extrínseca	Econômico	5	4,60
25	Qualificação e competência pessoal para o desenvolvimento da pesquisa	Extrínseca	Econômico	5	4,55
18	Instalações e condições de infraestrutura para o desenvolvimento de pesquisas disponíveis na instituição	Intrínseca	Desafio	5	4,54
2	Apoio administrativo institucional no processo de patenteamento, especialmente dos Escritórios de Transferência de Tecnologia (TTOs) da Universidade	Mista	Reputação	5	4,41
1	Acesso a meios e insumos, incluindo tecnológicos, para o desenvolvimento da pesquisa	Extrínseca	Econômico	5	4,39
15	Independência e autonomia no desenvolvimento da pesquisa	Intrínseca	Desafio	5	4,36
24	Proteção e apropriação intelectual de resultados de pesquisa	Intrínseca	Desafio	5	4,33
12	Experiência profissional prévia como acadêmico e pesquisador	Intrínseca	Desafio	5	4,23
26	Relações de mercado e cooperação com a indústria, incluindo a atração de recursos financeiros	Mista	Reputação	5	4,23
3	Cumprimento de indicadores de produção e possibilidades de publicação de pesquisas cujos resultados são patenteáveis	Mista	Reputação	5	4,21
19	Licenciamento e Comercialização de resultados de pesquisa	Extrínseca	Econômico	5	4,19
17	Influência dos membros da pesquisa a partir da troca de experiências	Intrínseca	Desafio	4	4,18

11	Disponibilidade de materiais, dados e recursos da indústria para o desenvolvimento de atividades de pesquisa	Mista	Reputação	5	4,11
7	Posição profissional e nível de responsabilidade que estimule o desenvolvimento da pesquisa	Extrínseca	Econômico	5	4,05
20	Momento profissional da carreira acadêmica ou pesquisadora	Extrínseca	Econômico	4	3,88
22	Influência de viagens e/ou experiências em outros lugares/culturas para o desenvolvimento de pesquisas	Mista	Reputação	5	3,66
5	Maior prestígio e visibilidade acadêmica com reconhecimento e reputação por pares	Mista	Reputação	4	3,58
9	Contribuição de promoção e/ou desenvolvimento de carreira com impacto na remuneração	Extrínseca	Econômico	4	3,40
16	Influência de colaboradores e parceiros estrangeiros	Mista	Reputação	4	3,39
8	Concorrência e colaboração entre pares no desenvolvimento de redes pessoais e profissionais	Mista	Reputação	3	3,30
4	Aumento da renda pessoal e/ou benefícios econômicos resultantes da atividade de pesquisa	Extrínseca	Econômico	3	3,15
14	Capacidade de financiar e conduzir negócios	Extrínseca	Econômico	3	3,04
6	Baixa mobilidade: apego e permanência a um local fixo e específico para o desenvolvimento da pesquisa.	Mista	Reputação	3	2,79

Tabela 5 - Fatores motivacionais por ordem de importância.

Fonte: Elaborado pelo autor

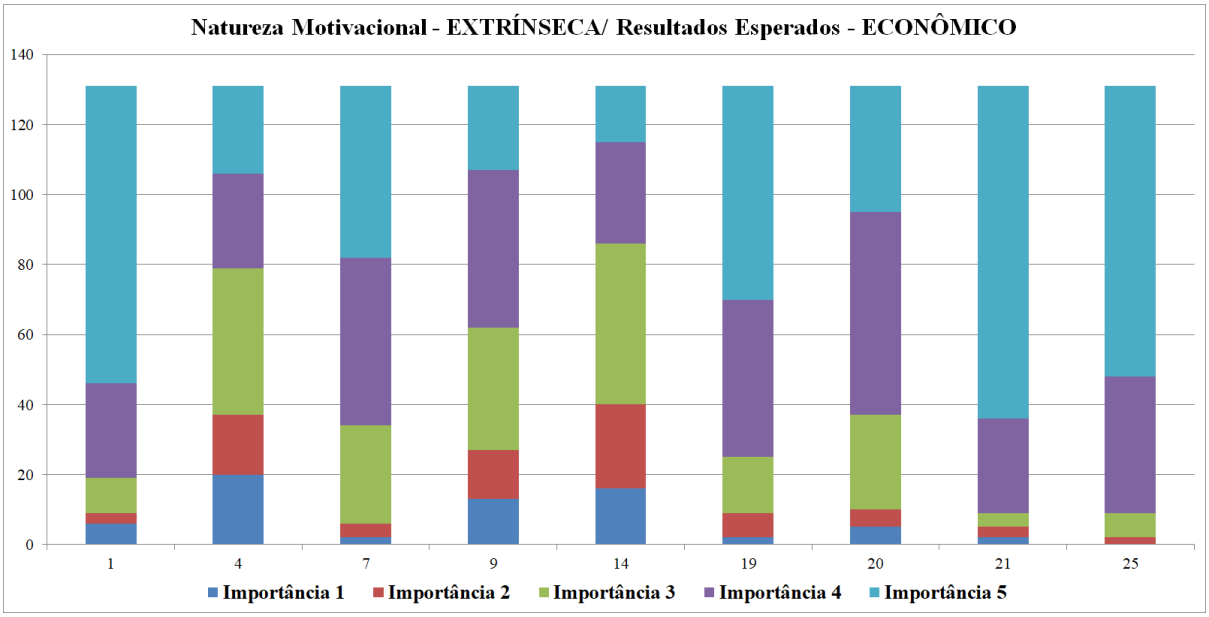


Gráfico 14 - Distribuição de importância do grupo fator Extrínseco/Econômico
Fonte: Elaborados pelo autor

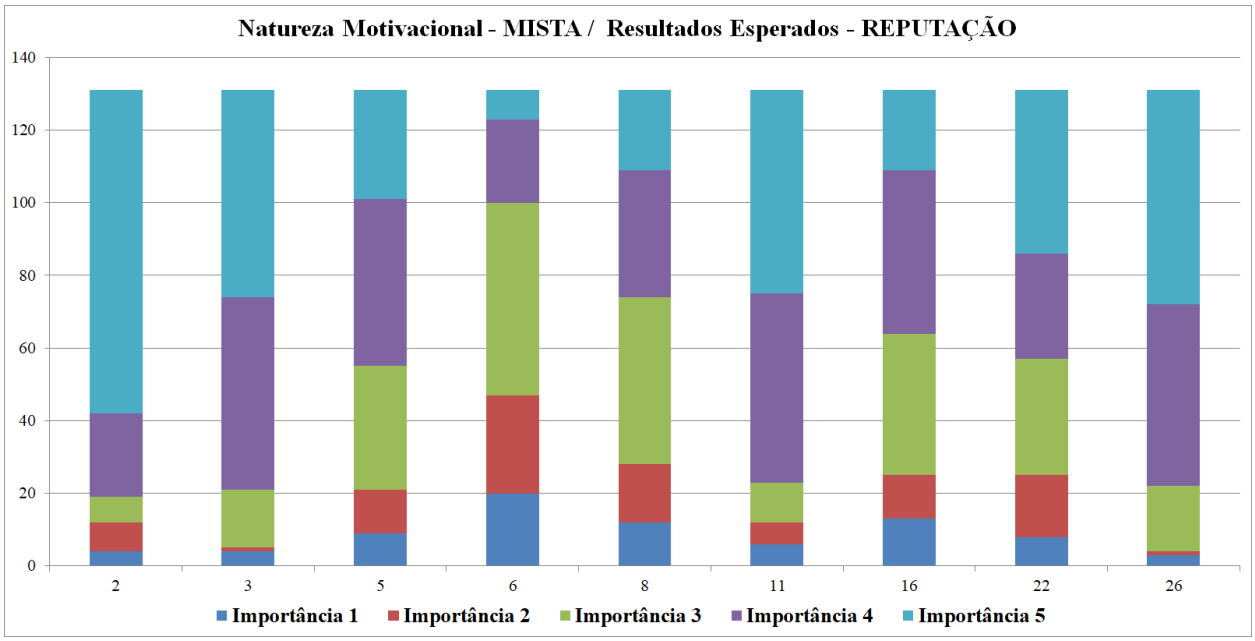


Gráfico 15 - Distribuição de importância do grupo fator Misto/Reputação
Fonte: Elaborados pelo autor

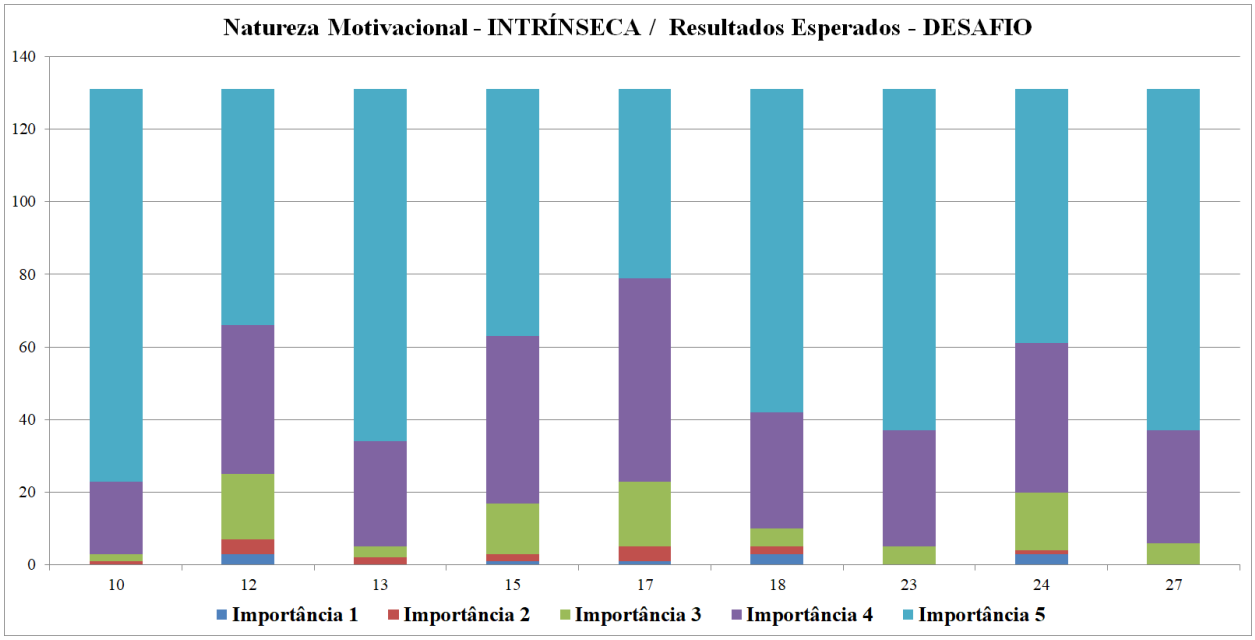


Gráfico 16 - Distribuição de importância do grupo fator Intrínseco/Desafio
Fonte: Elaborados pelo autor

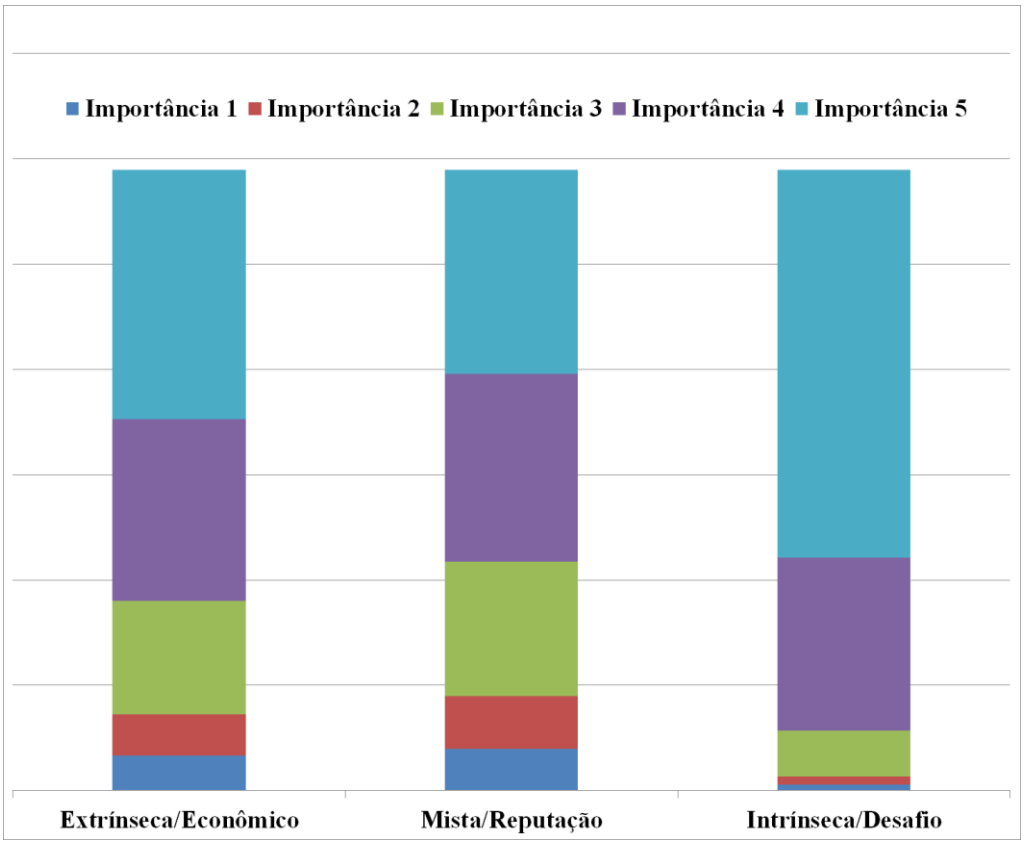


Gráfico 17 - Distribuição de importância comparativo dos grupos de fatores
Fonte: Elaborados pelo autor

Considerando a atribuição por ordem de importância dos fatores e seu agrupamento, se verifica que a série ligada à natureza motivacional intrínseca e ao resultado esperado desafio foi a que recebeu os maiores níveis de atribuição de importância. Na sequência a série extrínseca/econômico tem superação à série mista/reputação. É possível afirmar que a amostra de patenteadores tem um perfil que se realiza nas atividades inerentes ao processo de fazer pesquisa, nele se envolvendo intelectualmente a fim de prover soluções para desafios científicos que resultam em inovações patenteáveis.

Nesta direção colabora a análise individual dos fatores em que quatro com maior atribuição de importância/influência pela amostra são, pela ordem decrescente, também pertencem à série intrínseca/desafio:

10 - Contribuição dos resultados da pesquisa para o desenvolvimento social e econômico da sociedade

13 - Atração por desafios, resolução de problemas de pesquisa e curiosidade intelectual

23 - Potencial para a aplicação de resultados de pesquisa
27 - Satisfação pessoal e realização profissional na realização da pesquisa

Ainda na análise individual se tem que os fatores que tiveram menor atribuição de influência estão relacionados às séries extrínseca/econômico e mista/reputação. Os quatro fatores seguintes estão organizados da maior para a menor importância atribuída:

08 - Concorrência e colaboração entre pares no desenvolvimento de redes pessoais e profissionais
04 - Aumento da renda pessoal e/ou benefícios econômicos resultantes da atividade de pesquisa
14 - Capacidade de financiar e conduzir negócios
06 - Baixa mobilidade: apego e permanência a um local fixo e específico para o desenvolvimento da pesquisa.

Diferente da relação anterior, esta última tem composição mais heterogênea, trazendo temas como colaboração entre pares, resultados e habilidades financeiros e, por fim, o de menor importância, o fator mobilidade, indicando uma propensão da amostra em admitir mudanças em função da evolução acadêmica e profissional.

Esses resultados podem refletir o atual estágio de maturidade dos NITs das universidades brasileiras. Como mencionado anteriormente, o processo de licenciamento e recebimento de *royalties* pelos pesquisadores ainda é incipiente, o que pode desestimular os inventores em potencial que são motivados por fatores financeiros (JENSEN e THURSBY, 2001). De acordo com Baldini (2001), casos de sucesso são incentivos importantes para promover o interesse em patenteamento e licenciamento nas universidades.

A legislação brasileira permite que os pesquisadores de universidades públicas recebam até 1/3 (um terço) dos *royalties* do licenciamento de patentes (Lei 10.973, de 2 de dezembro de 2004), mas essa prática ainda não é muito difundida nas instituições nacionais. Poucas instituições assinaram contratos de transferência de tecnologia (PARANHOS *et al*., 2018). Apenas 5% das patentes ativas foram licenciadas na USP, que é uma universidade de referência no Brasil e possui um dos mais antigos e consolidados escritórios de transferência de tecnologia (BARBOSA, 2019). Deve-se

acrescentar que muitas universidades ainda não têm NITs ou regulamentos internos bem definidos (PARANHOS *et al*., 2018).

Outro aspecto a ser considerado é que o valor do licenciamento de uma patente está associado à expectativa do potencial retorno financeiro da tecnologia protegida, que é um processo complexo (TUKOFF-GUIMARÃES *et al*., 2021). Há uma deficiência de pessoas qualificadas nas universidades para tratar dessas questões (TORKOMIAN, 2009, QUERIDO *et al*., 2011; PARANHOS *et al*., 2018). Portanto, é possível que as universidades estejam protegendo e/ou comercializando tecnologias sem avaliar adequadamente seu potencial comercial.

Ainda existe um preconceito no meio acadêmico com a relação universidade-indústria, especialmente nas universidades públicas, que são as que mais registram patentes no Brasil (Closs *et al*., 2013). Entretanto, essa realidade vem mudando nos últimos anos (ANDRADE *et al*., 2022). Conforme declarado por um participante da Conferência *SciBiz*, o maior evento de inovação e empreendedorismo do Brasil:

> "Os pesquisadores precisam pensar na inovação como um possível caminho natural para suas pesquisas (...) É necessário treinar os pesquisadores para trabalhar com inovação, para que eles saibam como identificar oportunidades e o que fazer com elas, quando aparecerem. Por exemplo, como lidar com questões relacionadas à propriedade intelectual (patentes), relacionamento com empresas, desenvolvimento tecnológico, etc., coisas que não fazem parte do processo tradicional de pesquisa. - coisas que não fazem parte do portfólio tradicional de habilidades de um pesquisador acadêmico" (BARBOSA, 2019).

Considerando a ordem importância atribuída aos ***fatores motivacionais*** que norteiam os patenteadores brasileiros, têm-se a razão do protagonismo universitário observado no número de pedidos de patentes. Expressiva parcela dos pesquisadores tem viés empreendedor ou híbrido, o que indica a valorização e perspectiva de interação da universidade com a indústria, admitindo e visando o aproveitamento comercial no avanço científico e social, principalmente no que transborda como benefícios para a sociedade e, no limite, de retorno financeiro para o inventor de patentes, embora, ainda não se declare tal percepção como um alvo primário, A proeminência dos fatores

intrínsecos e de superação de desafios dá conta da satisfação pessoal indicada pelos pesquisadores na atividade de pesquisa e patenteamento de suas invenções, seguida, ainda que em menor importância atribuída, pelas motivações de natureza econômica e que aufira ganhos de prestígio para si, equipes e para as instituições a que estão vinculados.

CAPÍTULO V – DESENVOLVIMENTO E APLICAÇÃO DO MODELO DE CLASSIFICAÇÃO DO PERFIL MOTIVACIONAL

Considerando o volume de dados consolidado no *dataset* quanto aos atributos da amostra de pesquisa – fatores motivacionais e principais características sociodemográficas dos pesquisadores, o presente capítulo avança para a fase de processamento e análise dos dados com a aplicação de ferramentas computacionais, visando alcançar o terceiro objetivo de pesquisa.

São desenvolvidas a metodologia e a estratégia de elaboração de *scripts* de processamento, tratamento e análise*,* culminando na mineração de dados quando é desenvolvido e aplicado um modelo classificador via algoritmo de árvore de decisão na linguagem *Python* à luz do *framework* considerado no capítulo II.

Na evolução deste capítulo são considerados a metodologia e os conceitos próprios de cada seção e/ou etapa afeita à ciência da informação em que se fundamentam as ferramentas e técnicas computacionais aplicadas. Ao fim são apresentadas e comentadas as classificações obtidas a partir do modelo desenvolvido.

V.1 Considerações sobre a amostra e organização dos dados

As atividades de patenteamento, principalmente em universidades e entidades de pesquisa equivalentes, resultam de um conjunto de fatores individuais, institucionais e contextuais (TIJSSEN, 2018). O foco da presente pesquisa tem sua fonte de dados em uma amostra de indivíduos pesquisadores, reconhecidos como atores de grande impacto desse processo, notadamente por conta de sua capacidade intelectual e inventiva. Olaya-Escobar *et al*. (2020) destacam que o perfil dos pesquisadores tem maior impacto no processo de patenteamento do que a estrutura disponível nos escritórios de transferência de tecnologia, devendo, portanto, ser mais bem analisado sob uma perspectiva multidimensional, como a proposta no *framework* conceitual considerado no capítulo II.

Para tanto, a consolidação e tratamento de dados sobre a atuação dos pesquisadores consiste em um relevante recurso na análise e aferição de perfis. A amostra tomada para esta pesquisa reúne atributos sociodemográficos, acadêmicos, profissionais e outros afeitos às áreas de pesquisa e produção científica, além de características

relacionadas a aspectos e fatores comportamentais e motivacionais da atividade de patenteamento.

O conjunto de dados originais coletado na presente pesquisa foi organizado em um *dataset* em formato de planilha eletrônica, cuja estrutura contempla um total de 48 atributos (colunas) e 450 pesquisadores (linhas), além de cabeçalhos de dados (uma linha adicional cujo conteúdo é a descrição de cada atributo). A planilha apresenta os dados brutos obtidos dos pesquisadores que compõem a amostra. Distinção há entre os que responderam ao questionário cujas informações foram colhidas integralmente, e os não respondentes que contam com informações colhidas da plataforma *Lattes*. São dados qualitativos e quantitativos primários referentes à amostra total de 450 pesquisadores. O *dataset* está disponível em repositório indicado no Apêndice C.

Para efeito de tratamento de pesquisa foi adotada a divisão dos dados em **públicos** e **institucionais** em função de sua caraterística e disponibilidade, sempre tomando como referência o *dataset*.

I. *Dados públicos* – disponibilizados na plataforma *Lattes* e acessível ao público em geral sem a necessidade de protocolos ou autorização (do atributo "*Research áreas*" a "*Doctorate's degree location*");
II. *Dados institucionais*– não disponibilizados ao público em geral e obtidos por meio das respostas ao questionário *on-line* (do atributo *"Birth Interval"* ao *"Classification regarding professional orientation");*
    a. *Dados funcionais* – acessíveis a gestores e equivalentes, pois disponíveis nas bases de dados das entidades a que os pesquisadores estão filiados (do atributo *"Birth Interval"* ao *"Have any patent request been the result of interaction with the industry?");*
    b. *Dados da escala Likert* – impressões dos pesquisadores sobre os fatores referentes ao processo de patenteamento (do atributo *"Access to means and inputs, including technological, for the research development"* ao *"Personal satisfaction and professional achievement in conducting the research");*
    c. *Classificação secundária* – baseada em respostas anteriores do conjunto de dados (do atributo "*Interaction in patenting process. Active or passive?"* e "*Patenting Frequency. Serial or Occasional?");*

d. *Autoclassificação* – Pesquisador indica sua posição baseado em definições disponibilizadas (atributo "*Classification regarding professional orientation*").

V.2 O processo de descoberta de conhecimento em base de dados

Colecionar, organizar e tratar um volume de dados não é uma tarefa trivial. O ganho de performance se dá na aplicação de ferramentas computacionais. Esta solução emerge em resposta à crescente demanda do cotidiano, permeado por informações que regem as tomadas de decisões de entidades e indivíduos.

Considerando o tratamento de um volume de dados significativo se tem a área denominada Descoberta de Conhecimento em Base de Dados (DCBD) ou *Knowledge Discovery in Databases* (KDD). Segundo FAYAAD *et al.*(1996), o *KDD* é um processo de extração de conhecimento de grandes bases de dados. Os autores o descrevem como um processo não trivial, com relevante grau de complexidade, interativo e iterativo. A interatividade se dá pela participação com o elemento humano – o analista – que intervém e controla na medida em que define objetivo e, por fim, avalia o resultado. A iteratividade, por sua vez, se dá pela possibilidade de tantas repetições da aplicação do processo ou de suas fases quanto a necessidade de refinamento dos dados. O objetivo do *KDD* é identificar informações inéditas, validadas, úteis à prospecção, e passíveis de análise e interpretabilidade (FAYAAD *et al*., 1996b).

Autores como, Adriaans e Zantinge (1996), Brachman e Anand (1996), Fayaad *et al*. (1996a); Fayaad *et al*. (1996b); Cabena *et al*. (1997) e Castro e Ferrari (2016), dentre outros, abordam o tema concebendo o processo do *KDD* em fases correlacionadas, interdependentes e sobrepostas. A figura 15, apresentada em inglês a fim de manter sua originalidade, representa o fluxo das cinco fases, descritas sucintamente à luz dos citados autores. As fases:

- seleção dos dados;
- pré-processamento;
- transformação;
- mineração e
- interpretação/análise.

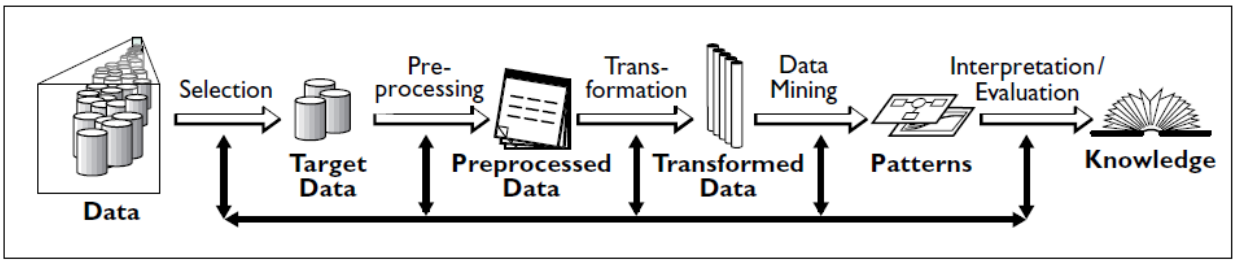


Figura 15 - Representação das fases do *Knowledge Discovery in Databases* (KDD)
Fonte: Fayyad *et al*. (1996b)

Seleção de dados – selecionar um conjunto ou um subconjunto de dados, identificando a amostra onde se pretende extrair conhecimento.

Pré-processamento – também denominada de fase de limpeza, inclui operações de remoção de elementos estranhos ou inconsistentes denominados ruídos, além da eliminação de discrepâncias e da realização de eventuais incrementos e formatações.

Transformação – também denominada de fase de projeção, compreende operações de redimensionamento dos dados visando sua melhor representação, armazenamento e comparabilidade.

Mineração – fase de determinação e desenvolvimento dos parâmetros e do próprio algoritmo de mineração de dados visando encontrar padrões presentes nos dados e/ou a entrega de modelos preditivos.

Interpretação/análise – fase de conhecer e entender a descoberta, empreendendo análises e leituras à luz do conhecimento obtido a fim de validá-lo ou retroalimentar o processo.

Em todas as fases da obtenção do conhecimento de dados é necessária a aplicação de ferramentas computacionais, desde a extração até a mineração de dados, etapa considerada crucial no processo *KDD* quando é desenvolvido um algoritmo ou uma combinação deles para se obter o resultado e proceder a análise necessária a partir do conhecimento gerado.

Frise-se, de acordo com Nimer e Spandri (1998), que o *KDD* dá conta do processo de descoberta dos dados úteis aos objetivos definidos, enquanto a mineração de dados (*data mining*) focaliza o desenvolvimento e aplicação de algoritmos para extração de modelos e padrões dos dados. No resultado do processo, como salientam Fayyad *et al*. (1996b), além dos dados que serão evidentemente objeto de análise, está a entrega

do *pattern,* expressão em inglês cuja melhor tradução se aproxima do termo ‘padrão’. Porém, no contexto do *KDD*, a interpretação da expressão como um ‘modelo’ seria mais indicada, pois sugere mais do que uma estrutura definida ou finalizada, apontando, por fim, às saídas como um resultado do processo do qual decorrem.

Neste sentido o presente trabalho, considerando os conceitos da aprendizagem de máquina (*machine learning*), se propõe ao desenvolvimento de um algoritmo na linguagem *Python* que, modelado e representado preditiva e simbolicamente em uma árvore de decisão, visa reconhecer características e classificar um conjunto de dados a fim de verificar determinados padrões que culminem em um modelo classificador pertinente ao *dataset* considerado, podendo servir como referência conceitual e metodológica para investigações similares.

Tomando como referência a representação do *KDD* na figura 15 e as fases do pré-processamento, detalhadas na próxima seção, é trazida a figura 16 cujo fluxo detalha as ênfases e adaptações pertinentes ao escopo e delimitações da pesquisa em curso.

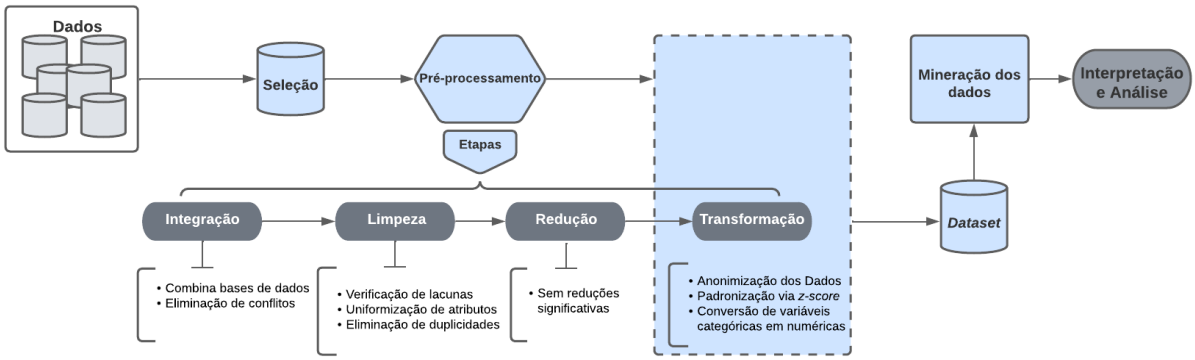


Figura 16 - Fluxo *KDD* da pesquisa
Fonte: Elaborado pelo autor.

Cumpre salientar que a fase de **seleção de dados** se deu no escopo do capítulo III onde também o pré-processamento (integração, limpeza e redução) teve grande avanço quando da realização de rotinas próprias da composição do *dataset*. No capítulo em tela, por decisão metodológica e pelo emprego de ferramentas estatísticas e computacionais mais avançadas, a fase do pré-processamento denominada **transformação** é complementada principalmente quanto às tarefas de padronização e conversão das variáveis categóricas em numéricas (considerando que a anonimização dos dados foi realizada e apresentada no capítulo anterior).

A seguir a fase de pré-processamento e seus diferentes métodos são mais bem explicitadas, assim como rotinas inerentes ao seu desenvolvimento com a apresentação dos respectivos conceitos e técnicas empregados.

V.3 Pré-processamento dos dados

É na fase de pré-processamento de dados em que se concentram os recursos de tempo e esforço material e intelectual a fim de preparar, organizar e estruturar o volume de dados. Segundo assinalam Kotsiantis *et al*. (2006) e Garcia *et al*., (2016) esta fase impacta diretamente na integridade e qualidade de dados, sendo sensível principalmente ao aprendizado de máquina e à extração de conhecimento via mineração dos dados. Dentre as principais virtudes de um conjunto de dados de qualidade estão a precisão, a completude, consistência, pontualidade, credibilidade, atualização e interpretabilidade.

O objetivo principal do pré-processamento de dados é identificar e afastar problemas comuns, como classes desbalanceadas, incompletude e outras inconsistências (GARCIA *et al*., 2015; KRAWCZYK, 2016). Como organizado, de forma mais ampla por Garcia *et al*., (2016) e mais sintetizada por Sivakumar e Gunasundari (2017), versão aqui adotada, quatro são os principais métodos ou etapas da fase de pré-processamento dos dados: limpeza, integração, transformação e redução. A ordem, principalmente quanto às fases de transformação e redução, não é obrigatória.

Os métodos apresentados são descritos na ordem em que as rotinas são cumpridas, sendo descritas de forma sucinta.

A **integração** se dá pela agregação de várias fontes ou arquivos cujos dados podem ter padrões não assemelhados ou a princípio não identificados, combinando-os e eliminando eventuais redundâncias a fim lhes conferir coerência no armazenamento.

Ações de integração no *dataset*:

- Combinação de dados oriundos da plataforma *Lattes* e do questionário *on-line*, utilizando o atributo nome dos participantes, posteriormente removido;
- Remoção e/ou correção de grafia nos rótulos e dados dos atributos.

A **limpeza** empreende algumas rotinas nos dados visando tratar ou gerar valores ausentes, corrigir inconsistências e eliminar ruídos e discrepâncias, além da uniformização dos dados.

Ações de limpeza no *dataset*:

- Verificadas eventuais faltas de dados de atributos referentes aos não respondentes do questionário de pesquisa;
- Eliminação de conflitos de dados pela uniformização de formatos quanto aos locais de nascimento, data de nascimento e outras informações sociodemográficas;
- Conferência e eliminação de duplicidades de respostas/linhas;

A **redução** se dá pelo ajuste no tamanho do volume de dados, reduzindo sua dimensão sem perder, contudo, qualidade e integridade. É possível a eliminação de atributos eventualmente irrelevantes, gerando uma representação reduzida do conjunto de dados e, consequentemente, a redução da quantidade de exemplos a que o algoritmo será exposto.

Considerando o volume da amostra e por análise metodológica não foram identificadas necessidades de ajustes significativos no *dataset* considerado.

Na **transformação** se ajusta as variáveis para um formato, classificação, intervalo e/ou escala que permita as comparações e relações necessárias em uma base comum, além do redimensionamento dos dados, empregando, por exemplo, a técnica de padronização.

Ações de transformação no *dataset*:

- Realização do processo de anonimização de dados sensíveis.
- Conversão das variáveis categóricas em numéricas.
- Aplicação da técnica de padronização de atributos via z-score;

Considerado o processo *KDD* para a geração de conhecimento sobre os dados e suas respectivas fases, dando prosseguimento à fase de transformação do pré-processamento, o passo seguinte é a categorização dos atributos a partir dos parâmetros necessários ao tratamento e mineração dos dados.

V.3.1 Parâmetros e marcadores para categorização da amostra

A partir deste ponto o trabalho recupera os parâmetros metodológicos sustentados no Capítulo II, tomando os elementos, camadas e dimensões do *framework* como

categorias (variáveis de saída) e marcadores a fim de consubstanciarem a formatação do perfil pretendido, categorizando cada indivíduo da amostra.

O denominado *framework*, representado na figura 9, consiste em uma abstração teórica sustentada e composta a partir de um arcabouço bibliográfico, reunindo e articulando conceitos, métodos e abordagens e, ao mesmo tempo, servindo como instrumento orientador da aferição do perfil motivacional de pesquisadores, cujos atributos estão armazenados e dispostos no *dataset*.

Os dados de cada pesquisador estão dispostos nas linhas do *dataset* e as colunas correspondem a atributos/ categorias. As informações dos pesquisadores que estavam categorizadas em variáveis de saídas (categorias do *framework*) não sofreram alteração. Para os não categorizados foram gerados processos de normalização a partir das informações de entrada disponíveis no *dataset.*

Doravante o trabalho busca ampliar os atributos constantes no *dataset* a fim de atender à mensuração necessária das dimensões que estruturam o *framework* e, finalmente, classificar os pesquisadores nas categorias de saída de cada camada do modelo. Para tanto, a partir do modelo são admitidos os parâmetros para as camadas: *natureza motivacional, resultados esperados, orientação profissional, frequência e parceria, relação entre normas e valores pessoais.*

Para cada dimensão/camada mensurada novas colunas são adicionadas ao *dataset*, compondo, assim, um volume de atributos tal que sirva de parâmetro para a instrução do modelo classificador.

Quanto aos dados coletados, se tem que os respondentes do questionário têm suas informações registradas no *dataset*, indicando seu posicionamento quanto a duas das cinco camadas (frequência e parceria e orientação profissional), estando, assim, já configurada sua categorização. Restam, então, três categorias para apuração e categorização: natureza motivacional, resultados esperados e relação entre normas e valores pessoais.

Para determinação da categorização referente a estas três camadas são empreendidas análises e o desenvolvimento de um *script* em *Python*, resultando no incremento do *dataset.*

V.3.2 Redimensionamento de dados para a categorização

A apresentação do *dataset* visa consolidar um volume de dados validados, aspirando sua submissão ao treinamento de um algoritmo de aprendizado de máquina na elaboração de um modelo classificador. Um desafio ao trabalhar com informações no campo da ciência de dados e aprendizado de máquina, sobretudo na fase de pré-processamento, mais precisamente na etapa de transformação visando à categorização da amostra, é lidar com o desbalanceamento das variáveis de entrada. Para tanto é realizado o redimensionamento dos dados por meio de técnicas denominadas normalização ou padronização (FAWCETT e PROVOST, 1997; WEISS, 2004).

O desbalanceamento se deve principalmente à natureza dos próprios dados, da forma ou fonte de coleta. Também há influência dos custos, restrições de privacidade, dentre outros, além, claro, da natureza diversa dos dados entre si, muito comum em informações, por exemplo, de diagnósticos, apuração de fraudes e dados do ambiente natural.

O redimensionamento mitiga os riscos inerentes aos problemas de classificação, promovendo melhores condições de treino e conferindo maior acurácia ao resultado obtido no desenvolvimento do algoritmo (KRAWCZYK, 2016). O desbalanceamento gera maior exposição do modelo de treinamento de aprendizado de máquina à classe com maior representação e, consequentemente, menor exposição da classe menos representada, afetando diretamente o resultado. Aqui o foco é preparar a estrutura e agregar qualidade aos dados, cuja dimensão é mais ajustada ao algoritmo em desenvolvimento, evitando a escassez ou desequilíbrio de exemplos de treinos.

O processo de redimensionamento não altera os dados em si, mas atua na convergência das escalas, de modo que as variáveis redimensionadas contribuam de forma mais equitativa para o resultado, evitando, por fim, o enviesamento da classificação, permitindo melhor categorização, comparação e condição de treino. Dentre as técnicas mais comuns no redimensionamento estão a padronização, ranqueamento e normalização, que organizam os dados em um intervalo de valores comuns, importando em menor ou maior acurácia a se obter no modelo pretendido (KOTSIANTIS *et al*., 2006; CASTRO e FERREIRA, 2016).

A normalização é uma técnica em que os valores são redimensionados na escala conhecida como Min-Máx. em intervalos entre 0 e 1.

Fórmula da normalização $Xc = \frac{X - Xmin}{X\ max - X\ min}$

A padronização segue a lógica da normalização, porém o redimensionamento se dá considerando a média e desvio padrão, este variando em função do valor obtido da média tomada como valor padrão igual a 0.

Segundo Castro e Ferrari (2016), dentre as medidas de posição relativa está o escore z ou escore padronizado, servindo à comparação de valores de diferentes conjuntos de dados relativa. O também denominado *z-score* toma o valor da média como padrão e mensura quantos desvios-padrão um ponto está abaixo ou acima da média, gerando a distribuição Z.

A fórmula da padronização via *z-score* é $Z = \frac{(x - \mu)}{\sigma}$ para qualquer dado da amostra, sendo que: (x) ponto da amostra; (μ) média; (σ) desvio padrão.

Dentre as ferramentas de redimensionamento a que os dados são submetidos, são empregadas para efeito de comparação, a padronização via *z-score*, a normalização via mín. - máx. e o ranqueamento visando obter medidas de comparação das amostras.

O valor Z tem resultado melhor quando comparado ao ranqueamento ou atribuição normalizada de faixa 0 a 1, gerando maior equilíbrio e balanceamento entre os atributos e minimizando a ocorrência de empates ou ambiguidades.

Considerados estes conceitos, a partir da seção seguinte foram realizados os redimensionamentos dos dados referentes às camadas que estruturam o *framework*, mas que ainda não têm o valor representado no *dataset* (natureza motivacional, resultados esperados e relação entre normas e valores pessoais). Para tanto é realizada a padronização empregando o *z-score* via *script* em linguagem *Python*.

Considerando as referências metodológicas e de escopo apresentadas no *framework*, para duas (orientação profissional e frequência e parceria) das cinco camadas já se

tem no *dataset* a categorização pretendida. Para as camadas restantes, considerando a correlação entre as camadas natureza motivacional e resultados esperados é demonstrado a seguir o processo de padronização dos dados visando à categorização da amostra.

V.3.2.1 Balanceamento da camada Natureza motivacional

A determinação da camada de natureza motivacional está baseada nas 27 questões de escala *Likert* do questionário *on-line* e seguem os parâmetros já apresentados. O endereço eletrônico do repositório em que o *dataset* está armazenado é disponibilizado no Apêndice C.

O primeiro passo é o cálculo das médias de respostas às perguntas de cada uma das categorias possíveis associadas: extrínseca, intrínseca e mista.

A seguir é apresentado um panorama do redimensionamento dos dados da categoria da natureza motivacional, apresentando inicialmente os valores originais em comparação com as versões balanceadas via *z-score,* ranqueamento e mín.-máx.

Para este fim um *script* desenvolvido em *Python* é aplicado no redimensionamento dos dados referentes à natureza motivacional. O código-fonte desta seção está disponibilizado no Apêndice D, no quadro D1.

A tabela 6 e o gráfico 17 apresentam os dados originais, sem o redimensionamento e, portanto, desbalanceados.

Destaca-se o emprego do diagrama em caixa (*box plot*) no gráfico 17, entre outros, modelo frequentemente utilizado em descrições de dados. Como analisam Castro e Ferrari (2016), este modelo agrega informações relevantes à visualização dos dados, como a exibição do centro, da dispersão e da distribuição, empregando na sua formulação os valores mínimos e máximos e os quartis da distribuição. Seu formato de caixa se deve à forma geométrica cujas linhas verticais têm os pontos extremos do valor mínimo ao máximo, com a mediana apontando o valor central da distribuição e com linhas horizontais no primeiro e terceiro quartis, delimitando a caixa.

	Conta-gem	Média	Desvio Padrão	Mín.	25% 1º quartil	50% Mediana	75% 3º quartil	Máx.	Identificados com a categoria
Extrínseca	131.00	3.92	0.55	2.11	3.56	3.89	4.22	5.00	21
Intrínseca	131.00	4.50	0.38	2.78	4.33	4.56	4.78	5.00	110
Mista	131.00	3.74	0.57	1.67	3.44	3.78	4.11	5.00	0

Tabela 6 - Dados originais da natureza motivacional sem a padronização

Fonte: Elaborado pelo autor

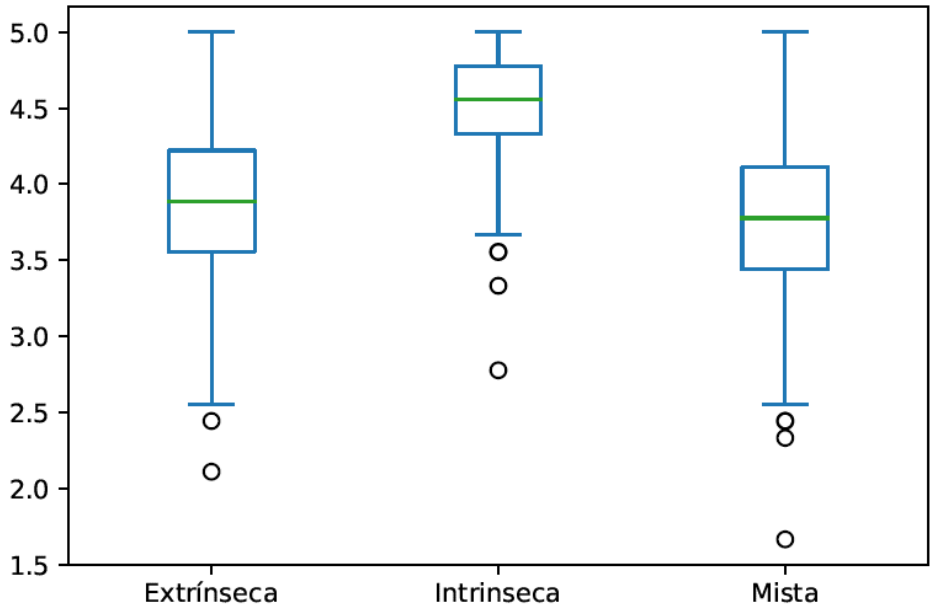


Gráfico 17 - Distribuição dos dados originais da natureza motivacional sem padronização

Fonte: Elaborado pelo autor

Os valores dos atributos não redimensionados são disponibilizados atualizados no *dataset nos* atributos *"Extrinsic", "Intrinsic"* e *"Mixed"*. Os pesquisadores identificados nos perfis da categoria apresentam esperado desbalanceamento, como exibido na coluna final da tabela 6. A tabela 7 e o gráfico 18 apresentam os dados balanceados pelo *z-score.*

	Conta-gem	Média	Desvio Padrão	Mín.	25% 1º quartil	50% Mediana	75% 3º quartil	Máx.	Identificados com a categoria
z-Extrínseca	131.00	0.00	1.00	-3.32	-0.67	-0.05	0.56	1.99	44
z-Intrínseca	131.00	-0.00	1.00	-4.56	-0.43	0.16	0.75	1.34	43
z-Mista	131.00	0.00	1.00	-3.64	-0.52	0.06	0.65	2.21	44

Tabela 7 - Dados da natureza motivacional balanceados via *z-score*.

Fonte: Elaborado pelo autor

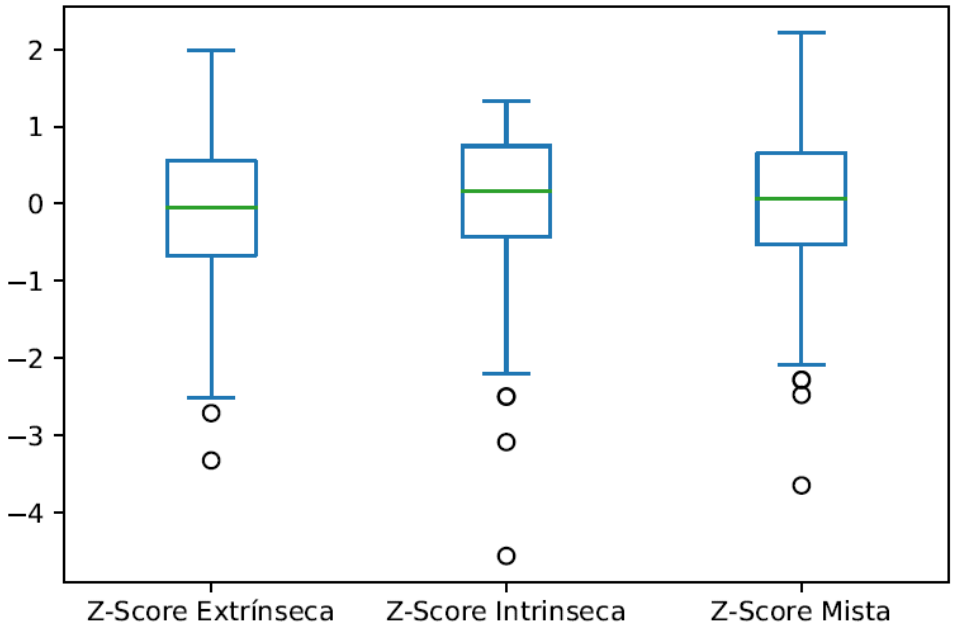


Gráfico 18 - Distribuição dos dados balanceados da natureza motivacional via *z-score*.

Fonte: Elaborado pelo autor

A tabela 8 e o gráfico 19 apresentam os dados balanceados via ranqueamento quanto à natureza motivacional.

	Contagem	**Média**	**Desvio**	**Contagem**	**Média**	**Desvio**	**Contagem**	**Média**	**Identificados com a categoria**
Ranq. Extrínseca	131.00	66.00	37.82	1.00	29.00	59.50	94.00	131.00	46
Ranq. Intrínseca	131.00	66.00	37.75	1.00	38.00	64.50	99.50	126.50	44
Ranq. Mista	131.00	66.00	37.87	1.00	38.00	67.50	96.50	131.00	41

Tabela 8 - Dados ranqueados da natureza motivacional

Fonte: Elaborado pelo autor

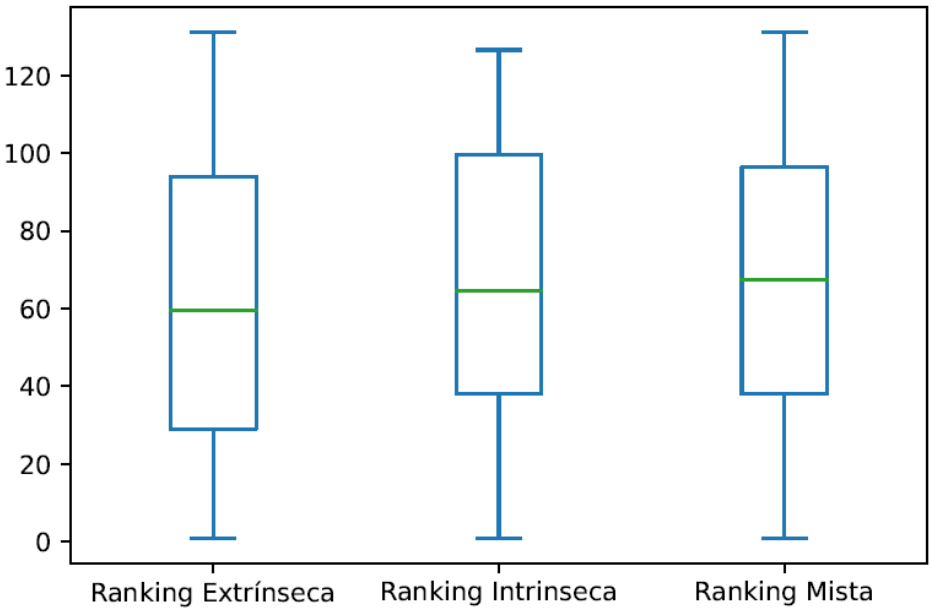


Gráfico 19 - Distribuição dos dados ranqueados quanto à natureza motivacional

Fonte: Elaborado pelo autor

A tabela 9 e o gráfico 20 apresentam os dados balanceados via mínimo-máximo quanto à natureza motivacional.

	Contagem	Média	Desvio	Contagem	Média	Desvio	Conta-gem	Média	Identifica-dos com a categoria
mm Extrínseca	131.00	0.63	0.19	0.00	0.50	0.62	0.73	1.00	22
mm Intrínseca	131.00	0.77	0.17	0.00	0.70	0.80	0.90	1.00	95
mm Mista	131.00	0.62	0.17	0.00	0.53	0.63	0.73	1.00	14

Tabela 9 - Dados balanceados via mínimo-máximo da natureza motivacional

Fonte: Elaborado pelo autor

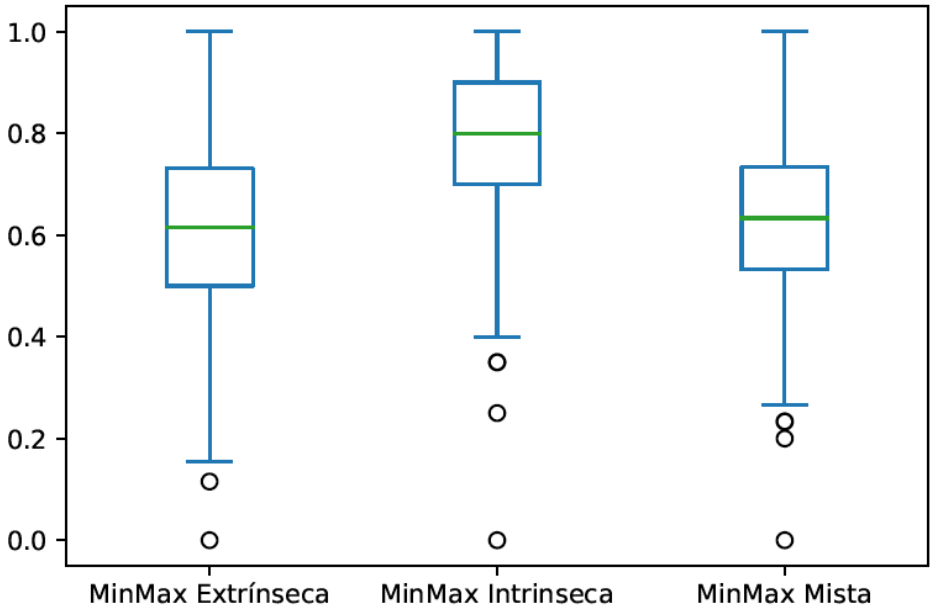


Gráfico 20 - Dados da natureza motivacional balanceados via mínimo-máximo.
Fonte: Elaborado pelo autor

Os dados balanceados estão disponibilizados no *dataset* nos atributos "*z Extrínsic"*, "*z Intrínseca"* e "*z Mixed"* no *dataset*, alcançando a pretendida categorização da amostra quanto à camada correspondente do *framework*. No *dataset* o atributo "*Nature of motivation*" apresenta a indicação do perfil de cada linha referente ao respectivo pesquisador, considerando a padronização e, portanto, o balanceamento dos dados.

A partir dos resultados expostos, primeiramente os dados originais, não balanceados, se observa uma predominância dos respondentes quanto à identificação da natureza intrínseca de motivação. O fator de desbalanceamento parece se demonstrar um viés natural da amostra, pois é composta por acadêmicos cuja motivação tem majoritariamente natureza intrínseca, sendo orientada pela superação de desafios, sobretudo os intelectuais, presentes nas atividades de ensino. Tal viés é eliminado quando da normalização dos dados, permitindo a diferenciação de parte da amostra de patenteadores, ainda que minoritários, mas afeitos a outra natureza motivacional quando comparados a seus pares.

O balanceamento pelo *z-score*, enfim, comparado com as demais técnicas, apresentou melhor resultado na eliminação de empates e conferiu melhor balanceamento aos dados, agregando, por fim, maior qualidade aos atributos e permitindo melhor emprego do aprendizado de máquina na concepção do modelo classificador.

Cumpre salientar que o balanceamento de dados fornece elementos calibradores do modelo classificador, importando menos na atribuição de um perfil exclusivo ao pesquisador e mais em uma distribuição em ordem de preponderância útil ao treinamento do algoritmo.

V.3.2.2 Balanceamento da camada relação entre normas/ valores pessoais

A determinação da camada referente à relação entre normas e valores pessoais está baseada nas proposições de Lam (2011) e já apresentada na seção do questionário on-line e seguem os parâmetros já apresentados na seção IV.3.1 Os dados desta categoria são submetidos, então, ao processo do balanceamento.

Os valores da média original quanto à introjeção, identificação e integração têm seus valores expostos nos atributos "*introjection*", "*identification*" e "*integration*" no *dataset*.

O código-fonte desenvolvido nesta seção está disponibilizado no Apêndice D, no quadro D2. A tabela 10 e o gráfico 21 apresentam os dados originais da camada não balanceados.

	Contagem	Média	Desvio Padrão	Mín.	25% 1º quartil	50% Mediana	75% 3º quartil	Máx.	Identificados com a categoria
Introjeção	131.00	0.33	0.27	0.00	0.00	0.33	0.67	1.00	34
Identificação	131.00	0.63	0.21	0.33	0.33	0.67	0.67	1.00	79
Integração	131.00	0.37	0.28	0.00	0.17	0.33	0.67	1.00	18

Tabela 10 - Dados da relação entre normas e valores, sem o balanceamento.
Fonte: Elaborado pelo autor

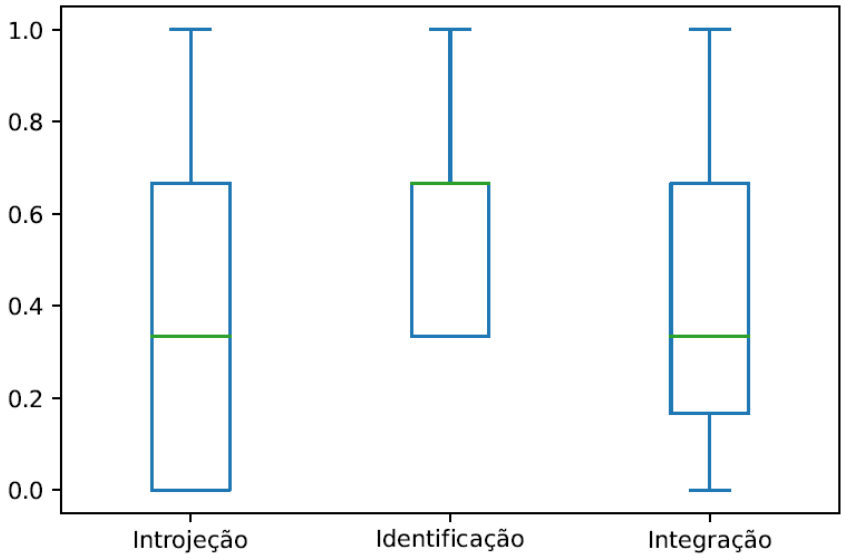


Gráfico 21 - Dados da relação entre normas e valores, sem o balanceamento.

Fonte: Elaborado pelo autor

A tabela 11 e o gráfico 22 apresentam os dados balanceados pelo *z-score*.

	Conta-gem	Mé-dia	Desvio Padrão	Mín.	25% 1º quartil	50% Mediana	75% 3º quartil	Máx.	Identifica-dos com a categoria
z-Introjeção	131.00	0.00	1.00	-1.25	-1.25	0.00	1.25	2.50	34
z-Identificação	131.00	0.00	1.00	-1.42	-1.42	0.18	0.18	1.78	55
z-Integração	131.00	0.00	1.00	-1.34	-0.74	-0.14	1.06	2.26	42

Tabela 11 - Dados da relação entre normas e valores balanceados via *z-score.*

Fonte: Elaborado pelo autor

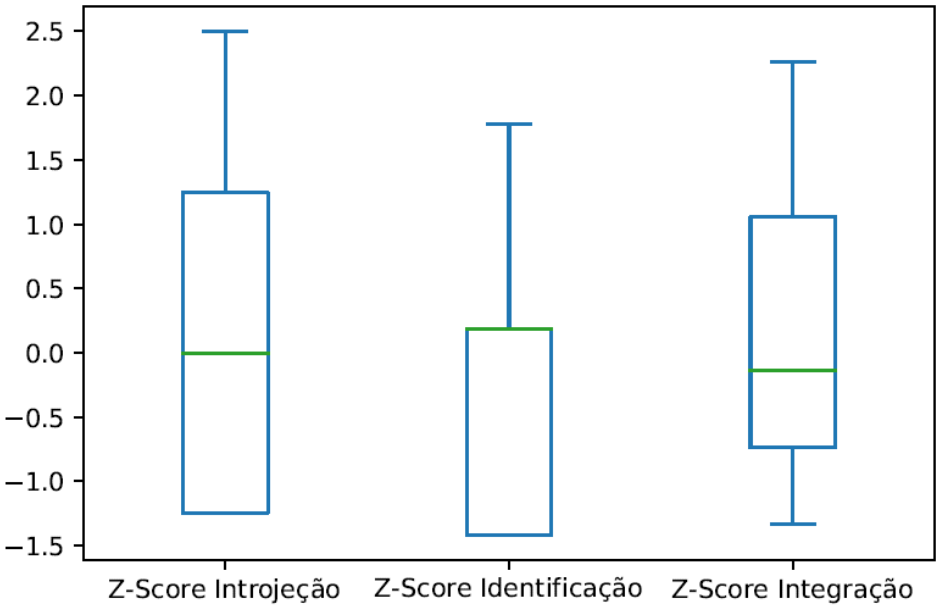


Gráfico 22 - Distribuição dos dados da relação entre normas e valores via *z-score*.

Fonte: Elaborado pelo autor

A tabela 12 e o gráfico 23 apresentam os dados ranqueados quanto à relação entre normas e valores

	Contagem	Média	Desvio	Contagem	Média	Desvio	Contagem	Média	Identificados com a categoria
Ranq. Introjeção	131.00	66.00	35.45	19.50	19.50	68.00	112.50	129.5	34
Ranq. Identificação	131.00	66.00	33.28	17.50	17.50	73.50	73.50	122.0	55
Ranq. Integração	131.00	66.00	35.71	17.00	39.25	61.50	107.50	128.5	42

Tabela 12 - Dados ranqueados da relação entre normas e valores

Fonte: Elaborado pelo autor

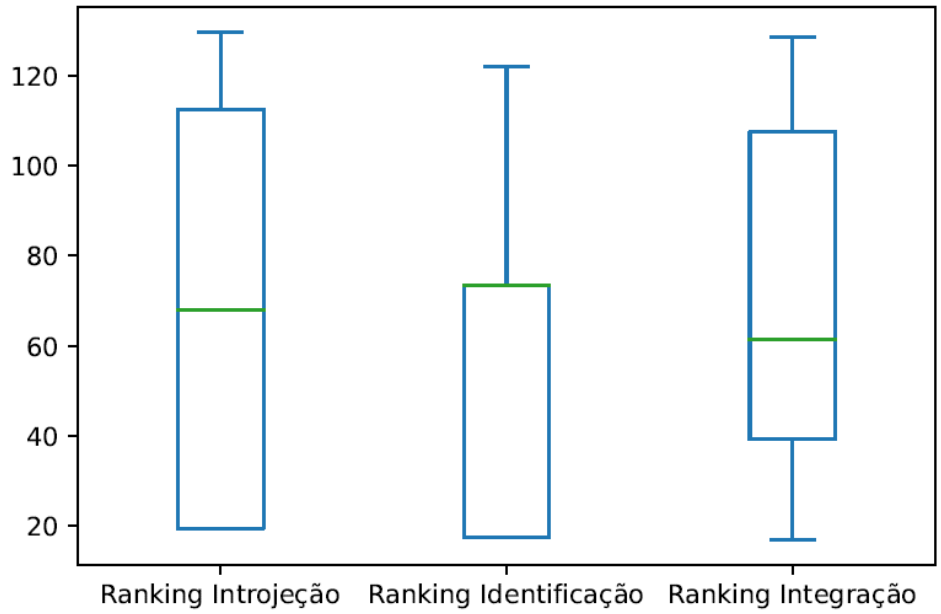


Gráfico 23 - Distribuição dos dados ranqueados quanto à relação entre normas e valores

Fonte: Elaborado pelo autor

A tabela 13 e o gráfico 24 apresentam os dados *mm* quanto à relação entre normas e valores

	Contagem	Média	Desvio	Contagem	Média	Desvio	Contagem	Média	Identifica-dos com a categoria
mm Introjeção	131.00	0.33	0.27	0.00	0.00	0.33	0.67	1.00	34
mm Identificação	131.00	0.44	0.31	0.00	0.00	0.50	0.50	1.00	55
mm Integração	131.00	0.37	0.28	0.00	0.17	0.33	0.67	1.00	42

Tabela 13 - Dados mm da relação entre normas e valores

Fonte: Elaborado pelo autor

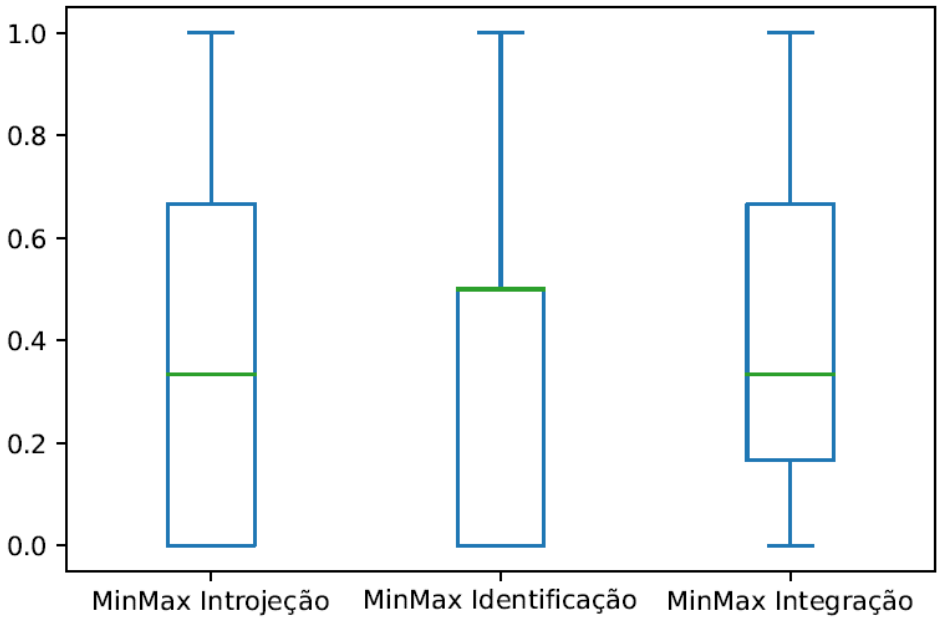


Gráfico 24 - Distribuição dos dados mm quanto à relação entre normas e valores
Fonte: Elaborado pelo autor

Com o balanceamento se verifica que os dados assumiram padrões mais equilibrados na categoria. Considerando os dados expostos, sem o balanceamento se observa que o perfil identificação tem protagonismo em relação aos demais. Isto porque, segundo o parâmetro adotado para sua definição, o referido perfil admite duas opções possíveis de resposta – ativo ou passivo – quando verificada a interação no processo patentário. Os dados balanceados reduzem a ocorrência do desequilíbrio.

Os dados balanceados atualizados estão expostos no *dataset* nos atributos *"z introjection"*, *"z identification"* e *"z integration"*. O atributo *"Relação entre normas e valores pessoais"* traz a indicação do perfil de cada linha referente ao respectivo pesquisador.

O atributo *"Has strict relationship between standards / personal values? (Existe uma relação direta entre padrões/valores pessoais?)"* é adicionado no *dataset* tendo derivado de uma apuração mais aprofundada, pois são tomados os pesquisadores em cujos dados atendem aos parâmetros para atribuição do perfil, indicados como 'verdadeiro' e, os que têm presença parcial como 'falso'. Dentre os indicados como 'verdadeiro' se tem a seguinte distribuição: Identificação: 19 / Integração: 6 / Integração:4.

Superada a tarefa de balanceamento e tendo já sido realizada a anonimização dos dados resta considerar o *dataset* como um todo e avançar para a conclusão da etapa de transformação dos dados, parte final da fase de pré-processamento que antecede a mineração dos dados. A tarefa seguinte é a conversão dos dados categóricos em variáveis numéricas.

V.3.3 Conversão das variáveis categóricas em numéricas

Ainda na fase de pré-processamento de dados e dando sequência à etapa de transformação, tem-se que os algoritmos operam com atributos quantitativos ou qualitativos como entrada. O atributo é uma propriedade de um objeto. Um conjunto de atributos, por sua vez, congrega elementos descritivos de um objeto. Os atributos eleitos para serem convertidos são aqueles que mantêm relação direta com as variáveis de saída do *Framework*. Nota-se, então, que os atributos categóricos denominados no *dataset*, a saber, *"Research areas"* e *"Research subareas"* (áreas de pesquisa), assim também os atributos *"Bachelor's degree location"*, *"Master's degree location"*, *"Doctorate's degree location"* (locais de formação acadêmica), *"Birthplace"* (local de nascimento) do *dataset*, não serão convertidos.

As razões para a não conversão neste momento se dão, primeiramente, porque tais atributos são considerados variáveis de entrada sem relação direta com as categorias do *framework*. Assim, os atributos se manifestam em quantidade de opções que tornariam a tarefa mais complexa e com baixo potencial de resultado. O ponto aqui é direcionar esforços metodológicos em variáveis cuja conversão tenha maior contribuição para o resultado pretendido.

O *dataset* é integrado por atributos quantitativos e qualitativos ou categóricos, não suscetíveis, portanto, em sua totalidade, à contagem ou medição já que em parte contêm representações de classificação ou categoria. Daí a necessidade de, além da padronização dos dados, da conversão dos dados representados de forma categórica para o formato numérico (CORMEN *et al*., 2012).

O tipo ou formato de um atributo considera as propriedades que possui. Os atributos que compõem o *dataset* podem ser classificados, segundo Castro e Ferrari (2016), em categórico (binário, nominal e ordinal) e numeral (discreto, contínuo e razão), sendo que:

Atributo binário – assumem apenas dois valores, como “sim” ou “não”, podendo ser simétricos se as alternativas forem igualmente relevantes ou assimétricas caso não tenham a mesma relevância.

Atributo nominal – tem rótulos distintos e conhecidos que denominam ou simbolizam algo ou alguém sem indicar ordem ou posição, como “vermelho”, “preto”, “azul”.

Atributo ordinal – ordena as categorias indicando sequência ou classificação em que a ordem é importante e não se indica a magnitude entre elas, como em uma distribuição convencional, como “a, b, c, d, ...”.

Atributo discreto – representa um número finito de valores, geralmente representado por uma variável inteira, como idade, código CEP.

Atributo contínuo – representa um número infinito representado pelos números reais, podendo existir muitos valores entre os rótulos, como peso e temperatura.

Atributo razão – estabelecem medidas e padrões em geral, como “peso”, “distância”, “velocidade”, “remuneração”.

No *dataset* há atributos, principalmente os oriundos da plataforma *Lattes*, que estão no formato categórico e outros derivados de questões no padrão binário sim/não, principalmente oriundas, neste caso, da consulta via questionário *on-line*, e, portanto, se referem aos respondentes (n=131), cujas respostas foram convertidas para o padrão numérico binário 1 (sim) e 0 (não). Isso foi realizado para a resposta em relação a gênero, por exemplo, atribuindo o seguinte padrão: 1 (feminino) e 0 (masculino).

Considerando o *dataset* da presente pesquisa é a seguir desenvolvida a tarefa de conversão dos atributos categóricos. No caso em tela são chamados os seguintes atributos correspondentes às respostas binárias “sim” ou “não”, dispostos no *dataset*, convertidos para “1” ou “0”, respectivamente. São referentes às perguntas com respostas binárias. São os atributos:

- *Have you ever had a patent awarded? (Você já teve uma patente concedida?)*
- *Have you ever had any patents licensed? (Você já teve alguma patente licenciada?)*
- *Have you ever had a patent deposited abroad through Patent Cooperation Treaty (PCT)? (Você já teve uma patente depositada no exterior por meio do Tratado de Cooperação de Patentes (PCT)?)*

- *Have any patent request been the result of interaction with the industry?(Algum pedido de patente foi resultado de interação com o setor?)*

O código-fonte desenvolvido nesta seção está disponibilizado no Apêndice D, no quadro D3.

Os dados convertidos e já ilustrados no gráfico 12 no Capítulo IV são apresentados na tabela 14.

Atributos (perguntas)	**"1" – Sim**	**%**	**"0" – Não**	**%**
Você já teve uma patente concedida?	90	69%	41	31%
Você já teve alguma patente licenciada?	44	33%	87	67%
Você já teve uma patente depositada no exterior por meio do Tratado de Cooperação de Patentes (PCT)?	52	40%	79	60%
Algum pedido de patente foi resultado de interação com o setor?	59	45%	72	55%

Tabela 14 - Atributos binários convertidos para variáveis numéricas

Fonte: Elaborado pelo autor

Quanto ao gênero atributo *"Gender"* do *dataset*, também binário, foi convertido, conforme tabela 15.

Atributo Gênero	**Qte.**	**%**
Mulher "1"	47	36%
Homem "0"	84	64%

Tabela 15 - Conversão do atributo binário relativo a gênero.

Fonte: Elaborado pelo autor

O atributo relativo à data de nascimento, no *dataset "Birth Interval"*, representado no pela indicação do intervalo de cinco anos, o ano de início do intervalo foi utilizado como base para geração da conversão em contagem, gerando o gráfico 25.

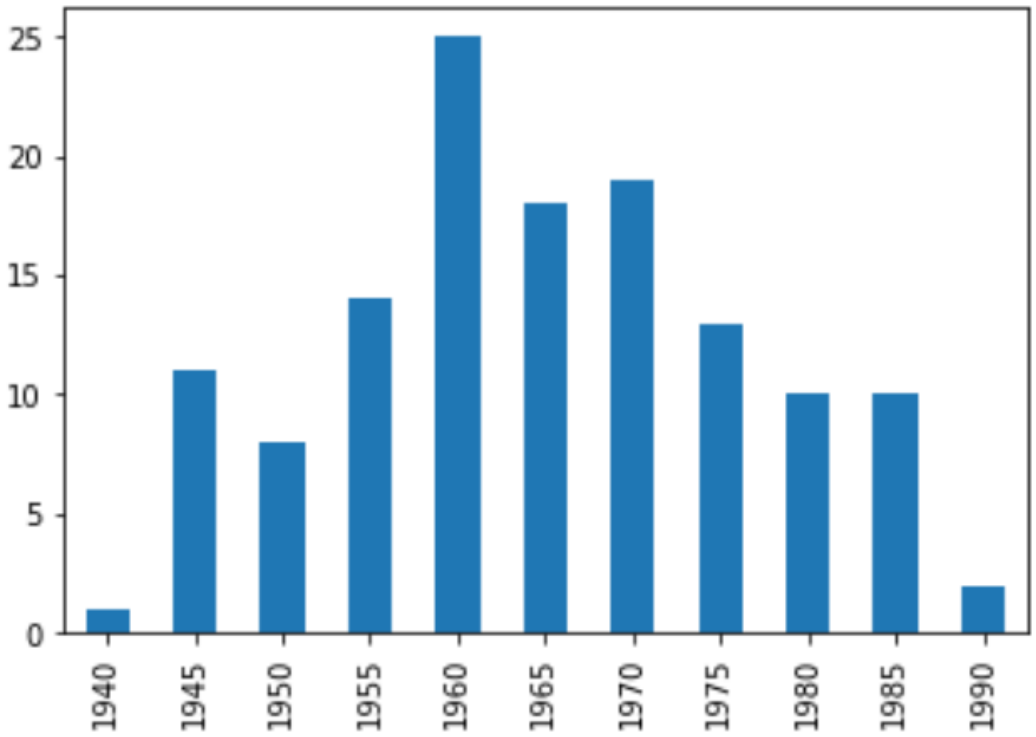


Gráfico 25 - Conversão em contagem por intervalo de ano de nascimento.
Fonte: Elaborado pelo autor

Este ponto, por fim, marca a finalização da fase de transformação e, consequentemente, do pré-processamento. Considerando o fluxo disposto na figura 16, têm-se o *dataset* ajustado e pronto para a fase de mineração de dados. A seção seguinte traz conceitos pertinentes à mineração de dados e à construção do algoritmo para então serem apresentados os perfis classificados à luz do *framework*.

O *dataset* atualizado pela conversão dos atributos categóricos e o código-fonte com as aplicações desenvolvidas via *script* em *Python* nesta e em outras seções estão disponíveis conforme informações nos apêndices C e D, respectivamente.

V.4 A mineração de dados

Na esteira do fluxo de descobrimento do conhecimento em base de dados, após o término do pré-processamento, conforme já apresentado na Figura 15, se tem a fase de mineração de dados (FAYYAD *et al*., 1996b).

A mineração é a fase em que se empreende, segundo Castro e Ferrari (2016), a depuração de grandes bases de dados. O processo se dá na exploração do *dataset,* podendo se empregar algoritmos como ferramentas a fim de extrair conhecimento.

Para tanto esta fase conta com instrumentos computacionais e técnicas de aprendizagem de máquina e modelos de algoritmos.

São duas as principais tarefas da mineração de dados. A *descritiva* se dá pelo reconhecimento e caracterização das propriedades gerais dos dados. A *preditiva* infere a partir dos dados com o objetivo de prever padrões. Para Fayyad *et al*. (1996b), é nesta fase que é tomada a decisão de quais algoritmos serão aplicados aos dados para reconhecimento de padrões, aprendizado de máquina e de banco de dados. Os autores descrevem uma combinação de três elementos presentes nestes algoritmos: *um modelo desenhado para uma função; um critério de parametrização; o algoritmo de pesquisa em si.*

Para tanto se ressalta que o algoritmo é um procedimento que sistematiza e sequência operações finitas para a partir de dados de entrada oferecer soluções para um problema computacional (CORMEN *et al*., 2012).

O aprendizado de máquina (*machine learning*) é uma área de pesquisa afeita a campos de conhecimentos como estatística, ciência da informação, biologia, dentre outros. Seu objetivo é desenvolver sistemas computacionais cujo desempenho se dá a partir de experiências cujo aprimoramento é constante e gradativo por meio de mudanças e melhoramentos no seu padrão comportamental a exemplo do modelo da aprendizagem humana (CASTRO; FERRARI, 2016). Dentre os seus principais métodos ou técnicas estão as regras de associação e classificação, nesta última se ressalta a árvores de decisão. Na presente pesquisa, considerando os objetivos, se desenvolve um modelo de classificação baseado em árvore de decisão.

V.4.1 O algoritmo de árvore de decisão

O aprendizado de máquina é um campo central da inteligência artificial com proeminência em algumas áreas de estudo. Sua aplicação automatiza o desenvolvimento de modelos com a capacidade de aprendizagem, identificando padrões e tomando decisões a partir de dados de entrada e saída. O sistema simula a dinâmica de aprendizado e consegue prever a partir do seu próprio desempenho a fim de aprimorá-lo. Esse modelo de treino pode inclusive oferecer possibilidades/caminhos para decisões múltiplas, como aquelas próprias da técnica *decision tree* (árvore de decisão) (MERTENS *et al*., 2018).

Tem sido cada vez mais recorrente a aplicação da linguagem de programação associada à aprendizagem de máquina. Kadiyala e Kumar (2018), por exemplo, aplicam os três pacotes de aprendizados de máquina (*scikit-learn*, *lightgbm* e *xgboost*) no desenvolvimento de quatro algoritmos em *Python* baseados em árvore de decisão visando prever as concentrações de carbono em ônibus urbanos a partir de bancos de dados de emissão de CO_2. Também Mertens *et al*., (2018) desenvolveram um pacote *Python*, *Abrox*, para computação bayesiana aproximada com foco na comparação de modelos. Raychev *et al*., (2016) apresentam um modelo probabilístico construído a partir de códigos que automatiza tarefas de programação, usando árvores de decisão no contexto de programas de aprendizagem.

Inspirado em um fenômeno orgânico observado na natureza os algoritmos baseados no modelo denominado árvore de decisão contam com a simplicidade e eficácia do ambiente natural para representar, modelar e classificar dados. Como um instrumento de abstração o formato de árvore tem sua importante virtude em fornecer um modo intuitivo e mais compreensível de arranjo e análise de dados (QUINLAN,1993).

Na esteira do conceito de inteligência artificial, sobretudo em um cenário que demanda decisões múltiplas visando à classificação de uma amostra, uma forma de abstração é o modelo trazido na técnica da árvore de decisão que se apresenta como uma forma de representação que viabiliza a verificação de padrões em grandes bancos de dados, permitindo sua categorização, predição e modelagem (QUINLAN, 1986; MERTENS *et al*., 2018).

O modelo agrega o conceito de aprendizagem de máquina à extração de informações, explícitas ou não. Dentre as técnicas de mineração de dados empregadas em um problema de classificação se tem a árvore de decisão e rede neural, por exemplo. Esta última geralmente mais utilizada em maiores volumes de dados, além de outros que corroboram para o tratamento e modelagem dos dados com a necessária representação, evidenciação e validação.

Esses modelos, segundo Kovács (1996), têm a possibilidade de extrair e reconhecer padrões e correlações imperceptíveis ao usuário. Podem extrair características ocultas em dados brutos que são tomados como exemplos, permitindo seu agrupamento e classificação. Por meio da inteligência artificial, a aprendizagem se dá mediante a

exposição do modelo a um volume de dados, atingindo contínuas e progressivas melhorias.

Quando se considera bancos de dados robustos (*Big Data*), além de tratar e reconhecer padrões ou metadados, os algoritmos são empregados no alcance de soluções. Neste caso, há necessidade de não somente obter os dados, mas evidenciar o caminho traçado neste objetivo. O modelo da árvore de decisão, segundo Martins (2021), segue o conceito *white-box* ou *white-model*, que, ao contrário do modelo *black-box*, permite, além de ajuste de parâmetros, uma maior explicitação e transparência das informações da estrutura e da implementação do algoritmo (DELIBAŠIĆ *et al*.,2012).

A árvore de decisão é um classificador que permite o tratamento e modelagem de dados baseado no conceito trazido por Quinlan (1993) de ‘dividir para conquistar’, ou seja, dividir em sucessivos conjuntos e subconjuntos válidos até que restem classes únicas ou majoritárias.

Concebe-se, assim, um modelo classificador de dados baseado em algoritmo de aprendizagem, tendo como entrada uma base de informações qualitativas e quantitativas que submetida a um modelo analítico organiza as saídas em classes que têm como base o valor de atributos. A árvore de decisão tem ainda seu valor na conjugação de tempo de processamento, na forma intuitiva de examinar os resultados e, finalmente, em representar de maneira simples e compreensível a estrutura final do classificador (QUINLAN, 1986, 1993; CABENA *et al.,*1997).

De acordo com Quinlan (1993), o livro de Earl Hunt 1966 intitulado “*Experiments in Induction*” introduziu os classificadores baseados em árvore de decisão. Os autores também citam o algoritmo CART, desenvolvido por Friedman em 1984, que faculta a construção de algoritmos de aprendizado de máquina supervisionado para classificação (prever categorias discretas, tipo sim ou não, ou com mais categorias) e regressão (prever valores numéricos, como lucros, medidas, etc.). Posteriormente Quinlan (1993) desenvolveu o algoritmo ID3 que foi sucedido pelos algoritmos C4.5 e C5.0.

Quanto ao formato da árvore de decisão, a sua estrutura se compõem de nodos (atributos) de onde derivam os arcos ou ramos (ligações) e, por fim, os nodos que são as folhas (classes). O algoritmo seleciona um atributo ou nodo principal, valorando e

dividindo o conjunto de casos, conferindo um ramo para cada valor possível deste atributo. Assim, cada novo ramo, se passível de subdivisão, é atribuído um novo nodo, representando, por fim, a estrutura em que seja possível a realização de um modelo treino para a classificação.

Tem-se, assim, conforme Freitas (1997), um algoritmo representado na Figura 77 em formato genérico para a classificação por árvore de decisão, onde **S** está representando um conjunto de dados a ser classificado e um modelo ou caminho para treino do algoritmo:

---

(1) Se todos os exemplos no atual conjunto de exemplos S satisfazem um critério de parada
então
(2) cria um nodo folha com algum nome da classe e pára;
senão
(3) seleciona um atributo A para ser utilizado como um atributo de particionamento e cria um nodo com o nome do atributo de particionamento;
(4) escolhe um teste sobre os valores de A, com resultados mutualmente exclusivos e coletivamente exaustivos $R_1,...,R_k$, e cria um ramo, a partir do nodo recentemente criado, para cada teste;
(5) particiona S nos subconjuntos $S_1,...S_k$, tal que cada S_i, i=1..,k, contenha todos os exemplos em S com resultado R_i do teste escolhido;
(6) aplica este algoritmo recursivamente para cada subconjunto S_i, i=1,...k;
fim_senão
fim_se

Figura 77 - Algoritmo genérico para construção de árvores de decisão
Fonte: Freitas (1997)

Observa-se que as árvores de decisão se constituem em um método prático utilizado na aprendizagem intuitiva. O algoritmo ou função aprendida é representado no formato de árvore em uma estrutura do tipo *top-down*. O modelo de análise é preditivo, caracterizando-se por um método de busca e classificação em um intervalo de hipóteses, orientado por uma hipótese que atende ao modelo de treinamento.

No que tange ao modelo eleito, traz-se a inspiração original - o exemplo primeiro apresentado por Quinlan que utiliza a necessidade de classificar as manhãs de sábado mais adequadas segundo as condições climáticas para jogar tênis, representado na Figura 88, apresentada em inglês a fim de manter sua originalidade.

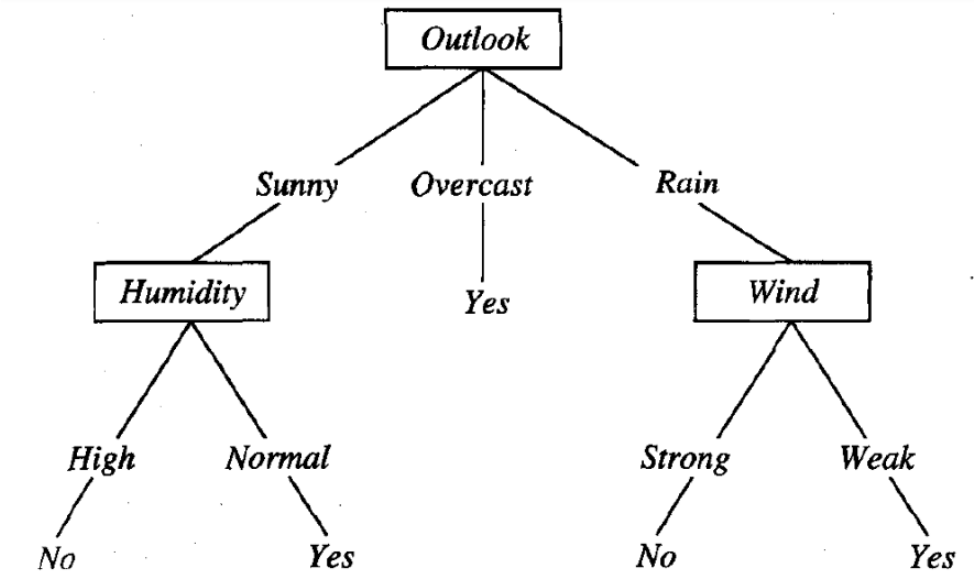


Figura 88 - Jogar tênis ou não? Um modelo de árvore de decisão
Fonte: Quinlan (1993; 1986)

Neste exemplo, considerando as diferentes informações e relações possíveis, tem-se, a partir do dicotomizador iterativo 3 ou ID3, o clássico modelo de decisão que, por fim, oferece o resultado esperado baseado nas referências: *se o clima for de sol e a umidade for normal é possível jogar tênis.* O ID3, cujo aprendizado usa a estratégia *top-dowm*, é o conjunto de árvores de decisão possíveis no modelo. Os modelos C4.5 e C5.0, sucessores do modelo anterior, são mais recentes.

Têm-se, então, por um conjunto de regras (se – então) a testagem dos atributos (*Outlook, Temperature, Humidity, Wind*):

Outlook – é o melhor atributo e, portanto, está na raiz da árvore. É o ***Nó Raiz*** e testa o atributo principal. Neste modelo, realizado o teste de ganho de informação, revelou qual atributo deve ser testado primeiro na estrutura.

Humidity; Wind – ***nós*** descendentes.

Sunny; Overcast; Rain – **Ramos** ou **linhas** que partem de um nó e corresponde a um valor ao atributo testado.

Yes, No – **Folhas** ou **classes**, atribuindo uma classificação, sendo a última possibilidade de classificação.

Assim, os **nós** testam os atributos. Os **ramos** correspondem ao valor do atributo. As **folhas** atribuem uma classificação.

Por fim, a consideração e adoção do modelo e estrutura da árvore de decisão se revelam mais próprios para o objetivo da presente pesquisa, já que a técnica agrega a simplificação desejada de concepção, desenvolvimento e apresentação do conhecimento gerado. É especialmente eficiente na construção de algoritmos classificadores com saídas de predição de classes, tomando por base os valores de características de um volume de dados.

Seguindo o escopo da pesquisa em tela, como a ferramenta computacional de classificação, o aprendizado de máquina supervisionado é aplicado por meio do algoritmo de árvore de decisão.

V.4.2 A avaliação do modelo classificador

Na sequência das etapas do processo de descoberta de conhecimento de dados, posterior à mineração de dados é realizada a interpretação e análise, neste caso, do modelo classificador a fim de verificar a confiabilidade da inferência realizada. Para tanto, se realiza um ranqueamento a fim de verificar a importância dos dados de entrada para a classificação de atributos de saída.

Ressalta-se que foram destacados os atributos determinantes para a classificação, considerando o escopo definido pelo *framework* e que, portanto, contribuem para o modelo classificador pretendido – os denominados atributos alvos, principalmente os numéricos e binários – com aplicação da validação cruzada na etapa de treino.

No desenvolvimento de modelos preditivos em que o algoritmo é treinado para gerar determinada classificação, os dados da própria amostra são utilizados na dinâmica de treino e teste do modelo. O processo se dá pela divisão dos dados em subconjuntos destinados para treino e teste/avaliação do modelo.

No caso de aprendizado supervisionado em que se prediz o resultado a partir do histórico, o emprego do total de dados para treino e teste não é o recomendado, sendo mais comum separar parte do conjunto para cada uma destas tarefas a fim de indicar o modelo mais eficiente de treino e com melhor resultado validado. A validação cruzada (*cross validation*) é uma técnica que, segundo Castro e Ferrari (2016), aperfeiçoa a dinâmica treino/teste. São utilizadas partes da amostra dedicada ao treino, visando realizar predições se valendo de dados não expostos ao treinamento. A técnica consiste em dois passos: dividir os dados em subgrupos e variar o treinamento e teste

entre os grupos em diferentes rodadas. Tais condições são fundamentais: os grupos devem ter o mesmo porte ou tamanho de variáveis; todos os grupos são utilizados para treinar o modelo, exceto um, destinado à validação.

O grupo de dados é divido em *k folds* em que cada seção é denominada *fold*, representando a quantidade de interações em que cada grupo da amostra se sucede no processo de treinamento, enquanto cada seção sucessiva e exaustivamente é reservada exclusivamente para o teste a cada rodada de interação. A figura 19 revela essa dinâmica.

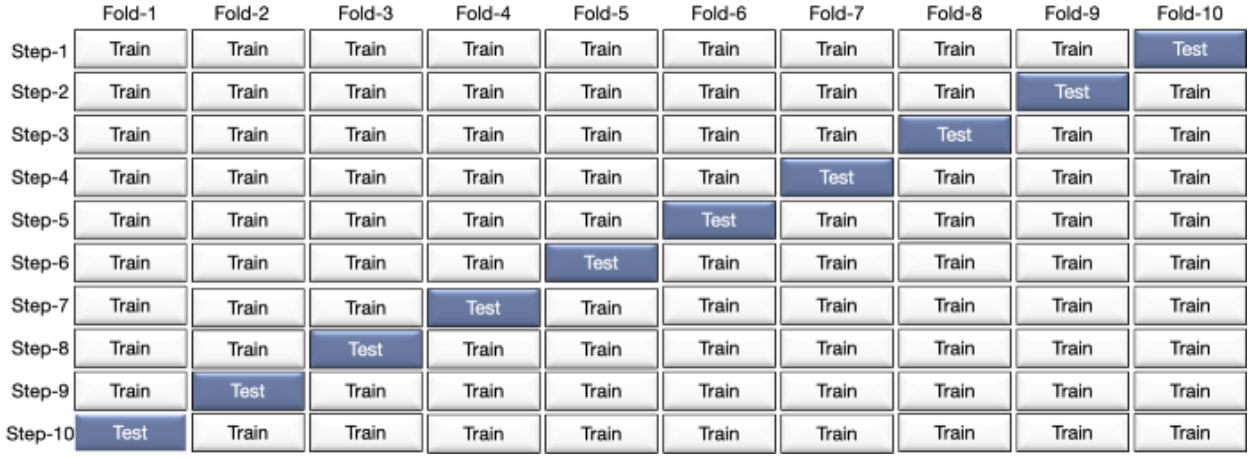


Figura 99 - Sequência do processo da validação cruzada.

Fonte: Disponível em: <https://towardsdatascience.com/what-is-cross-validation-60c01f9d9e75>. Acesso em: 10 abr. 2023.

Assim todos os dados são utilizados, ao fim das interações, tanto para treino como para teste, sendo o todo validado. A cada interação é gerado um resultado de acerto, um *score* de resultados. A soma do valor obtido de cada interação é dividida pelo número de interações - *k* (rodadas de treino/teste).

Quanto ao *dataset* considerado, nesta etapa, são tomados os dados dos respondentes para treino (dados públicos e institucionais). Então é realizada a validação cruzada para cada camada dos atributos alvos com relação direta no resultado pretendido no *framework* a fim de servir ao treino do algoritmo e obter a importância dos dados de entrada para cada atributo alvo – as saídas.

O código-fonte do algoritmo de árvore de decisão que emprega a validação cruzada para a determinação de importância e geração do ranqueamento dos atributos de entrada está disponível em repositório conforme informações apresentadas no Apêndice D.

A seguir são expostas as posições de importância dos atributos de entradas referentes às diferentes saídas, camadas do *framework*.

A escala de importância está definida entre “0” – pouco importante e “1” – muito importante.

“Interação no processo de patenteamento. Ativo ou passivo?”

	Qte.	**%**
Passivo	72	55%
Ativo	59	45%

Ranqueamento – Escala de importância 0 a 1

0	Artigos completos publicados em periódicos	0.0
1	Capítulos de livros publicados	0.0
2	Trabalhos completos publicados em anais de congressos	0.0
3	Resumos publicados em anais de conferências	0.0
4	Número de patentes e solicitações de registro	0.0
5	Você já teve uma patente concedida?	0.0
6	Você já teve alguma patente licenciada?	0.0
7	Você já teve uma patente depositada no exterior por meio do Tratado de Cooperação de Patentes (PCT)?	0.0
8	Gênero	0.0
9	Intervalo de nascimento	0.0
10	Algum pedido de patente foi resultado de interação com a indústria?	1.0

Como o atributo testado está vinculado a uma pergunta direta do questionário se tem uma espécie de teste de sanidade ou validade do algoritmo, já que a pergunta “Algum pedido de patente foi resultado de interação com a indústria?” apresentou relação de importância máxima, considerando o atributo do qual origina o dado.

Importante ressaltar que a camada em tela atesta a frequência e parceria, considerando o corte amostral que admitiu os pesquisadores com pelo menos 05 patentes. Assim, todos são considerados ativos, restando aferir a relação com a indústria no processo de patenteamento, em caso positivo ativo, no contrário, passivo.

“Classificação quanto à orientação profissional”

	Qte.	**%**
Híbrido	72	55%
Empreendedorismo	44	34%
Tradicional pragmático	11	08%
Tradicional puro	04	03%

Ranqueamento – Escala de importância 0 a 1

0	Você já teve uma patente depositada no exterior por meio do Tratado de Cooperação em Matéria de Patentes (PCT)?	0.0
1	Tratado de Cooperação em Matéria de Patentes (PCT)?	0.017636
2	Algum pedido de patente foi resultado de interação com a indústria?	0.027491
3	Gênero	0.042437
4	Você já teve alguma patente licenciada?	0.04409
5	Você já teve uma patente concedida?	0.071494
6	Capítulos de livros publicados	0.081526
7	Intervalo de nascimento	0.10655
8	Número de patentes e solicitações de registro	0.163064
9	Resumos publicados em anais de conferências	0.203017
10	Artigos completos publicados em periódicos	0.242695

Nesta camada em que se afere a orientação profissional se tem o quadro claro de desbalanceamento que prejudica o treino/ processo de aprendizagem do algoritmo.

O atributo mais importante para esta classe alvo foi o “Artigos completos publicados em anais de congressos e “Artigos completos publicados em periódicos”.

“Natureza da motivação”

	Qte.	**%**
Extrínseca	*44*	*34%*
Mista	*44*	*34%*
Intrínseca	*43*	*32%*

Ranqueamento – Escala de importância 0 a 1

0	Você já teve uma patente concedida?	0.0
1	Você já teve alguma patente licenciada?	0.0
2	Gênero	0.0
3	Algum pedido de patente foi resultado de interação com a indústria?	0.009369
4	Você já teve uma patente depositada no exterior por meio do Tratado de Cooperação de Patentes (PCT)?	0.050956
5	Artigos completos publicados em anais de congressos	0.112619
6	Resumos publicados em anais de congressos	0.124976
7	Intervalo de nascimento	0.157203
8	Número de patentes e solicitações de registro	0.157784

9	Artigos completos publicados em periódicos	0.191233
10	Capítulos de livros publicados	0.195860

Nesta classe de Natureza de motivação a distribuição da classe está balanceada. Os atributos que mais contribuem são “Artigos completos publicados em periódicos” e “Capítulos de livros publicados”. Interessante notar a contribuição do ano de nascimento “Intervalo de nascimento para esta classe alvo. Ressalta-se ainda a baixa importância do atributo “Gênero”, por exemplo.

“Relação entre normas / valores pessoais”

	Qte.	**%**
Identificação	55	*42%*
Integração	42	*32%*
Introjeção	34	*26%*

Ranqueamento – Escala de importância 0 a 1

0	Você já teve uma patente concedida?	0.0
1	Gênero	0.0
2	Você já teve alguma patente licenciada?	0.0276
3	Você já teve uma patente depositada no exterior por meio do Tratado de Cooperação em Matéria de Patentes (PCT)?	0.03205
4	Número de patentes e solicitações de registro	0.058065
5	Capítulos de livros publicados	0.078352
6	Intervalo de nascimento	0.115256
7	Artigos completos publicados em anais de congressos	0.13853
8	Resumos publicados em anais de congressos	0.157464
9	Artigos completos publicados em periódicos	0.195984
10	Algum pedido de patente foi resultado de interação com a indústria?	0.196699

Para a classe alvo que mensura a relação entre normas e valores pessoais há uma distribuição balanceada e os atributos com relação com à publicação científica. Detalhe que o atributo que representa a interação com a indústria no processo de detalhamento tem a maior importância.

Na construção de um modelo classificador com aprendizado de máquina é necessário estabelecer métricas de avaliação do treino e de desempenho. A métrica depende do

problema em análise, mas sempre busca avaliar a performance do modelo em predizer. Segundo Castro e Ferrari (2016), uma medida de avaliação de classificadores que apura a confiabilidade é a acurácia (*accuracy*), um indicador simples, mas importante, dado que considera o percentual de acertos, relacionando-o com o total de entradas.

O objetivo da acurácia é de representar quantitativamente a qualidade do modelo classificador, apurando como o classificador desempenha em dados não usados para treinamento e para tanto se emprega a técnica da validação cruzada, tendo como resultado a acurácia média em cada conjunto testado.

O código-fonte desenvolvido nesta seção está disponibilizado no Apêndice D, no quadro D4. Por fim, há a análise da acurácia dos atributos alvos com relação às camadas do *framework* com a aplicação da validação cruzada.

Foram realizados um total de 25 repetições de *4-fold cross validation*, cada uma delas sorteando a divisão dos dados nos *folds*. Procedimento que visou à robustez estatística propiciada pela avaliação exaustiva do classificador considerando diferentes amostras de treino e teste. A acurácia média e outros indicadores de desempenho de cada variável de saída do modelo evidenciados estão na tabela 16 e gráfico 26.

ACURÁCIA	Interação no processo de patenteamento. Ativo ou passivo?	Classificação quanto à orientação profissional	Natureza da motivação	Relação entre normas / valores pessoais
Contagem	100.0	100.0	100.0	100.0
Média	**1.0**	**0.443324**	**0.344754**	**0.435559**
Desvio Padrão	0.0	0.081308	0.076904	0.071311
Mínimo	1.0	0.242424	0.156250	0.242424
Primeiro quartil - 25%	1.0	0.372159	0.303030	0.393939
Mediana - 50%	1.0	0.430871	0.343750	0.454545
Terceiro quartil - 75%	1.0	0.515152	0.393939	0.484848
Máximo	1.0	0.636364	0.500000	0.625000

Tabela 16 - Apuração da acurácia obtida nos atributos alvos.

Fonte: Elaborado pelo autor

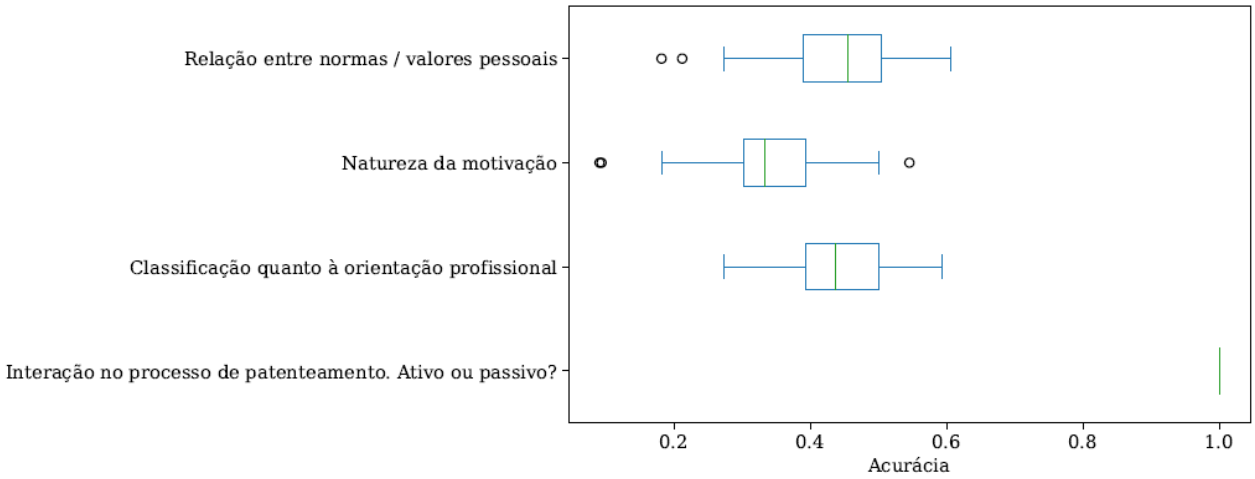


Gráfico 26 - Representação gráfica da acurácia obtida nos atributos alvos
Fonte: Elaborado pelo autor

O que se depreende dos dados é que, por exemplo, a acurácia da camada natureza da motivação está próxima a 30%, enquanto a que afere a interação com a indústria no processo de patenteamento é de 100%. A disparidade indica que a aferição da precisão depende do número de categorias de classificação possíveis (variáveis de saída) – mais categorias, maior dificultade - e dos atributos tomados como parâmetros (variáveis de entrada) – se as variáveis são binárias, maior a assertividade.

V.5 Perfil motivacional

Concluído o tratamento e padronização dos dados e extração de informações do *dataset* via *KDD* se tem a execução do modelo classificador na amostra de pesquisadores a fim de se definir o perfil motivacional da amostra.

Assim, são tomadas as classificações de perfis referentes à amostra, considerando a parcela dos respondentes (n=131), contando com dados públicos e institucionais e, portanto, com quantidade maior de atributos como variáveis de entrada. A parcela de não respondentes (n=319), por sua vez, conta com os dados públicos e, assim, com quantidade inferior de variáveis de entrada.

Para a classificação da parcela de respondentes (n=131) e considerando possuir a completude dos dados, basta a aplicação do método de submissão ao *framework* cujos parâmetros foram anteriormente apresentados, sem a necessidade do emprego de ferramentas computacionais nesta tarefa.

Na classificação dos não respondentes (n=319), justamente por conta da incompletude dos dados, cabe a aplicação do algoritmo desenvolvido e treinado que toma o

padrão/modelo de respostas/comportamento extraído dos respondentes para aprender e inferir/predizer o comportamento e operacionalizar a classificação da amostra não respondente.

V.5.1. Classificação dos respondentes (n=131)

Quanto aos respondentes são trazidas duas versões de modelo classificador. A tabela 17, ilustrada na figura 20, apresenta os dados não balanceados, colhidos da base de currículos que combinados com as respostas do questionário são submetidos aos parâmetros do *framework*.

Dimensões	Camadas	Atributos	Classificação	
Pessoal	**Natureza motivacional**	Extrínseca	21	16%
		Intrínseca	110	84%
		Mista	0	0%
	Resultados esperados	Econômico	21	16%
		Desafio	110	84%
		Reputação	0	0%
Comportamental	**Orientação profissional**	Tradicional Puro	04	03%
		Tradicional Pragmático	11	08%
		Híbrido	72	55%
		Empreendedor	44	34%
	Frequência e parceria	Ocasional Passivo	0	0%
		Ocasional Ativo	0	0%
		Serial Passivo	72	55%
		Serial Ativo	59	45%
Ambiental	**Relação entre normas e valores pessoais**	Introjeção	34	26%
		Identificação	79	60%
		Integração	18	14%

Tabela 17 - Classificação não balanceada da parcela respondente da amostra.

Fonte: Elaborado pelo autor

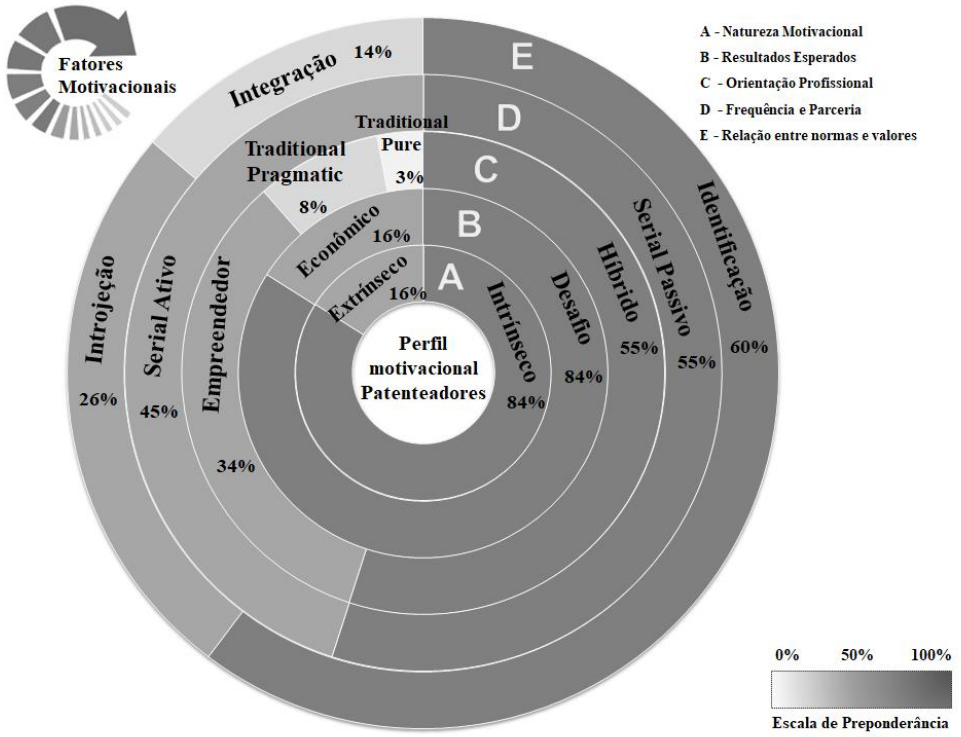


Figura 20 - Perfil motivacional não balanceado da parcela respondente da amostra.
Fonte: Elaborado pelo autor

A tabela 18, ilustrada na figura 21, apresenta os dados balanceados referentes aos respondentes.

Dimensões	Camadas	Atributos	Classificação	
Pessoal	Natureza motivacional	Extrínseca	44	34%
		Intrínseca	43	33%
		Mista	44	34%
	Resultados esperados	Econômico	44	34%
		Desafio	43	33%
		Reputação	44	34%
Comportamental	Orientação profissional	Tradicional Puro	04	03%
		Tradicional Pragmático	11	08%
		Híbrido	72	55%
		Empreendedor	44	34%
	Frequência e parceria	Ocasional Passivo	0	0%
		Ocasional Ativo	0	0%
		Serial Passivo	72	55%
		Serial Ativo	59	45%

Ambiental	**Relação entre normas e valores pessoais**	Introjeção	34	26%
		Identificação	55	42%
		Integração	42	32%

Tabela 18 - Classificação balanceada da parcela respondente da amostra.
Fonte: Elaborado pelo autor

Comparando os dois modelos classificadores, distintos em função do balanceamento dos dados, a assertividade da classificação é quase absoluta, já que se dispõe de todos os dados (públicos e institucionais) submetidos ao *framework* a fim de definir o perfil dos pesquisadores. Porém, se nota um viés natural no modelo não balanceado, mais perceptível na camada natureza da motivação e resultados esperados. Isso pode resultar do fato de se tratar de uma amostra vinculada à área da educação em que as variáveis intrínsecas e afeitas à solução de desafios, no caso, acadêmicos, têm maior ocorrência. Isto fica menos evidenciado no modelo balanceado que, no limite, expressa melhor a realidade e equilíbrio dos dados. Isto é especialmente útil e confere qualidade à amostra utilizada como treino, aumentando o nível de confiabilidade do algoritmo na inferência.

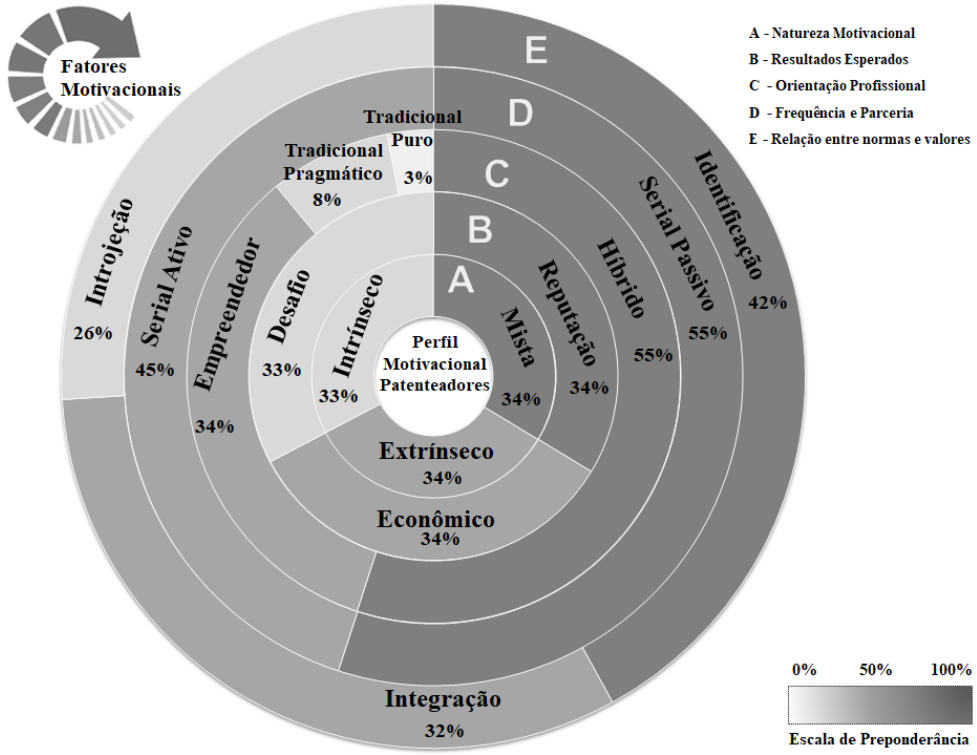


Figura 21 - Perfil motivacional balanceado da parcela respondente da amostra.

Fonte: Elaborado pelo autorV.5.2. Classificação dos não respondentes (n=319)

Quanto à amostra dos não respondentes a classificação resulta de um processo de inferência realizado pelo algoritmo treinado. A tabela 19, ilustrada na figura 22 traz a classificação inferida da amostra não respondente (n=319).

Dimensões	Camadas	Atributos	Classificação	
Pessoal	**Natureza motivacional**	Extrínseca	120	38%
		Intrínseca	115	36%
		Mista	84	26%
	Resultados esperados	Econômico	120	38%
		Desafio	115	36%
		Reputação	84	26%
Comportamental	**Orientação profissional**	Tradicional Puro	9	3%
		Tradicional Pragmático	20	6%
		Híbrido	163	51%
		Empreendedor	127	40%
	Frequência e parceria	Ocasional Passivo	0	0%
		Ocasional Ativo	0	0%
		Serial Passivo	180	56%
		Serial Ativo	139	44%

Ambiental	**Relação entre normas e valores pessoais**	Introjeção	92	29%
		Identificação	144	45%
		Integração	83	26%

Tabela 19 - Classificação da parcela não respondente da amostra.
Fonte: Elaborado pelo autor

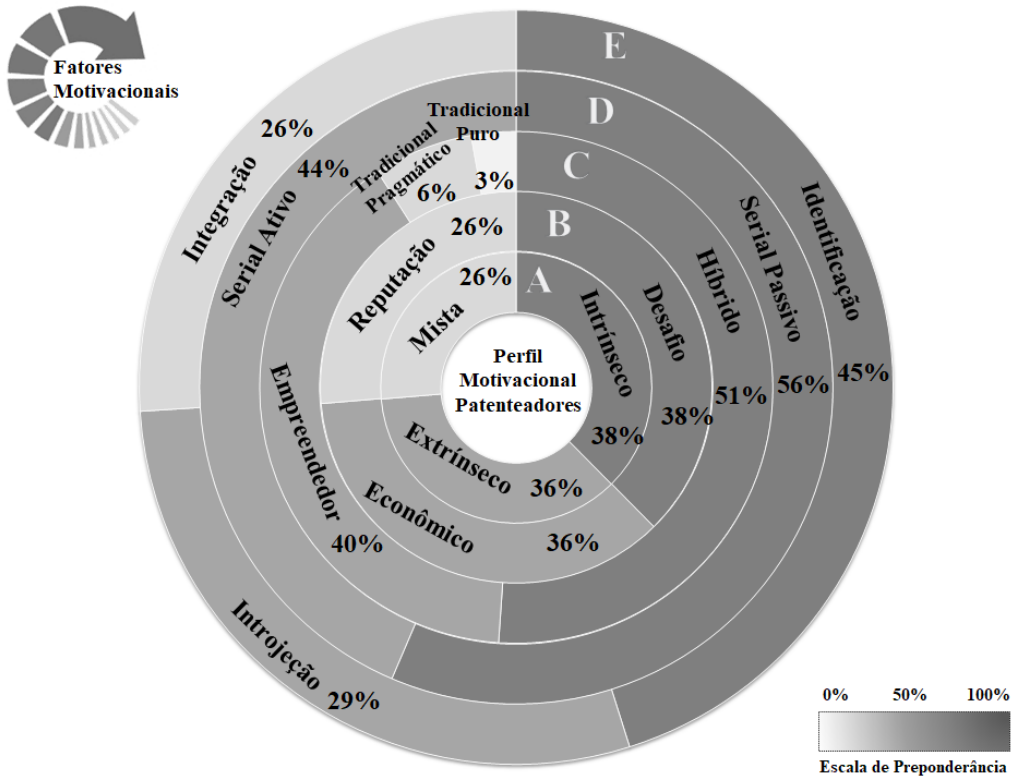


Figura 22 - Perfil motivacional da parcela não respondente da amostra.
Fonte: Elaborado pelo autor

A tabela 20 é tomada como apoio e legenda da figura 22 trazendo as medidas de avaliação da classificação. Foram realizadas avaliações de desempenho da predição do algoritmo para cada das cinco camadas do modelo, referenciadas de A a E.

CAMADAS		Árvore de decisão Padrão		Árvore de decisão Otimizada
		Acurácia	**F1 Macro**	**F1 Macro**
A	Natureza Motivacional	31%	30%	32%
B	Resultados Esperados	31%	30%	32%
C	Orientação Profissional	41%	23%	31%
D	Frequência e Parceria	53%	53%	62%
E	Relação normas/valores	34%	31%	32%

Tabela 20 - Medidas de avaliação da classificação da parcela não respondente.
Fonte: Elaborado pelo autor

Tomando a árvore de decisão em sua configuração padrão, pré-definida do módulo *scikit-learn*, se tem as médias de acurácia. Quando avaliada pela *Escore F1 macro-ponderado*, medida de desempenho mais sensível e que penaliza mais os dados desbalanceados (Lipton *et al*., 2014), se apresenta uma redução no desempenho em algumas dimensões do *framework*. Foram feitos ajustes de configuração na árvore de decisão buscando uma alternativa melhorada se comparada à padrão. Trata-se de uma versão otimizada, modelo que, por fim, aprimora o desempenho, mas tem uma predição que privilegia os elementos/variáveis mais comuns em cada camada.

Os códigos-fonte dos testes de avaliação, dos ajustes realizados e das classificações obtidas estão disponibilizados conforme informações apresentadas no apêndice D.

Considerando e combinando a classificação obtida das parcelas respondentes e não respondentes se tem o seguinte cenário quanto à amostra de patenteadores vinculados às universidades admitidas na pesquisa:

- **Suas motivações têm natureza predominantemente intrínseca e quanto aos resultados esperados é orientada pela superação de desafios inerentes à pesquisa.** Neste ponto, há alinhamento com o que Göktepe (2008), em pesquisa que considerou a realidade da universidade *Lund*, na Suécia, identificou como uma curiosidade inata dos cientistas a solução de problemas e um fascínio pelo processo de pesquisar e resolver questões que supera o interesse em benefícios econômicos e em retornos reputacionais. O trabalho de Göktepe fez menção quanto à satisfação dos inventores em resolver desafios inerentes à pesquisa. O estudo de Lam (2011), realizado no Reino Unido, salienta que as motivações intrínsecas são mais promissoras para o engajamento comercial de pesquisadores. No cenário nacional a pesquisa de

Oliveira (2011), também focada em universidades públicas, apesar de restrita às paulistas, identifica como um relevante fator motivador o fato de saber o pesquisador que sua atividade contribui para a geração de conhecimento na sua área de estudos. Por fim, as motivações intrínsecas atreladas à superação de desafios intelectuais são, em tese, consideradas as 'mais nobres' segundo a regra moral comum e, portanto, sustentar essa posição seja mais atrativo.

- **Quanto à orientação profissional há predominância do perfil híbrido seguido da identificação com o perfil empreendedor.** Assim como considerado por Lam (2011) e também Closs (2013), esta camada afere a percepção das parcerias com o meio privado e as implicações comerciais e sociais no processo de pesquisa. Ambos os perfis (híbrido e empreendedor) representam pesquisadores atuantes na academia e que valorizam a relação ciência-indústria, reconhecendo e colaborando nos efeitos desta interação no avanço científico, com maior atenção pelos empreendedores quanto às apropriações comerciais de seus inventos. Aqui parece se evidenciar algo que destoa da realidade internacional: a razão do protagonismo universitário brasileiro no patenteamento, observado no número de pedidos de patentes em comparação a instituições privadas nacionais, detentoras de maiores recursos, orientadas pelo mercado e com menos entraves burocráticos. Mormente há tanto apontamentos negativos aos denominados entraves e burocracias próprias do ambiente público, quanto menções ao predomínio de expectativas comerciais pela indústria. Os relatos colhidos no questionário de pesquisa que remontam aos sucessos e dissabores desta relação foram trazidos na apresentação dos dados sociodemográficos no capítulo IV.
- **Quanto à frequência de patenteamento, considerando que todos os patenteadores admitidos têm pelo menos cinco pedidos de patentes, toda a amostra é considerada serial. Na aferição da parceria no desenvolvimento da pesquisa e, sobretudo, no patenteamento, se tem um paradoxo com a camada orientação profissional: baixa interação com a indústria no processo de pesquisar e patentear (passivo), mas uma elevada demonstração de espírito empreendedor de que deriva a expectativa de comercialização dos inventos patenteados (ativo).** Nota-se que, na parceria, há distribuição quase igualitária, revelando o que pode ser uma

fase de transição em direção, ao que indica, a uma maior aproximação com o ambiente privado.

- **Quanto à relação entre as normas institucionais e valores pessoais há predomínio do perfil denominado identificado.** Neste perfil há voluntária intenção de harmonizar sua conduta com a expectativa do meio. Indica maior consciência e apreciação do seu papel institucional que pode resultar numa escolha de agregar valor, inclusive satisfação pessoal, no processo de pesquisa. Este perfil, contudo, é mais sensível a estímulos a fim de orientar seu comportamento e, portanto, responder bem a políticas gerenciais e treinamentos. Na sequência há maior ocorrência do perfil introjetado, mais orientado pelo atendimento compulsório de normas, seguido do perfil integrado, mais identificado naturalmente com regulamentos e de comportamento mais conservador quanto às relações institucionais.

CAPÍTULO VI – REVISÃO E CONSIDERAÇÕES FINAIS

Considerando o referencial teórico consultado, o escopo da investigação, a metodologia empregada e o volume de dados que culminaram na geração do modelo classificador, o presente capítulo revisa os objetivos e perguntas de pesquisa à medida que são apresentadas as principais contribuições, os pontos de originalidade e os resultados da pesquisa. São expostos, por fim, os desafios, limitações, lacunas e possibilidades de avanços em estudos futuros.

O estudo se delineou nas fases exploratória e descritiva, empregando diferentes abordagens, estratégias e métodos a fim de atender aos objetivos propostos e responder às questões de pesquisa formuladas na introdução (Quadro 1), tendo como foco a amostra de pesquisadores vinculados a universidades públicas brasileiras: a) *Como o conhecimento sobre esse tema pode ser organizado e sistematizado de modo a relacionar os fatores que influenciam no patenteamento com os perfis dos pesquisadores que atuam nas universidades? b) Quais fatores influenciam e qual o perfil dos pesquisadores das universidades brasileiras? c) Como identificar as motivações e o perfil dos pesquisadores das universidades?*

VI.1. Revisão dos objetivos e questões de pesquisa

A hipótese que orientou a investigação toma as motivações como determinantes no comportamento do pesquisador, cujo perfil influencia no fazer pesquisa e culmina na decisão em patentear.

A **primeira questão de pesquisa** demandou o levantamento da literatura sobre os temas centrais – patenteamento e motivação - e seus satélites: inovação, propriedade intelectual, relação ciência-indústria, entre outros. O foco está no mapeamento das principais contribuições científicas nas temáticas, visando à organização e sistematização para identificar fatores que motivam o patenteamento e pudessem compor um perfil dos inventores patenteadores vinculados a universidades. Para tanto, caracterizando a fase exploratória do estudo, se empreenderam revisões bibliográficas como o **método de investigação**, a partir de buscas em bases de artigos científicos.

A revisão da literatura permitiu obter o referencial teórico e o panorama do tema nos cenários nacional e internacional. A revisão bibliográfica proveu arcabouço teórico e

permitiu a coleta de conceitos, teorias, abordagens e metodologias. Proveu ainda variáveis e fatores motivacionais que instrumentalizaram e orientaram a elaboração dos instrumentos de pesquisa, notadamente o questionário elaborado que serviu à coleta de dados sociodemográficos e de informações sobre a atividade pantentária e quanto à percepção sobre a importância dos fatores no processo de patenteamento. Da revisão crítica do referencial teórico resultou o atendimento ao **primeiro objetivo específico** na confecção de um *framework* que consolidou os principais achados da revisão e balizou o tratamento dos dados e a identificação dos perfis na fase descritiva.

A representação da estrutura do *framework* é dinâmica e aborda o tema de uma perspectiva multidimensional, como indicado por Olaya-Escobar e Berbegal-Mirabent (2020). Visualmente há a representação sintética em cinco camadas que compõem o perfil, organizadas sucessivamente, contemplando as variáveis nas dimensões ambiental, comportamental e pessoal, permitindo classificar os pesquisadores a partir de uma relação estabelecida com os fatores e variáveis que integram o modelo.

No nível individual, cada pesquisador é classificado em uma categoria por camada. Existem várias possibilidades para diferentes perfis, refletindo a complexidade da natureza e comportamento humano e as especificidades geográficas e institucionais da amostra. Embora a possibilidade de composição de perfis seja grande, o modelo se propõe identificar e agrupar pesquisadores conforme características, interesses e motivações assemelhados.

A classificação se dá em ordem de preponderância, não tendo o condão de diagnóstico único e exclusivo em determinada categoria, mas estabelece uma ordem de preponderância nas categorias com maior identificação. O modelo identifica como um determinado pesquisador ou um grupo de pesquisadores, setores ou instituições são motivados, permitindo ao tomador de decisões desenvolver incentivos personalizados e estabelecer políticas e ações na promoção da inovação, do empreendedorismo acadêmico e de um patenteamento de forma mais assertiva. A identificação de perfis se revela uma importante aliada à prevenção e mediação de conflitos de papéis e atuações na vida acadêmica, bem como pode subsidiar a promoção de uma alocação mais efetiva e satisfatória de recursos humanos a partir das motivações identificadas.

Embora o *framework* tenha como principal foco pesquisadores afiliados às universidades é possível adaptá-lo para aferir perfis de outros grupos e organizações envolvidas em outras etapas do processo de inovação e atividades de pesquisa.

Já dispondo dos 27 fatores motivacionais, classificados por natureza da motivação e resultados esperados, a **segunda questão de pesquisa** ensejou maior precisão na relação desses fatores motivacionais com os perfis dos pesquisadores. Nesta etapa, de natureza predominantemente descritiva, **o método empregado** foi o levantamento de dados pelo acesso às bases de patentes e patenteadores. Devido a sua complexidade e visando maior rigor metodológico foram adotadas ferramentas computacionais como o *Web Scraping,* além da aplicação de análises estatísticas no levantamento e descrição dos dados. O atendimento do segundo objetivo específico se deu na seleção dos mais profícuos inventores de patentes vinculados às 12 universidades públicas brasileiras ranqueadas pela quantidade de depósitos de patentes no INPI, buscando uma amostra mais homogênea no tocante às práticas de patenteamento, sendo possível saber a quais universidades esses pesquisadores estavam vinculados. Neste contexto, um algoritmo na linguagem *Python* foi desenvolvido para aplicação tanto na coleta, recuperação, cruzamento e análise dos dados dos documentos de patentes e currículos dos pesquisadores. Essa amostra de inventores de patentes foi submetida ao questionário que aferiu aspectos pessoais, acadêmicos e motivacionais, consultando diretamente os principais atores no processo de pesquisa e patenteamento.

Os dados oriundos da plataforma *Lattes* e do questionário compõem um *dataset*, sendo possível por meio de tratamento estatístico levantar as principais características, o perfil sociodemográfico e a atribuição de importância dos fatores motivacionais no processo de patenteamento.

Quanto ao **perfil sociodemográfico** da amostra, primeiramente no que tange à carreira, os integrantes apresentam maturidade que, como também identificado por Cativelli e Lucas (2016), é um importante insumo para o desempenho profissional, resultado, como se aduz da amostra, de uma longa e produtiva carreira acadêmica. Dado o universo amostral considerado na pesquisa, os pesquisadores estão vinculados às principais universidades públicas brasileiras, lecionando, orientando e pesquisando principalmente nas áreas biológicas, saúde, engenharias e tecnológicas. Contam com

uma produção científica compatível com sua qualificação, sem concorrer, conforme correlação demonstrada, com os números de envolvimento no processo patentário. Quanto ao gênero, a amostra é majoritariamente masculina, mas a participação de mulheres gradualmente se aproxima de uma participação igualitária, o que se evidencia na parcela mais jovem da amostra em que o gênero feminino se destaca. Quanto à formação acadêmica é realizada predominante em território nacional, com tendência a internacionalização nos níveis mais altos de especialização.

Quanto aos 27 fatores motivacionais identificados, a parcela da amostra que respondeu ao questionário atribuiu a eles uma ordem de importância. Os fatores de natureza intrínseca, ao encontro do que concluíram Blind e Gauch (2008), os vinculados à superação de desafios que demandam soluções técnicas, tiveram os maiores níveis de importância atribuídos. O fator econômico e outros afeitos a objetivos financeiros, por sua vez, foram apontados de menor importância, o que pode se dar por razões culturais e/ou de conhecimento reduzido quanto aos benefícios financeiros que podem decorrer da pesquisa. À medida que houver uma mudança cultural no ambiente das universidades brasileiras e a transferência de tecnologia se tornar um processo consolidado e valorizado, é possível que os aspectos financeiros se tornem um fator motivacional mais presente e atraiam mais pesquisadores para se envolverem em atividades de patenteamento e licenciamento, conforme identificado em alguns estudos realizados em universidades de outros países (D'ESTE e PERKMANN, 2011; VEER *et al*., 2012).

Entretanto, é importante destacar que o impacto dos *royalties* e de outros benefícios financeiros sobre os pesquisadores no desenvolvimento de pesquisas e outros mecanismos de transferência de tecnologia ainda não são bem conhecidos e não pode ser generalizado. Conforme apontado por Baldini (2010) e Blind *et al.*(2018), a literatura sobre o assunto apresenta resultados controversos. Entretanto, os resultados deste estudo indicam que as universidades brasileiras precisam oferecer outros incentivos capazes de atrair mais pesquisadores para atividades de promoção da inovação, com maior interação com as empresas (BALDINI, 2011).

Relevante parcela da amostra é aberta a movimentações geográficas, com notada migração ou concentração para a região Sudeste, principalmente para São Paulo devido a sua centralidade econômica, populacional e de formação e absorção de

pesquisadores. Isto pode apontar, em proporções nacionais, a um fenômeno que Ackers (2005) denomina como fuga de cérebros e Ge *et al.*(2016) tratam como realocação de capital intelectual. A mobilização acadêmica da amostra ocorre ainda para o exterior, parecendo almejar, como também apontado pelo estudo de Bolli e Schläpfer (2015), a melhores recursos para formação e ascensão na carreira.

A **terceira questão de pesquisa** provoca a elaboração de uma aplicação que instrumentalizada pelo levantamento das motivações afeitas ao patenteamento sirva à identificação do perfil motivacional dos pesquisadores. Para tanto o objetivo específico sugere o desenvolvimento de um modelo classificador baseado em soluções computacionais.

O plano é trazer à prática o *framework* teórico, orientado pelas variáveis e fatores motivacionais e alimentado pelo *dataset*. Esta etapa oportuniza a validação do *framework* se valendo do volume de dados dos inventores patenteadores considerando o contexto de universidades públicas brasileiras. Esta tarefa teve dois caminhos possíveis que dependeram diretamente do volume e característica dos dados disponíveis da amostra.

Para a definição do perfil motivacional da parcela da amostra que respondeu ao questionário (n=131), bastou a consolidação e submissão ao *framework* teórico do volume de dados públicos, colecionados da plataforma *Lattes* e corroborado pelas percepções colhidas do questionário, os dados institucionais.

Quanto à parcela restante (n=319), em maior número, e da qual se dispunha apenas dos dados públicos, colhidos do *Lattes* e, considerando que a tarefa de definir perfis caracteriza um problema de classificação, foi empregado como **método** o desenvolvimento de um algoritmo classificador. Baseado em árvore de decisão e desenvolvido em linguagem *Python* à luz do aprendizado de máquina, a modelagem teve como fim realizar a inferência dos perfis à luz do modelo aferido da amostra respondente que serviu de treino ao referido algoritmo de aprendizado.

Avaliações do modelo preditivo do algoritmo foram realizadas a fim de se aferir sua precisão. As taxas de avaliação teriam melhor desempenho se o modelo contasse com um volume de dados maior para sua exposição e treinamento. As taxas obtidas, no entanto, não invalidam o método e muito menos os resultados. Isso de conclui ao se

observar a similaridade entre o perfil obtido pela aplicação estatística a partir dos parâmetros do *framework* com o obtido pela inferência dos dados.

Os resultados do modelo classificador a que foram submetidos a amostra foram demonstrados nas figuras 20 a 22, considerando o balanceamento ou não dos dados, seguidos dos resultados refentes a cada camada correspondente.

Os resultados emergem do volume de informações consolidadas no *dataset*, a que a pesquisa agregou originalidade pela aplicação de ferramentas computacionais no tratamento e extração de conhecimento dos dados e, finalmente, na definição dos perfis baseados nos parâmetros presentes no *framework* teórico. A adoção de tais ferramentas consiste em um fator de ineditismo e contribuição metodológica, ampliando as possibilidades de reprodutibilidade e validação do método e seus resultados.

Quanto à amostra submetida ao modelo classificador, correspondendo à identificação do **perfil motivacional**, tem-se, a partir dos resultados da aplicação do modelo, que os pesquisadores quanto ao tema da relação ciência-indústria, seja os de perfil mais tradicional ou empreendedor, entendem o valor do apoio institucional no processo de pesquisa, mas não percebem o mesmo suporte institucional das universidades públicas na pavimentação do caminho de aproximação com o meio privado. Nesta rota há a percepção de importantes entraves burocráticos, desde o processo de apoio ao patenteamento até atingir o licenciamento ou eventual aproveitamento comercial dos inventos. Os resultados colhidos no processo de pesquisa e no patenteamento vêm sendo incentivados por regulamentações governamentais e fortalecimentos dos NITs, mas, no limite, parecem resultar muito mais de esforços individuais do que de políticas institucionais sustentáveis.

Neste ponto é que o aspecto motivacional é trazido, respondendo à indagação quanto ao que motiva os pesquisadores ao patenteamento. Dentre o que mais motiva os pesquisadores ao patentear suas invenções há proeminência dos fatores intrínsecos, aqueles relacionados ao fazer pesquisa. A superação de desafios dá conta da satisfação pessoal indicada pelos pesquisadores na atividade de pesquisa e patenteamento de suas invenções. Há ainda o reconhecimento e aspiração de eventuais vantagens. Aqui se observa menos os efeitos dos resultados econômico-financeiros, mas há maior sensibilidade quanto ao incremento de prestígio para si, para as equipes e instituições que mantêm vinculo.

Os resultados encontrados sugerem considerar que os pesquisadores que desempenham suas atividades acadêmicas com maior satisfação são mais produtivos, engajados e mais motivados ao patenteamento de invenções de maior impacto social e comercial.

A pesquisa toma as patentes como um resultado objetivo da atividade acadêmica, em cuja base estão as motivações que orientam o esforço intelectual e inventividade do pesquisador. Assim, os resultados apresentados expressam um posicionamento ou uma classificação pontual referente à amostra considerada. A classificação obtida é um retrato de um ponto da realidade dinâmica em que os pesquisadores estão inseridos. Quanto ao modelo proposto, validado na amostra tomada do contexto de universidades públicas brasileiras, este sim, ainda que sujeito a ajustes e atualizações, além de carecer de maior exposição a um maior volume de dados para aprimorar sua precisão, está consolidado enquanto método e definido como uma iniciativa inovadora que reúne os principais e atuais elementos, fatores e variáveis da temática, podendo ser reaplicado nesta e em outras amostras.

VI.2. Originalidade das contribuições

Na esteira de trabalhos similares que tenham abordado temas relacionados à propriedade intelectual, inovação, motivação e empreendedorismo acadêmico, a presente pesquisa avança no conhecimento científico das temáticas citadas na medida em que faz uma revisão da literatura, até então fragmentada, principalmente, quanto à atividade patentária. O estudo serve para ampliar a incipiente literatura sobre atividades de transferência de tecnologia no contexto de economias emergentes como o Brasil.

No que tange ao perfil sociodemográfico e, especificamente, quanto aos aspectos motivacionais afeitos ao patenteamento o trabalho enriquece a literatura acadêmica ao reunir e sintetizar teorias, conceitos, modelos, metodologias e, especialmente, fatores motivacionais, dando centralidade ao tema da aferição dos perfis motivacionais de pesquisadores ligados às atividades de pesquisa e patenteamento.

Quanto aos fatores motivacionais, baseando-se em consulta à literatura, o trabalho agrega contribuição acadêmica ao reunir e classificar 27 fatores motivacionais que instrumentalizaram o questionário que é uma contribuição estratégica adicional, pois afere, entre outras questões, as principais motivações de inventores de patentes.

Para além da entrega de um perfil sociodemográfico da amostra composta pelos mais profícuos patenteadores vinculados às mais bem classificadas universidades públicas nacionais conforme a quantidade de pedidos de patentes, o estudo fornece um retrato atual das principais motivações de pesquisadores brasileiros para patentear, cujos dados estão colecionados no *dataset* que consiste em mais um produdo do estudo, podendo ser utilizado para outras aplicações e interesses de investigação.

Os resultados do presente estudo sugerem que a identificação de perfis é uma questão complexa e orientada por diferentes valores e motivações que se traduzem em comportamentos. Nesse sentido, a partir do mapeamento e relato de um conjunto de fatores motivacionais que se traduz em variáveis de aferição, o trabalho é pioneiro no desenvolvimento de um *framework* que delineia e representa, ao mesmo tempo em que sintetiza e disponibiliza, um modelo para identificação de perfis. O modelo aplica e ilustra de forma clara e facilmente perceptível suas dimensões e elementos estruturantes, permitindo tanto sua compreensão teórica quanto a aferição de perfis com finalidade de estabelecer ou aprimorar políticas públicas de recursos humanos na academia.

O trabalho agrega ao tradicional e eficiente método estatístico, as estratégias, abordagens e notações computacionais, estas especialmente delineadas para a classificação de perfis acadêmicos e profissionais, conferindo maior precisão aos objetivos e ao método de pesquisa. Neste sentido e para o avanço nas discussões propostas é disponibilizado no apêndice A o modelo de questionário de pesquisa de onde se obteve os principais dados qualitativos e quantitativos. Nos apêndices C e D, respectivamente, estão disponibilizados as informações e *links* para repositórios que armazenam duas entregas importantes da pesquisa: o *dataset* e *os* códigos-fonte desenvolvidos em *Python.*

O trabalho, por fim, entrega um modelo classificador estruturado, aberto e dinâmico, baseado em fatores motivacionais e outras variáveis, sendo validado na realidade de uma qualificada amostra de inventores patenteadores brasileiros vinculados a universidades públicas nacionais, podendo ser aprimorado e reproduzido em aplicações similares.

VI.3. Principais desafios e limitações

O insumo principal da pesquisa está no volume de dados referente às patentes, instituições e pesquisadores. A possibilidade de pesquisa em ambiente *web* agrega velocidade e qualidade à coleta, mas traz um desafio quando o conteúdo está disperso e fragmentado. Quanto às instituições com maior número de pedidos de patentes e, finalmente, o acesso às informações cadastrais das patentes a consulta foi realizada na *Derwent Innovation* Index. Neste ponto a pesquisa se alinha com outras inicitivas de investigação, pois se vale dos dados que representam os pedidos de depósitos de patentes e não das patentes concedidas. Isto se deve, principalmente, ao *backlog* do processo que compreende o pedido de depósito, o exame, a emissão de pareceres e, finalmente, a concessão do registro de patente. No limite, a depender do prazo decorrido, que em alguns casos pode superar cinco anos, adotar as patentes concedidas poderia implicar na consideração de tecnologias superadas.

Quanto aos pesquisadores a plataforma de currículos *Lattes* foi consultada. A base foi eleita em função de seu uso abrangente como fonte de informações pelas instituições e pesquisadores brasileiros. No acesso aos dois sítios eletrônicos foi adotado o emprego da técnica *Web Scraping* a fim de automatizar e otimizar a coleta das informações. No caso da *DII* o principal obstáculo se deu na dificultade em identificar os nomes dos inventores nos arquivos de patentes. Isso se deu por conta da interface do arquivo de registro de patente em si, em formato pouco amigável e o emprego de codificações, comuns em arquivos de natureza técnica, mas que afetou principalmente na validação dos nomes que se apresentaram fragmentados e fora do padrão utilizado nas citações utilizadas pelos próprios pesquisadores. Foi necessário fazer ampla consulta a outras bases de dados acadêmicos para que os nomes fossem confirmados.

Quanto à plataforma *Lattes* o *captcha* é um recurso importante de segurança e previne a estabilidade do ambiente virtual, mas impede o acesso direto automatizado, sendo necessário baixar indireta e manualmente cada currículo na extensão XML e então aplicar o *Web Scraping*. Outro ponto relevante, comum em plataformas de livre preenchimento, é a ocorrência de casos de incompletudes e inconsistências nas informações que precisaram ser validadas ou confirmadas em consulta a outros sítios institucionais.

Outro desafio se deu no contato e envio do questionário à amostra de pesquisa. Primeiramente quanto à obtenção dos endereços eletrônicos dos pesquisadores, o que também demandou ampla consulta a fim de coletar pelo menos um endereço valido, preferencialmente institucional. O envio foi realizado em plena pandemia do COVID-19, período em que as instituições tiveram seu funcionamento afetado e muitos pesquisadores afastados de seus centros de trabalhos por razões logísticas e/ou de saúde. Para superar a situação foram realizadas pelo menos cinco ondas de envio, cuja taxa de respondência superou 30%, proporção que poderia ter sido superior em condições mais favoráveis.

Apesar do compartilhamento de todas as informações pertinentes à pesquisa e seus objetivos, bom como termos de compromissos acadêmicos, o acesso ao *link* do questionário via *e-mail* foi reportado como um ponto de preocupação por alguns pesquisadores mais sensíveis às questões de segurança cibernética. A intermediação das instituições neste processo poderia ter aumentado a taxa de respostas na medida em que a validação institucional agregaria confiabilidade ao método de consulta e conteúdo.

Estes fatores foram considerados no planejamento da pesquisa, culminando na definição do corte populacional e de amostra da pesquisa. Outros pontos também foram considerados em função da limitação de recursos humanos, prazos institucionais e outras questões afeitas à dinâmica de um trabalho acadêmico. Oportunamente o estudo pode ter sua amostra ampliada no cenário brasileiro, podendo, inclusive, considerar amostras internacionais a fim de realizar eventuais comparações e testes de cenários.

VI.4. Possibilidades de avanços em estudos futuros

A adoção de ferramentas computacionais nas fases de consulta, estruturação, anonimização, tratamento e mineração de dados permitem a reprodutibilidade da metodologia e apropriação do método aplicado. Tanto o data*set* quanto o modelo computacional cujo código-fonte é disponibilizado com vistas ao potencial de servir de inspiração metodológica para investigações similares. A ideia é que outras fontes de dados possam ser submetidas ao modelo estatístico e ao algoritmo a fim de gerar e comparar perfis.

Neste sentido é necessário ampliar a reflexão, respeitadas as fronteiras legais, quanto à facilitação ao acesso a dados institucionais e acadêmicos que se demonstraram fragmentados, dispersos e despadronizados. Tal cenário impacta na acurácia do modelo desenvolvido e em qualquer outra iniciativa, computacional ou não, que aspire gerar conhecimentos a partir dos dados públicos que são insumos fundamentais ao tipo de investigação desenvolvida. Uma solução pode estar na adoção pelas instituições brasileiras de um formato padronizado de armazenamento e disponibilização dos dados. Um caminho é a interação e convergências com outras plataformas internacionais de consulta e gerenciamento de dados e publicações acadêmicas (*Research gate, ORCID, Linkedin*, etc.) com estrutura e recursos para adesão pela comunidade acadêmica e instituições de pesquisa.

Como desdobramento da metodologia empregada e para agregar validação ao método e resultados se pretende a submissão da pesquisa em suas estratégias, ferramentas e resultados aos NITs e a outras entidades de fomento à pesquisa e inovação com impacto no patentamento. Esta ação visa atender ao escopo inicial de uma pesquisa que se pretende aplicada, permitindo a apropriação e aprimoramento do diagnóstico do perfil sociodemográfico e principalmente motivacional que deriva do *framework* teórico e conferir maior eficiência aos instrumentos de investigação e ao modelo estatístico e computacional proposto.

Aos tomadores de decisões cumpre a oportunidade de considerar a importância e influência dos fatores motivacionais e outras variáveis apurados quando do estabelecimento de políticas públicas e no aprimoramento do processo de inovação e patenteamento com foco no desenvolvimento de pessoas. Também há a ocasião de promover melhorias nas bases de dados de universidades e centros de pesquisa, admitindo e aferindo informações sobre comportamento e motivações para conhecer o perfil de seus pesquisadores e desenvolver políticas e estímulos institucionais personalizados.

O estudo ainda busca oferecer parâmetros, variáveis e conhecimento no tema para o aprimoramento da atuação dos NITs em mediar e promover o avanço das pesquisas acadêmicas no processo de inovação, sobretudo ao considerar e adotar aspectos motivacionais em suas políticas e planos de desenvolvimento de pessoal. Dos NITs ou equivalentes, na esteira do aperfeiçoamento do método de pesquisa, podem vir novas

referências, dados e, principalmente, novos interesses e necessidades que impliquem na ampliação das possibilidades de diagnóstico e processamento oferecidas pelo modelo.

Visando a apropriação e o aprimoramento do modelo está em um futuro próximo o desenvolvimento de uma aplicação de interface amigável e útil para sua adoção. Em seu escopo está buscar facilitar a importação de dados e a geração do diagnóstico de forma interativa e adaptável às necessidades e padrões tanto dos pesquisadores quanto dos gestores públicos nas instituições de pesquisa e universidades.

Um aprofundamento de um tema, tratado aqui de forma adjacente, é o mapeamento e estudo das causas e efeitos da mobilidade acadêmica ou realocação de capital intelectual em âmbito nacional e no exterior. Tal análise pode suscitar reflexões de natureza acadêmica e econômica, com desdobramentos para aspectos regionais na medida em que seja aferido o papel de incentivos privados e públicos neste fluxo e com impacto no desenvolvimento tecnológico e científico, principalmente para as instituições de pesquisa nacionais. Este fenômeno pode apresentar nuances bem interessantes se comparados os cenários do pré e pós-pandemia.

Espera-se, por fim, que a pesquisa sirva de inspiração conceitual e metodológica para sua ampliação e replicação, principalmente do questionário em outras amostras e bases de dados acadêmicos, instituições e regiões, nacionais e internacionais. Isto oportunizaria eventuais comparações de cenários, discussões, aperfeiçoamentos e disseminação de boas práticas, gerando análises similares ampliadas quanto à relação entre os atributos objetivos e subjetivos de pesquisadores em seu desenvolvimento como patenteador.

REFERÊNCIAS

ACKERS, L. Moving people and knowledge: Scientific mobility in the European Union1. **International migration (Geneva, Switzerland)**, v. 43, n. 5, p. 99–131, 2005.

ADRIAANS, P.; ZANTINGE, D. **Data Mining**. Harlow: Addison-Wesley, 158p, 1996,

ALEXANDER, A., MARTIN, D. P., MANOLCHEV, C., MILLER, K.University–industry collaboration: using meta-rules to overcome barriers to knowledge transfer. **The Journal of technology transfer**, v. 45, n. 2, p. 371–392, 2020.

AMADEI, J. R. P.; TORKOMIAN, A. L. V. As patentes nas universidades: análise dos depósitos das universidades públicas paulistas (1995-2006). **Ciência da Informação**, v. 38, n. 2, p. 9–18, 2009.

ANDERSEN, B.; ROSSI, F. Intellectual property governance and knowledge creation in UK universities. **Economics of innovation and new technology**, v. 20, n. 8, p. 701–725, 2011.

ANDRADE, E. P. *et al*. An exploratory analysis of Brazilian universities in the technological innovation process. Technological forecasting and social change, v. 182, n. 121876, p. 121876, 2022.

ASSUMPÇÃO, E. Universidades brasileiras e patentes: utilização do sistema nos anos 90. Diretoria de Articulação–Dart, Centro, *Documentação e Informação Tecnológica–Cedin, Divisão de Estudos e Programas-Diespro,* I*NPI*, Rio de Janeiro, 2000.

BACKS, S.; GÜNTHER, M.; STUMMER, C. Stimulating academic patenting in a university ecosystem: an agent-based simulation approach. **The Journal of technology transfer**, v. 44, n. 2, p. 434–461, 2019.

BALDINI, N. Do royalties really foster university patenting activity? An answer from Italy. Technovation, v. 30, n. 2, p. 109–116, 2010.

BALDINI, N. University patenting: patterns of faculty motivations. **Technology Analysis and Strategic Management**, v. 23, n. 2, p. 103–121, 2011.

BALDINI, N. University patenting: patterns of faculty motivations. Technology Analysis and Strategic Management, v. 23, n. 2, p. 103–121, 2011.

BALDINI, N.; GRIMALDI, R.; SOBRERO, M. To patent or not to patent? A survey of Italian inventors on motivations, incentives, and obstacles to university patenting. **Scientometrics**, v. 70, n. 2, p. 333–354, 2007.

BARBASTEFANO, R. G.; DE SOUZA, C. G.; DE ARAUJO, F. O. **UNIVERSITÁRIO NO BRASIL**. **..** *Interação universidade-empresa: análise de padrões de depósitos conjuntos e desafios ao patenteamento universitário no brasil*. In: XXX ENEGEP, 2010, São Carlos (SP). Anais do ENEGEP. Rio de Janeiro: ABEPRO. v. 1. p. 1-12 Disponível em: <http://www.abepro.org.br/biblioteca/enegep2010_TN_STO_120_780_17252.pdf>. Acesso em: 21 jul. 2023.

BARBIERI, J. C. (1999). Os inventores no Brasil: tipos e modalidades de incentivos. *Revista de Administração de Empresas*, v. 39, n. 2, p. 54-63. http://www.scielo.br/pdf/rae/v39n2/v39n2a07.pdf.

BARBOSA, V. P. Inovação: o ingrediente que desafia as universidades. Disponível em: <https://www.inovacao.usp.br/inovacao-o-ingrediente-que-desafia-as-universidades/.>. Acesso em: 24 jul. 2023.

BEAUDRY, C.; KANANIAN, R. Follow the (industry) money – the impact of science networks and industry-to-university contracts on academic patenting in nanotechnology and biotechnology. **Industry and innovation**, v. 20, n. 3, p. 241–260, 2013.

BENTLEY, P. J.; GULBRANDSEN, M.; KYVIK, S. The relationship between basic and applied research in universities. **Higher education**, v. 70, n. 4, p. 689–709, 2015.

BLIND, K.; GAUCH, S. Research and standardisation in nanotechnology: evidence from Germany. **The Journal of technology transfer**, v. 34, n. 3, p. 320–342, 2009.

BLIND, K.; POHLISCH, J.; ZI, A. Publishing, patenting, and standardization: Motives and barriers of scientists. **Research policy**, v. 47, n. 7, p. 1185–1197, 2018.

BODAS FREITAS, I. M.; MARQUES, R. A.; SILVA, E. M. DE P. E. University–industry collaboration and innovation in emergent and mature industries in new industrialized countries. Research policy, v. 42, n. 2, p. 443–453, 2013.

BOGDAN, R.C.; BIKLEN, S. **Investigação qualitativa em educação – uma introdução à teoria e aos métodos.** Porto: Editora Porto, 2010.

BOH, W. F.; EVARISTO, R.; OUDERKIRK, A. Balancing breadth and depth of expertise for innovation: A 3M story. **Research policy**, v. 43, n. 2, p. 349–366, 2014.

BOLLI, T.; SCHLÄPFER, J. Job mobility, peer effects, and research productivity in economics. **Scientometrics**, v. 104, n. 3, p. 629–650, 2015.

BRACHMAN, R.; ANAND, T. The Process of Knowledge Discovery in Databases. In: FAYYAD, U. *et al*. **Advances in Knowledge Discovery and Data Mining**. Menlo Park, CA: AAAI Press. p.37-57, 1996.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Resumo Técnico: Censo Escolar da Educação Básica, 2021. Brasília, DF: Inep (2021). https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/resumo_tecnico_censo_escolar_2021.pdf

BRASIL. Lei nº 10.973, de 2 de Dezembro de 2004. Dispõe sobre incentivos à inovação e à pesquisa científica e tecnológica no ambiente produtivo e dá outras providências. *Diário Oficial da União,* 3 de dezembro de 2004, p. 2.

BURHAN, M.; SINGH, A. K.; JAIN, S. K. Patents as proxy for measuring innovations: A case of changing patent filing behavior in Indian public funded research organizations. **Technological forecasting and social change**, v. 123, p. 181–190, 2017.

CABENA, P., HADJINIAN, P., STADLER, R., VERHEES, J.eZANASI. A. **Discovering Data Mining from Concept to Implementation.** Upper Saddle River, New Jersey: Prentice Hall,. 195p,1997.

CASTRO, A. C.; JANNUZZI, C. A. S. C.; MATTOS, F. A. M. de. Produção e disseminação de informação tecnológica: a atuação da Inova - Agência de Inovação da UNICAMP. *Transinformação*, Campinas, v. 19, n. 3, p. 265-277, set./dez, 2007.

Castro, Leandro Nunes de; Ferrari, Daniel Gomes. **Introdução à Mineração de Dados: conceitos básicos, algoritmos e aplicações.** 1ª Ed. São Paulo: Saraiva, 2016.

CATIVELLI, A. S.; LUCAS, E. DE O. Patentes universitárias brasileiras: perfil dos inventores e produção por área do conhecimento. **Encontros Bibli Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 21, n. 47, p. 67–81, 2016.

CHANG, X.-H.; CHEN, Q.; FONG, P. S. W. University invention disclosure: balancing the optimal stage and type. **The Journal of technology transfer**, v. 42, n. 3, p. 510–537, 2017.

CHESBROUGH, H. W. **Open innovation: The new imperative for creating and profiting from technology.** Harvard Business Press, 2003.

CLOSS, L., FERREIRA, G., BRASIL, V., SAMPAIO, C., PERIN, M. What motivates Brazilian academic researchers to transfer technology? Journal of technology managementeinnovation, v. 8, n. 4, p. 13–14, 2013.

COMPAGNUCCI, L.; SPIGARELLI, F. The Third Mission of the university: A systematic literature review on potentials and constraints. **Technological forecasting and social change**, v. 161, n. 120284, p. 120284, 2020.

CORMEN, Thomas H.; LEISERSON, Charles E.; RIVEST, Ronald L.; STEIN, Clifford. **Algoritmos: Teoria e Prática.** 3rd ed. Rio de Janeiro : Elsevier., 2012.

CRESPO DE BARROS, W.; GOMES DE SOUZA, C. Patent documents as a source of technological information in Brazil: A user study of the program of automatic supply of technological information (PROFINT). **Information development**, v. 26, n. 1, p. 67–78, 2010.

CROITORU, A. SCHUMPETER, J.A., 1934. The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest ethe Business Cycle, translated from the German by Redvers Opie News Brunswick (U.S.A) eLondon (U.K.): Transaction Publishers. *Journal of Comparative Research in*

Anthropology eSociology, 3(2),137–148, 2012. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1496199.

CRUZ, H. N., Souza, R. F. Sistema Nacional De Inovação E a Lei DDa Inovação: Análise Comparativa Entre O Bahy-Dole Act E a Lei Da Inovação Tecnológica. *Review of Administration eInnovation - RAI*, 11(4), 329, 2015. http://www.revistas.usp.br/rai/article/view/110254.

D'ESTE, P.; IAMMARINO, S. The spatial profile of university-business research partnerships: The spatial profile of u-b research partnerships. **Papers in regional science: the journal of the Regional Science Association International**, v. 89, n. 2, p. 335–350, 2010.

D'ESTE, P.; PERKMANN, M. Why do academics engage with industry? The entrepreneurial university and individual motivations. **The Journal of technology transfer**, v. 36, n. 3, p. 316–339, 2011.

DA SILVA, R O.; SILVA, I. R. S. **Linguagem de Programação Python.** Tecnologias em Projeção, v. 10, n. 1, p. 55-71, 2019.

DANIEL, A. D.; ALVES, L. University-industry technology transfer: the commercialization of university's patents. **Knowledge management researchepractice**, v. 18, n. 3, p. 276–296, 2020.

DE MOURA, A. M. M. Motivação para a pesquisa, determinação de parcerias e divisão da coautoria e coinvenção: principais critérios utilizados pelos pesquisadores da área da Biotecnologia. **Em Questão**, v. 18, n. 3, p. 29–45, 2012.

DE SOUZA, C. G.; FERREIRA, M. L. A. Researchers profile, co-authorship pattern and knowledge organization in information science in Brazil. **Scientometrics**, v. 95, n. 2, p. 673–687, 2013.

DECI, E. L.; RYAN, R. M. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. **Psychological inquiry**, v. 11, n. 4, p. 227–268, 2000.

DEGL'INNOCENTI, M.; MATOUSEK, R.; TZEREMES, N. G. The interconnections of academic research and universities' "third mission": Evidence from the UK. **Research policy**, v. 48, n. 9, p. 103793, 2019.

DELIBASIC, B. *et al*. White-box or black-box decision tree algorithms: Which to use in education? **IEEE transactions on education**, v. 56, n. 3, p. 287–291, 2013.

DIOUF, R., SARR, E. N., SALL, O., BIRREGAH, B., BOUSSO, M., MBAYE, S. N. **Web Scraping: State-of-the-art and areas of application**. 2019 IEEE International Conference on Big Data (Big Data). **Anais**...IEEE, 2019.

EJERMO, O.; KÄLLSTRÖM, J. What is the causal effect of R&D on patenting activity in a professor's privilege country? Evidence from Sweden. **Papers in Innovation Studies**, 2015.

ERIKSON, T.; KNOCKAERT, M.; FOO, M. D. Enterprising scientists: The shaping role of norms, experience and scientific productivity. **Technological forecasting and social change**, v. 99, p. 211–221, 2015.

ETZKOWITZ, H.; ZHOU, C. Hélice Tríplice: inovação e empreendedorismo universidade-indústria-governo. **Estudos Avançados**, v. 31, n. 90, p. 23–48, 2017.

FAWCETT, T.; PROVOST, F. **Adaptive fraud detection. Data mining and knowledge** discovery, v. 1, n. 3, p. 291-316, 1997.

FAYYAD, U., Piatetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37-37. 1996a. https://doi.org/10.1609/aimag.v17i3.1230

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, n. 11, p. 27–34, 1996b.

FISCHER, B. B.; SCHAEFFER, P. R.; VONORTAS, N. S. Evolution of university-industry collaboration in Brazil from a technology upgrading perspective. Technological forecasting and social change, v. 145, p. 330–340, 2019.

FREEL, M.; PERSAUD, A.; CHAMBERLIN, T. Faculty ideals and universities' third mission. **Technological forecasting and social change**, v. 147, p. 10–21, 2019.

FREITAS, A. Generic, Set-Oriented **Primitives to Support Data-Parallel: Knowledge Discovery in Relational Database Systems.** Thesis, UK: University of Essex. 275 p, 1997.

GAGNÉ, M.; DECI, E. L. Self-determination theory and work motivation: SELF-DETERMINATION THEORY AND WORK MOTIVATION. **Journal of organizational behavior**, v. 26, n. 4, p. 331–362, 2005.

GAO, X.; GUO, X.; GUAN, J. An analysis of the patenting activities and collaboration among industry-university-research institutes in the Chinese ICT sector. **Scientometrics**, v. 98, n. 1, p. 247–263, 2014.

GARCEZ JÚNIOR, S. S.; MOREIRA, J. DE J. DA S. O backlog de patentes no Brasil: o direito à razoável duração do procedimento administrativo. Revista Direito GV, v. 13, n. 1, p. 171–203, 2017.

GARCÍA, S., RAMÍREZ-GALLEGO, S., LUENGO, J., BENÍTEZ, J. M., HERRERA, F. Big data preprocessing: methods and prospects. **Big data analytics**, v. 1, n. 1, 2016.

GARCÍA, Salvador; LUENGO, Julián; HERRERA, Francisco. **Data preprocessing in data mining**. Cham, Switzerland: Springer International Publishing, 2015.

GARNICA, L. A.; TORKOMIAN, A. L. V. Gestão de tecnologia em universidades: uma análise do patenteamento e dos fatores de dificuldade e de apoio à transferência de tecnologia no Estado de São Paulo. **Gestãoeprodução**, v. 16, n. 4, p. 624–638, 2009.

GE, C.; HUANG, K.-W.; PNG, I. P. L. Engineer/scientist careers: Patents, online profiles, and misclassification bias: Online Career Profiles and Misclassification. **Strategic management journal**, v. 37, n. 1, p. 232–253, 2016.

GENTZKOW, Matthew; Shapiro, Jesse M. **Code and data for the social sciences: A practitioner's guide.** Chicago, IL: University of Chicago, 2014.

GEUNA, A.; ROSSI, F. Changes to university IPR regulations in Europe and the impact on academic patenting. **Research policy**, v. 40, n. 8, p. 1068–1076, 2011.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social.** 6ª ed. São Paulo: Atlas, 2011.

GÖKTEPE-HULTÉN D. Inside the Ivory Tower- Inventors and Patents at Lund University. CIRCLE, Lund University. 266 p.,2008. https://vtechworks.lib.vt.edu/handle/10919/71615.

GÖKTEPE-HULTEN, D.; MAHAGAONKAR, P. Inventing and patenting activities of scientists: in the expectation of money or reputation? **The Journal of technology transfer**, v. 35, n. 4, p. 401–423, 2010.

GRIMALDI, R. *et al*. 30 years after Bayh–Dole: Reassessing academic entrepreneurship. **Research policy**, v. 40, n. 8, p. 1045–1057, 2011.

GUSMÃO, R. Práticas e Políticas Internacionais de Colaboração Ciência-Indústria. **Revista Brasileira de Inovação**, v. 1, n. 2, p. 327, 2009.

HAASE, H.; ARAÚJO, E. C. DE; DIAS, J. Inovações Vistas pelas Patentes: exigências frente às novas funções das universidade. **Revista Brasileira de Inovação**, v. 4, n. 2, p. 329, 2009.

HALILEM, N., AMARA, N., OLMOS-PEÑUELA, J., MOHIUDDIN, M. "To Own, or not to Own?" A multilevel analysis of intellectual property right policies' on academic entrepreneurship. **Research policy**, v. 46, n. 8, p. 1479–1489, 2017.

HARRISON, J.; TUROK, I. Universities, knowledge and regional development. **Regional studies**, v. 51, n. 7, p. 977–981, 2017.

HAYTER, C. S.; FEENEY, M. K. Determinants of external patenting behavior among university scientists. **Scienceepublic policy**, p. scw037, 2016.

HENDERSON, R.; JAFFE, A. B.; TRAJTENBERG, M. Universities as a source of commercial technology: A detailed analysis of university patenting, 1965–1988. **The Review of Economics and Statistics**, v. 80, n. 1, p. 119–127, 1998.

INDUSTRIAL, P.; FEDERAL, U.; GERAIS, M. **Protagonismo incomum**. p. 2–5, 2015.

Inovação Revista Eletrônica de P, DeI. Indústria 4.0: como a tecnologia fez surgir a Quarta Revolução Industrial. **Equipe Revista Inovação** em 16 de agosto de 2016 Inova / Labjor / Unicamp, Campinas. ISSN: 2359-4691, 2016.

INPI – Instituto Nacional da Propriedade Industrial. Rankings de depositantes residentes, 2021. https://www.gov.br/inpi/pt-br/central-de-conteudo/estatisticas/arquivos/estatisticas-preliminares/rankdepositantesresidentes-2020.pdf

Instituto Nacional da Propriedade Industrial (INPI**). Indicadores de Propriedade Industrial 2017: O uso do sistema de propriedade industrial no Brasil.** , p. 78, 2018.

JENSEN, R.; THURSBY, M. Proofs and prototypes for sale: The licensing of university inventions. American Economic Review, v. 91, n. 1, p. 240–259, 2001.

JESUS, C. S. ; SOUZA, C. G. ; ASSIS, L. S. ; JUNIOR, J. R. P. ; Cardoso, D. O. (2020). *Web Scraping para Coleta de Dados nos Anais do ENEGEP*. *Encontro Nacional de Engenharia de Produção.* http://www.abepro.org.br/biblioteca/TN_STO_347_1785_40717.pdf

JONGBLOED, B.; ENDERS, J.; SALERNO, C. Higher education and its communities: Interconnections, interdependencies and a research agenda. **Higher education**, v. 56, n. 3, p. 303–324, 2008.

KADIYALA, A.; KUMAR, A. Applications of python to evaluate the performance of decision tree-based boosting algorithms. **Environmental progressesustainable energy**, v. 37, n. 2, p. 618–623, 2018.

KOTSIANTIS, S. B.; KANELLOPOULOS, D.; PINTELAS, P. E. Data Preprocessing For Supervised Leaning. 2007.

KOVACS, Z. L. **Redes neurais artificiais: fundamentos e aplicações.** 2.ed. São Paulo: Colledium Cognitio. 174p, 1996.

KRAWCZYK, B. Learning from imbalanced data: open challenges and future directions. **Progress in artificial intelligence**, v. 5, n. 4, p. 221–232, 2016.

LAM, A. What motivates academic scientists to engage in research commercialisation: “gold,” “ribbon” or “puzzle”? **SSRN Electronic Journal**, 2010.

LAWSON, C.; STERZI, V. The role of early-career factors in the formation of serial academic inventors. **Scienceepublic policy**, v. 41, n. 4, p. 464–479, 2014.

LEITE, R. A. S. *et al*. Portfolio evaluation of academic patent: A proposal to Brazil. Journal of technology managementeinnovation, v. 14, n. 4, p. 66–77, 2019.

LEWENSOHN, D. *et al*. Applying patent survival analysis in the academic context. **Research evaluation**, v. 24, n. 2, p. 197–212, 2015.

LIAW, Y. C., CHAN, T. Y., FAN, C. Y., CHIANG, C. H. Can the technological impact of academic journals be evaluated? The practice of non-patent reference (NPR) analysis. **Scientometrics**, v. 101, n. 1, p. 17–37, 2014.

LIPTON, ZACHARY C.; ELKAN, CHARLES; NARYANASWAMY, BALAKRISHNAN. Optimal thresholding of classifiers to maximize F1 measure. In: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD, Nancy, France, September 15-19, 2014. Proceedings, Part II 14. Springer Berlin Heidelberg, 2014. p. 225-239, 2014.

LJUNGBERG, D.; BOURELOS, E.; MCKELVEY, M. Academic inventors, technological profiles and patent value: An analysis of academic patents owned by Swedish-based firms. **Industry and innovation**, v. 20, n. 5, p. 473–487, 2013.

LOMACHYNSKA, I. *et al*. Innovation potential: Impact on the national economy’s competitiveness of the Eu developed countries. **Baltic Journal of Economic Studies**, v. 4, n. 1, p. 262–270, 2018.

LOWE, R. A. Who develops a university invention? The impact of tacit knowledge and licensing policies. **The Journal of technology transfer**, v. 31, n. 4, p. 415–429, 2006.

MACHADO MENDES, L.; AMORIM-BORHER, B.; LAGE, C. Patent applications on representative sectors of biotechnology in Brazil: An analysis of the last decade. **Journal of technology managementeinnovation**, v. 8, n. 4, p. 15–16, 2013.

MARCONI, M. A., LAKATOS, E. M. **Fundamentos de metodologia científica.** 7. Ed. São Paulo: Editora Atlas, 2010.

MARTINS, G.A. ***Estudo de Caso*: uma estratégia de pesquisa.** São Paulo: Atlas, 2006.

MARTINS, T. G. DOS S.; SCHOR, P. Unpacking the black box. **Einstein (Sao Paulo, Brazil)**, v. 19, 2020.

MAULEON, E.; DARAIO, C.; BORDONS, M. Exploring gender differences in patenting in Spain. **Research evaluation**, v. 23, n. 1, p. 62–78, 2014.

MCT&I - Ministério de Ciência, TecnologiaeInovação. Indicadores Nacionais de Ciência, Tecnologia e Inovação 2022, 2023. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/indicadores/paginas/publicacoes/arquivos/indicadores_cti_2022.pdf

MERTENS, U. K.; VOSS, A.; RADEV, S. ABrox—A user-friendly Python module for approximate Bayesian computation with a focus on model comparison. **PloS one**, v. 13, n. 3, p. e0193981, 2018.

MILLER, M. **Visual Analytics of Spatio-Temporal Event Predictions: Investigating Causes for Urban Heat Islands**, 2018.

MOURA, A. M. M. (2012). Motivação para a pesquisa, determinação de parcerias e divisão da coautoria e coinvenção: principais critérios utilizados pelos pesquisadores da área da Biotecnologia. *Em Questão*, v. 18, p. 29-45. https://seer.ufrgs.br/EmQuestao/article/view/33025

MOUTINHO, P. S. F.; FONTES, M.; GODINHO, M. M. Do individual factors matter? A survey of scientists' patenting in Portuguese public research organisations. **Scientometrics**, v. 70, n. 2, p. 355–377, 2007.

MOUTINHO, R., AU-YONG-OLIVEIRA, M., COELHO, A., MANSO, J. P. Determinants of knowledge-based entrepreneurship: an exploratory approach. **International entrepreneurship and management journal**, v. 12, n. 1, p. 171–197, 2016.

MOWERY, D. C., NELSON, R. R., SAMPAT, B. N., ZIEDONIS, A. A. The growth of patenting and licensing by U.S. universities: an assessment of the effects of the Bayh–Dole act of 1980. **Research policy**, v. 30, n. 1, p. 99–119, 2001.

MOWERY, D. C.; ZIEDONIS, A. A. Academic patent quality and quantity before and after the Bayh–Dole act in the United States. **Research policy**, v. 31, n. 3, p. 399–418, 2002.

MUELLER, S. P. M.; PERUCCHI, V. Universidades e a produção de patentes: tópicos de interesse para o estudioso da informação tecnológica. **Perspectivas em Ciência da Informação**, v. 19, n. 2, p. 15–36, 2014.

NATALICCHIO, A.; MESSENI PETRUZZELLI, A.; GARAVELLI, A. C. The impact of partners' technological diversification in joint patenting: A study on firm-PRO collaborations. **Management decision**, v. 55, n. 6, p. 1248–1264, 2017.

NICOTRA, M.; DEL GIUDICE, M.; ROMANO, M. Fulfilling University third mission: towards an ecosystemic strategy of entrepreneurship education. **Studies in higher education**, v. 46, n. 5, p. 1000–1010, 2021.

NIMER, F.; SPANDRI, L.C. **Obtendo vantagens competitiva com o uso de data mining.** Developers Magazine, Ano 2, n.18, p.32, 1998.

OLAYA-ESCOBAR, E. S.; BERBEGAL-MIRABENT, J.; ALEGRE, I. Exploring the relationship between service quality of technology transfer offices and researchers' patenting activity. **Technological forecasting and social change**, v. 157, n. 120097, p. 120097, 2020.

OLIVEIRA, J. H. P. DE *et al*. Motivation for the development of patents in universities in the state of Pernambuco, Brazil. **Gestãoeprodução**, v. 27, n. 4, 2020.

OLIVEIRA, R. M. D. **Proteção e comercialização da pesquisa acadêmica no Brasil: motivações e percepções dos inventores.** 2011. 167 f (Doctoral dissertation, Tese (Doutorado) - Curso de Política Científica e Tecnológica, Instituto de Geociências, Universidade Estadual de Campinas, Campinas), 2011.

OLIVEIRA, R. M. de. ; VELHO, L. M. L. S. Patentes acadêmicas no Brasil: uma análise sobre as universidades públicas paulistas e seus inventores. *Parcerias Estratégicas*, v. 14, n. 29, p. 173-200, 2010. http://seer.cgee.org.br/index.php/parcerias_estrategicas/article/view/355

OLIVER, A. L.; SAPIR, A. Shifts in the organization and profession of academic science: the impact of IPR and technology transfer. **Journal of professions and organization**, p. jow012, 2017.

OUELLETTE, L. L.; TUTT, A. How do patent incentives affect university researchers? **International review of law and economics**, v. 61, n. 105883, p. 105883, 2020.

OWAN, H.; NAGAOKA, S. S. **Intrinsic and extrinsic motivation for inventors.** RIETI *discussion Paper Series* 11-E-022, 2011.

OWEN-SMITH, J.; POWELL, W. W. **The Journal of technology transfer**, v. 26, n. 1/2, p. 99–114, 2001.

PARANHOS, J.; CATALDO, B.; PINTO, A. C. DE A. CRIAÇÃO, INSTITUCIONALIZAÇÃO E FUNCIONAMENTO DOS NÚCLEOS DE INOVAÇÃO TECNOLÓGICA NO BRASIL: CARACTERÍSTICAS E DESAFIOS. REAd, v. 24, n. 2, p. 253–280, 2018.

PEREIRA, A.C.S. (2019) **Prospecção tecnológica em patentes sobre tecnologias militares. Dissertação** (Mestrado Profissional em Propriedade Intelectual e Inovação) – Instituto Nacional da Propriedade Industrial – INPI, Rio de Janeiro, 2019. https://www.gov.br/inpi/pt-br/servicos/academia/arquivo/arquivos-biblioteca/pereira-ana-carolina-de-souza.pdf

PERKMANN, M. *et al*. Academic engagement: A review of the literature 2011-2019. **Research policy**, v. 50, n. 1, p. 104114, 2021.

PERUCCHI, V.; MUELLER, S. P. M. Estudo com as patentes produzidas e o perfil dos inventores dos Institutos Federais de Educação, Ciência e Tecnologia. **RDBCI Revista Digital de Biblioteconomia e Ciência da Informação**, v. 12, n. 1, p. 191, 2014.

PESQUISA FAPESP. (Revista Eletrônica. **Protagonismo incomum**. Revista Pesquisa Fapesp**.** Ed. 249 de novembro de 2016. São Paulo. ISSN: 1519-8774, 2016. https://revistapesquisa.fapesp.br/wp-content/uploads/2016/11/042-045_Patentes_249.pdf

PESQUISA FAPESP. Revista Eletrônica. Soluções Microscópicas. Revista Pesquisa Fapesp**.** Ed. 257 de julho de 2017. São Paulo. ISSN: 1519-8774, 2017. https://revistapesquisa.fapesp.br/2017/07/18/folheie-a-edicao-257/ Acesso em 26 de novembro de 2019.

PHELAN, T. J. Striking the mother lode in science: The importance of age, place, and time. **The journal of higher education**, v. 65, n. 5, p. 627–629, 1994.

PHILPOTT, K. *et al*. The entrepreneurial university: Examining the underlying academic tensions. **Technovation**, v. 31, n. 4, p. 161–170, 2011.

PINHEIRO, I. A. **Perfil, motivações e demandas de uma amostra de inventores e empreendedores.** Read: Revista Eletrônica de Administração. 7(5), 1-14. 2001. https://seer.ufrgs.br/read/article/view/73231

PINTO, E. B.; FERNANDES, G. Collaborative R&D the key cooperation domain for university-industry partnerships sustainability – position paper. **Procedia computer science**, v. 181, p. 102–109, 2021.

PÓVOA, L. M. C. A universidade deve patentear suas invenções? **Revista Brasileira de Inovação**, v. 9, n. 2, p. 231, 2010.

PÓVOA, L. M. C.; RAPINI, M. S. Technology transfer from universities and public research institutes to firms in Brazil: what is transferred and how the transfer is carried out. **Scienceepublic policy**, v. 37, n. 2, p. 147–159, 2010.

QUERIDO, A. L. DE S.; SALGUEIRO LAGE, C. L.; GUIMARÃES VASCONCELLOS, A. What is the Destiny of patents of Brazilian universities? Journal of technology managementeinnovation, v. 6, n. 1, p. 46–57, 2011.

QUERIDO, A. L. S. (2011). **Destino das patentes das universidades brasileiras e mapeamento das atividades dos Núcleos de Inovação Tecnológica.** Rio de Janeiro, 2011. 147f. Tese (Doutorado em Biotecnologia Vegetal) - Centro de Ciências da Saúde - Universidade Federal do Rio de Janeiro, Rio de Janeiro. http://fenix3.ufrj.br/50/teses/d/CCS_D_AndreLuizDeSouzaQuerido.pdf

QUINLAN, J. **C4.5: Programs for machine learning.** San Mateo: Morgan Kaufmann. 302p, 1993.

QUINLAN, J. R. **Induction of decision trees**. Mach. Learn., 1 (1):81–106, Mar. ISSN 0885-6125. doi: 10.1023/A: 1022643204877, 1986.

RAFFERTY, M. The Bayh–Dole Act and university research and development. **Research policy**, v. 37, n. 1, p. 29–40, 2008.

RAMOS-VIELBA, I.; FERNÁNDEZ-ESQUINAS, M. Beneath the tip of the iceberg: exploring the multiple forms of university–industry linkages. **Higher education**, v. 64, n. 2, p. 237–265, 2012.

RAPINI, M. S. Interação universidade-empresa no Brasil: evidências do Diretório dos Grupos de Pesquisa do CNPq. Estudos economicos, v. 37, n. 1, p. 211–233, 2007.

RAUEN, C. V. O Novo marco legal da inovação no Brasil : o que muda na relação ICT-empresa? 2016.

RAYCHEV, V.; BIELIK, P.; VECHEV, M. Probabilistic model for code with decision trees. **SIGPLAN notices**, v. 51, n. 10, p. 731–747, 2016.

REVISTA BRASILEIRA DE INOVAÇÃO. Rio de Janeiro, v. 9, n. 2, p. 231-256, jul./dez.2010.

ROSE, M.; KITCHIN, J. R. scopus: Scriptable Bibliometrics Using a Python Interface to Scopus. **SSRN Electronic Journal**, 2019.

ROTHAERMEL, F. T.; AGUNG, S. D.; JIANG, L. University entrepreneurship: a taxonomy of the literature. **Industrial and corporate change**, v. 16, n. 4, p. 691–791, 2007.

RUSSO SERAFINI, M. *et al*. CARACTERÍSTICAS DA PROPRIEDADE INTELECTUAL NO NORDESTE ATRAVÉS DE SITES DE BUSCAS TECNOLÓGICAS. **Revista Gestão Inovação e Tecnologias**, v. 1, n. 1, p. 01–11, 2011.

RYAN, J. C. The work motivation of research scientists and its effect on research performance: Work motivation of research scientists. **R and D Management**, v. 44, n. 4, p. 355–369, 2014.

RYAN, J. C.; BERBEGAL-MIRABENT, J. Motivational recipes and research performance: A fuzzy set analysis of the motivational profile of high performing research scientists. **Journal of business research**, v. 69, n. 11, p. 5299–5304, 2016.

RYAN, R. M.; DECI, E. L. Intrinsic and extrinsic motivations: Classic definitions and new directions. **Contemporary educational psychology**, v. 25, n. 1, p. 54–67, 2000b.

RYAN, R. M.; DECI, E. L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. **The American psychologist**, v. 55, n. 1, p. 68–78, 2000a.

SAMPAIO, R. F.; MANCINI, M. C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Brazilian journal of physical therapy**, v. 11, n. 1, p. 83–89, 2007.

SAMPAT, B. N.; MOWERY, D. C.; ZIEDONIS, A. A. Changes in university patent quality after the Bayh–Dole act: a re-examination. **International journal of industrial organization**, v. 21, n. 9, p. 1371–1390, 2003.

SAUERMANN, H. *et al*. **Nber working paper series what makes them tick? Employee motives and firm innovation**. *Management Science*, v. 56, n. 12, p. 2134-2153 Disponível em: <https://www.nber.org/papers/w14443.pdf>. Acesso em: 21 jul. 2023.

SAURKAR, A. V.; PATHARE, K. G.; GODE, S. A. **An Overview On Web Scraping Techniques And Tools.** International Journal on Future Revolution in Computer ScienceeCommunication Engineering, V. 4, p. 363-367, 2018.

SCHMID, J.; WANG, F.-L. Beyond national innovation systems: Incentives and China's innovation performance. **Dang dai Zhongguo [The journal of contemporary China]**, v. 26, n. 104, p. 280–296, 2017.

SEVERINO, A. J. **Metodologia do trabalho científico.** São Paulo: Cortez, 2007.

SIEGEL, D. S.; VEUGELERS, R.; WRIGHT, M. Technology transfer offices and commercialization of university intellectual property: performance and policy implications. **Oxford review of economic policy**, v. 23, n. 4, p. 640–660, 2007.

JESUS, C. S. ,SOUZA, C. G., ASSIS, L. S., JUNIOR, JURAIR R. P., CARDOSO, D. O **WEB SCRAPING PARA COLETA DE DADOS NOS ANAIS DO ENEGEP. Contribuições da Engenharia de Produção para a Gestão de Operações Energéticas Sustentáveis**. Disponível em: <http://www.abepro.org.br/biblioteca/TN_STO_347_1785_40717.pdf>. Acesso em: 21 jul. 2023.

SILVA, K. *et al*. Academic patenting in Brazil: the role of academic inventors in PCT patent applications – 2002-2012. **Academia Revista Latinoamericana de Administración**, v. 30, n. 4, p. 529–546, 2017.

SINGH, A.; WONG, P.-K.; HO, Y.-P. The role of universities in the national innovation systems of China and the East Asian NIEs: An exploratory analysis of publications and patenting data. **Asian journal of technology innovation**, v. 23, n. 2, p. 140–156, 2015.

SIVAKUMAR, A.; GUNASUNDARI, R. A Survey on Data Preprocessing Techniques for Bioinformatics and Web Usage Mining. *International Journal of Pure and Applied Mathematics*, v. 117, n. 20, p. 785-794, 2017.

SJÖÖ, K.; HELLSTRÖM, T. University–industry collaboration: A literature review and synthesis. **Industry and higher education**, v. 33, n. 4, p. 275–285, 2019.

SKUTE, I. *et al*. Mapping the field: a bibliometric analysis of the literature on university–industry collaborations. **The Journal of technology transfer**, v. 44, n. 3, p. 916–947, 2019.

SOUZA C. G; SARMENTO, I. M. F. ; XAVIER, L. DE S.; Aguiar, R. A. A. Conhecimento sobre Patentes na Educação em Engenharia: uma experiência metodológica. *Revista de Ensino de Engenharia*, v. 25, p. 27-35, 2006. http://www.abenge.org.br/cobenge/legado/arquivos/14/artigos/RJ-9-98473000706-1118677972614.pdf

SOUZA, C. G. de, BARBASTEFANO, R. G., PEREIRA, F. de C. (2011). Estudo sobre o perfil dos inventores das patentes de universidades brasileiras. In: *Encontro Nacional de Engenharia de Producao*, 31., 2011, Belo Horizonte. Anais... Belo Horizonte: ABREPO. p. 1-12. Disponível em: http://www.abepro.org.br/.../enegep2011_TN_STO_142.pdf.

SOUZA, C. G.; FERREIRA, M. L. A. (2013). Researchers profile, co-authorship pattern and knowledge organization in information science in Brazil. Scientometrics (Print), v. 95, p. 673-687.

SOUZA, C.G.; BARBASTEFANO, R. G., PEREIRA, F.C. Estudo sobre o perfil dos inventores das patentes de universidades brasileiras. In: *Encontro Nacional de Engenharia de Produção*, 2011, Belo Horizonte. Anais do ENEGEP. Rio de Janeiro: ABEPRO. v. 1. p. 1-10, 2011. https://abepro.org.br/biblioteca/enegep2011_TN_STO_142_896_18859.pdf

STEK, P. E.; VAN GEENHUIZEN, M. S. Measuring the dynamics of an innovation system using patent data: a case study of South Korea, 2001–2010. **Qualityequantity**, v. 49, n. 4, p. 1325–1343, 2015.

STEPHAN, P. E. *et al*. Who's patenting in the university? Evidence from the survey of doctorate recipients. **Economics of innovation and new technology**, v. 16, n. 2, p. 71–99, 2007.

STEPHAN, P. E.; EVERHART, S. S. **Small business economics**, v. 10, n. 2, p. 141–151, 1998.

STEZANO, F. The role of technology centers as intermediary organizations facilitating links for innovation: Four cases of federal technology centers in Mexico: The role of technology centers as intermediary organizations. **The review of policy research**, v. 35, n. 4, p. 642–666, 2018.

STOKES, D. E. **Pasteur's quadrant Basic science and technological innovation**. Washington, DC Brookings Institution Press, 1997.

SVAGZDIENE, B.; KUKLYTE, J. The analysis of factors which have impact for summary innovation index in Germany, Estonia and Lithuania. *Transformations in Business and Economics*, v. 15, n. 2B, p. 784–799, 2016.

TIJSSEN, R. J. W. Anatomy of use-inspired researchers: From Pasteur's Quadrant to Pasteur's Cube model. **Research policy**, v. 47, n. 9, p. 1626–1638, 2018.

TORKOMIAN, A. L. V. Panorama dos Núcleos de Inovação Tecnológica no Brasil. In: Santos, M. E. R.; Toledo, P. T. M.; Lotufo, R. A. (Orgs.). Transferência de Tecnologia: Estratégias para a estruturação e gestão de Núcleos de Inovação Tecnológica. Campinas: Komedi, p.21-37, 2009.

TSENG, A. A.; RAUDENSKY, M. Performance evaluations of technology transfer offices of major US research universities. **Journal of technology managementeinnovation**, v. 9, n. 1, p. 93–102, 2014.

TSENG, J.-J. *et al*. A critical review of research on technological pedagogical and content knowledge (TPACK) in language teaching. **Computer assisted language learning**, v. 35, n. 4, p. 948–971, 2022.

TUKOFF-GUIMARÃES, Y. B. *et al*. Patents valuation in core innovation: case study of a Brazilian public university. Innovationemanagement review, v. 18, n. 1, p. 34–50, 2021.

VALENTE, L. V. Comprometimento cognitivo leve no pós-acidente vascular encefálico e conversão para demência em idosos: revisão sistemática. **CiênciaseCognição**, v. 22, n. 1, 2017.

VANSTEENKISTE, M.; LENS, W.; DECI, E. L. Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. **Educational psychologist**, v. 41, n. 1, p. 19–31, 2006.

VEER, T.; JELL, F. Contributing to markets for technology? A comparison of patent filing motives of individual inventors, small companies and universities. Technovation, v. 32, n. 9–10, p. 513–522, 2012.

VON PROFF, S.; BUENSTORF, G.; HUMMEL, M. University patenting in Germany before and after 2002: What role did the professors' privilege play? **Industry and innovation**, v. 19, n. 1, p. 23–44, 2012.

WEISS, Gary M. Mining with rarity: a unifying framework. *ACM Sigkdd Explorations Newsletter*, v. 6, n. 1, p. 7-19, 2004.

WOODSIDE, A. G. Moving away from bad practices in research toward constructing useful theory and doing useful research. Em: **Bad to Good**. [s.l.] Emerald Group Publishing Limited, 2016. p. 1–56.

WORLD INTELLECTUAL PROPERTY ORGANIZATION. **World intellectual property indicators 2022**. [s.l.] Unknown, [s.d.].

WRIGHT, M. *et al*. Mid-range universities' linkages with industry: Knowledge types and the role of intermediaries. **Research policy**, v. 37, n. 8, p. 1205–1223, 2008.

YANG, Q.; MINUTOLO, M. C. The strategic approaches for a new typology of firm patent portfolios. **International journal of innovation and technology management**, v. 13, n. 02, p. 1650012, 2016.

YIN, R. K. **Estudo de caso: planejamento e métodos.** 4ª ed. Porto Alegre: Bookman, 2010.

ZEMTSOV, S. *et al*. Determinants of regional innovation in Russia: Are people or capital more important? **Foresight-Russia**, v. 10, n. 20, p. 29–42, 2016.

ZHOU, H. (2018). Statistics relating to PCT applications by universities. Geneve: WIPO.

ZHOU, W.; JIANG, H.; KONG, Q. Technology transfer under China's foreign investment regime: Does the WTO provide a solution? **SSRN Electronic Journal**, 2019.

ZI, A.; BLIND, K. Researchers' participation in standardisation: a case study from a public research institute in Germany. **The Journal of technology transfer**, v. 40, n. 2, p. 346–360, 2015.

APÊNDICE A – Questionário *on-line*

QUESTIONÁRIO MODELO

PERFIL INVENTORES

Olá Pesquisador. Você e mais de 400 inventores que têm pedidos de depósito de patentes no Brasil estão sendo convidados a participar desta pesquisa que visa mapear o perfil motivacional dos inventores ligados às Universidades públicas brasileiras.

A pesquisa de doutorado está sendo desenvolvida pelo PPPRO (Programa de Engenharia de Produção e Sistemas) do CEFET-RJ na área de Propriedade Intelectual e Inovação.

Os dados são sigilosos e serão utilizados apenas com finalidades acadêmicas.

Obrigado por participar!

Informe seu nome completo:

Fique tranquilo! Seu nome não será divulgado, mas precisamos que o informe para cruzarmos os dados com o seu Currículo Lattes. Por favor, evite abreviações.

Ano de nascimento:

Por favor, informe o ano no formato XXXX (Exemplo: 1978)

Local de nascimento:

Por favor, informe cidade e estado conforme exemplo (São Paulo/SP)

GÊNERO:

( ) Feminino

( ) Masculino

( ) Outro

Você teve alguma patente concedida?

( ) Sim

( ) Não

Você teve alguma patente licenciada?

( ) Sim

( ) Não

Você teve alguma patente depositada no exterior ou via Tratado de Cooperação de Patentes (PCT)?

( ) Sim

( ) Não

Algum pedido de patente foi resultado de interação com a indústria?

( ) Sim
( ) Não

Por favor, indique qual das seguintes afirmações melhor descreve sua orientação profissional. (Lam, 2011)

( ) Acredito que a universidade e a indústria devem ser distintas. Assim, busco atuar e avançar profissionalmente apenas no campo acadêmico.

( ) Acredito que a universidade e a indústria devem ser distintas, mas busco desenvolver atividades ligadas à indústria principalmente para adquirir recursos para apoiar a pesquisa acadêmica.

( ) Acredito na importância da colaboração da academia-indústria e busco desenvolver atividades vinculadas a indústria visando ao avanço científico.

( ) Acredito na importância da colaboração da academia-indústria e busco desenvolver atividades vinculadas à indústria visando ao aproveitamento comercial.

Pure traditional - Penso que a universidade e a indústria devem ser distintas. Assim, busco atuar e avançar profissionalmente apenas no campo acadêmico.

Pragmatic traditional' - Penso que a universidade e a indústria devem ser distintas, mas busco desenvolver atividades ligadas à indústria principalmente para adquirir recursos para apoiar a pesquisa acadêmica.

Hybrid' - Acredito na importância da colaboração da academia-indústria e busco desenvolver atividades vinculadas a indústria visando ao avanço científico.

Entrepreneurial' - Acredito na importância da colaboração da academia-indústria e busco desenvolver atividades vinculadas à indústria visando ao aproveitamento comercial. (Lam, 2011)

Indique a importância dos seguintes fatores na sua motivação pessoal ao se engajar em atividades de patenteamento?

Por favor, considere a seguinte escala:

1 Nem um pouco importante
2 Pouco importante
3 Moderadamente importante
4 Muito importante
5 Extremamente importante

Acesso a meios e insumos, inclusive tecnológicos, para o desenvolvimento de pesquisas

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Apoio administrativo institucional no processo de patenteamento, principalmente do Núcleo de Inovação Tecnológica (NIT) da sua universidade

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Atendimento de indicadores de produção e possibilidades de publicação das pesquisas cujos resultados são patenteáveis

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Aumento da renda pessoal e/ou benefícios econômicos decorrente da atividade de pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Aumento do prestígio e visibilidade acadêmica com reconhecimento e reputação pelos pares

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Baixa mobilidade: vinculação e permanência a um local fixo e específico para o desenvolvimento da pesquisa.

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Colocação profissional e nível de responsabilidade que estimula o desenvolvimento da pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Competição e colaboração entre os pares na formação de redes pessoais e profissionais

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Contribuição da promoção e/ou progressão na carreira com impacto na remuneração

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Contribuição dos resultados da pesquisa para o desenvolvimento social e econômico da sociedade

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Disponibilidade de materiais, dados e recursos oriundos da indústria para o desenvolvimento das atividades de pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Experiência profissional anterior como acadêmico e pesquisador

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Gosto por desafios, resolução de problemas de pesquisa e curiosidade intelectual

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Habilidade para finanças e realização de negócios

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Independência e autonomia no desenvolvimento da pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Influência de colaboradores e parceiros estrangeiros

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Influência de membros do grupo de pesquisa por meio de troca de experiências

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Instalações e condições infraestruturais para o desenvolvimento da pesquisa disponíveis na sua instituição

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Licenciamento e Comercialização dos resultados da pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Momento profissional da carreira acadêmica ou de pesquisador

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Obtenção de financiamentos e recursos para a pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Participação em viagens e/ou experiências em outros lugares/culturas como influência para o desenvolvimento de pesquisas

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Potencial de aplicação dos resultados da pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Proteção e apropriação intelectual dos resultados da pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Qualificação e competência pessoal para o desenvolvimento da pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Relações e cooperação de mercado com a indústria, inclusive com atração de recursos financeiros

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Satisfação pessoal e realização profissional na realização da pesquisa

Nem um pouco importante ( ) ( ) ( ) ( ) ( ) Extremamente importante

Campo livre para, se desejar, complementar sua opinião, compartilhar impressões, sugestões e/ou dúvidas.

_____

APÊNDICE B – Termo Circunstanciado de Livre Esclarecimento (TCLE)

CEFET/RJ

Ministério da Educação
Secretaria de Educação Profissional e Tecnológica
Centro federal de Educação Tecnológica Celso Suckow da Fonseca
Diretoria de Pesquisa e Pós-Graduação
PROGRAMA DE ENGENHARIA DE PRODUÇÃO E SISTEMAS

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado(a) como voluntário(a) a participar da pesquisa **"Perfil motivacional dos inventores das universidades públicas brasileiras".** Neste trabalho pretendemos levantar o perfil motivacional dos inventores vinculados a universidades públicas brasileiras a fim de identificar a relação das motivações extrínsecas e intrínsecas no desenvolvimento de pesquisas e principalmente de seus impactos nos pedidos de depósitos de patentes no âmbito nacional. A pesquisa considera o protagonismo que as universidades públicas brasileiras têm obtido em depositar patentes, importantes indicadores de geração e transferência tecnológica e meio de atração de investimentos para a pesquisa acadêmica.

Com o aumento da atividade de patenteamento, parte dos estudos sobre o tema no cenário internacional e nacional, apesar de poucos e recentes, tem buscado levantar o perfil dos patenteadores com foco em características objetivas como formação, vínculo institucional, indicadores de produtividade e área de pesquisa. Uma lacuna no perfil está em conhecer as motivações pelas quais se patenteia e a quais objetivos se busca atender.

O que nos move a pesquisar é a consideração de que a capacidade inventiva e resolutiva dos pesquisadores é o principal insumo intelectual inerente ao ato de pesquisar e inovar. Assim, a hipótese de pesquisa é que as motivações pessoais têm influência e efeito determinante no desenvolvimento e resultado das pesquisas dos centros acadêmicos e no cumprimento das suas finalidades no processo de inovação tecnológica. Para além de apresentar o perfil motivacional do inventor vinculado à

universidade pública brasileira, a pesquisa pretende disponibilizar em domínio público uma metodologia de coleta, tratamento e relação de dados qualitativos e quantitativos, devidamente estruturada e validada, para, se reproduzida, viabilizar a identificação de perfis motivacionais.

Para esta pesquisa procederemos a extração de dados dos currículos disponibilizados na plataforma Lattes e a aplicação de questionário *on-line*, buscando obter dados (qualitativos e quantitativos) e medir a importância de alguns fatores na sua motivação pessoal ao se engajar em atividades de patenteamento. Visando enriquecer a pesquisa, se desejar complementar a sua opinião, poderá deixar comentários no campo (livre) destinado às observações.

Para participar, você não terá nenhum custo, nem receberá qualquer vantagem financeira. Apesar disso, caso sejam identificados e comprovados danos provenientes desta pesquisa, você tem assegurado o direito à indenização. Você será esclarecido (a) em qualquer aspecto que desejar e estará livre para participar ou recusar-se a participar e poderá retirar seu consentimento ou interromper sua participação a qualquer momento. Sua participação é voluntária e a recusa em participar não acarretará qualquer penalidade ou modificação na forma como é atendido(a) pelo pesquisador. Sua identidade será tratada com padrões profissionais de sigilo e não será identificado em nenhuma publicação.

Os riscos envolvidos na pesquisa consistem em eventual sensação de desconforto, como cansaço ou aborrecimento em responder o questionário, situação que poderá ser minimizada, bastando deixar de responder as questões, interromper a pesquisa momentaneamente ou interromper a pesquisa de forma definitiva ou, ainda, desistir da pesquisa, em qualquer fase, sem penalização alguma. É possível que haja algum constrangimento em responder a questões sensíveis sobre temas que envolvem interesses financeiros e outros ligados a comportamento no ato de pesquisar. Para minimizar tal risco a resposta ao questionário é facultada, além de serem garantidas a confidencialidade das respostas e a privacidade dos respondentes. O risco de divulgação de dados pessoais e outros, inclusive o nome do respondente, que possam caracterizar o respondente são minimizados na medida em que reiteramos o caráter confidencial das respostas e a finalidade estritamente acadêmica da pesquisa, cujos dados serão considerados no geral e não no individual, além dos instrumentos on-line

serão descontinuados após a coleta dos dados. A pesquisa tem o risco de ocupar o tempo do respondente em preencher o questionário, o que é minimizado pela utilização de linguagem que pretende transmitir fluidez na leitura e de questões dispostas de modo a tentar facilitar a visualização.

Para o saneamento de dúvidas em relação à pesquisa quanto aos objetivos, tratamento dos dados e outros é disponibilizado ao respondente um campo próprio no formulário para, se desejar, complementar sua opinião, compartilhar impressões, sugestões e/ou dúvidas, além de ter disponível, no fim do preenchimento do formulário, o endereço eletrônico do pesquisador por meio do qual é facultado o acesso a maiores informações e ao seu resultado individual e/ou coletivo, neste sendo preservada a identidade dos respondentes, a confidencialidade dos dados e os objetivos da pesquisa.

Se aceitar participar da pesquisa, o respondente terá o benefício de ter disponibilizado, mediante solicitação, dos dados que permitem identificar seu perfil de inventor, bem como do perfil agregado dos respondentes da entidade universitária a que está vinculado, neste caso sendo preservada a confidencialidade dos dados e a privacidade dos respondentes, podendo desenvolver o autoconhecimento e promoção de ações visando o aprimoramento de competências e objetivos. O desenvolvimento da pesquisa beneficia a comunidade científica na medida em que visa compartilhar uma metodologia que se pretende crível, reproduzível e validada para identificação e mapeamento de perfis motivacionais. Os resultados desta pesquisa estarão, portanto, a sua disposição quando finalizada. O nome ou o material que indique sua participação não será liberado sem a sua permissão.

Sua participação na presente pesquisa consiste no conhecimento e concordância com o presente termo, estando informado(a) dos objetivos do presente estudo de maneira clara e detalhada, tendo esclarecidas suas dúvidas. É sabedor de que a qualquer momento poderá solicitar novas informações e modificar sua decisão de participar se assim o desejar.

CONTATOS DO PESQUISADOR RESPONSÁVEL:
Nome: Carlos Silva de Jesus
Endereço:
E-mail:
Telefone:

APÊNDICE C – Repositório do *dataset*

O *dataset* está disponível no repositório no seguinte endereço eletrônico: https://data.mendeley.com/datasets/spjszyw8f5/1

APÊNDICE D – Códigos-fonte

Arquivos com o conjunto completo de códigos-fonte desenvolvidos em *Python* e aplicados na pesquisa estão disponíveis em repositório no seguinte endereço eletrônico: https://doi.org/10.5281/zenodo.8321898

A alguns códigos-fonte referentes a determinadas etapas estão destacados a seguir com as respectivas referências das seções do trabalho em que foram aplicados.

Quadro D1 – Balanceamento na categoria natureza motivacional conforme seção V.3.2.1 (*z-score*; ranqueamento; mínimo-máximo)

```
import pandas as pd
import numpy as np
import collections as cl
import scipy.stats as st
import matplotlib.pyplot as plt

#pd.options.display.float_format = '{:,.2f}'.format
#pd.set_option('display.precision', 2)
pd.set_option('display.float_format',  '{:,.2f}'.format)
```

```
data=pd.read_excel('patentees_data.xlsx')
data
```

```
attr_groups.append(['z ' + i for i in attr_groups[0]])
data[attr_groups[-1]] = data[attr_groups[0]].apply(st.zscore, nan_policy='omit')


attr_groups.append(['rank ' + i for i in attr_groups[0]])
data[attr_groups[-1]] = data[attr_groups[0]].rank()


attr_groups.append(['mm ' + i for i in attr_groups[0]])
data[attr_groups[-1]] = data[attr_groups[0]].apply(lambda x: (x - x.min())/(x.max() - x.min()))
```

```
attr_groups.append(['z ' + i for i in translation_dict.values()])

data[attr_groups[-1]] = data[translation_dict.values()].apply(st.zscore,
```

```
for i in attr_groups:
    #print(str(i))
    display(data[i].describe().T)
    data[i].plot.box()
    plt.savefig(f'{'' ''.join(i)}.pdf')
    plt.show()
    print('0: não-respondente, 1: mais identificado com apenas uma categoria, 2: empatado entre 2 categorias')
    display(cl.Counter((data[i].values == data[i].values.max(1)[:,np.newaxis]).sum(1)))
    print('Total identificado com cada categoria')
    display(data[i].idxmax(1).value_counts())
    print('-'*111)
```

Quadro D2 – Balanceamento da relação entre normas e valores pessoais conforme seção V.3.2.2.

```
introjection = pd.concat(axis = 1, objs = (
    data['Nature of motivation'] == 'Mixed',
    data['Interaction in patenting process. Active or passive?'] == 'Passive',
    data['Classification regarding professional orientation'].isin(['Pragmatic traditional', 'Pure Traditional']),
))

data['introjection'] = introjection.mean(1)
```

```
identification = pd.concat(axis = 1, objs = (
    data['Nature of motivation'] == 'Extrinsic',
    data['Interaction in patenting process. Active or passive?'].isin(['Passive', 'Active']),
    data['Classification regarding professional orientation'] == 'Hybrid',
))

data['identification'] = identification.mean(1)
```

```
integration = pd.concat(axis = 1, objs = (
    data['Nature of motivation'] == 'Intrinsic',
    data['Interaction in patenting process. Active or passive?'] == 'Active',
    data['Classification regarding professional orientation'] == 'Entrepreneurial',
))

data['integration'] = integration.mean(1)
```

```
non_respondents = data['Classification regarding professional orientation'].isna()
data.loc[non_respondents, ['introjection', 'identification', 'integration']] = np.nan
```

```
attr_groups = [['introjection', 'identification', 'integration']]
```

```
attr_groups.append(['z ' + i for i in attr_groups[0]])
data[attr_groups[-1]] = data[attr_groups[0]].apply(st.zscore, nan_policy='omit')

attr_groups.append(['rank ' + i for i in attr_groups[0]])
data[attr_groups[-1]] = data[attr_groups[0]].rank()

attr_groups.append(['mm ' + i for i in attr_groups[0]])
data[attr_groups[-1]] = data[attr_groups[0]].apply(lambda x: (x - x.min())/(x.max() - x.min()))
```

```
for i in attr_groups:
    #print(str(i))
    display(data[i].describe().T)
    data[i].plot.box()
    plt.savefig(f'{'' ''.join(i)}.pdf')
    plt.show()
    print('0: não-respondente, 1: mais identificado com apenas uma categoria, 2: empatado entre 2
categorias')
    display(cl.Counter((data[i].values == data[i].values.max(1)[:,np.newaxis]).sum(1)))
    print('Total identificado com cada categoria')
    display(data[i].idxmax(1).value_counts())
    print('-'*111)
```

Quadro D3 – Conversão das variáveis categóricas em numéricas conforme seção V.3.3

```
yes_no_attrs = [
        'Have you ever had a patent awarded?',
        'Have you ever had any patents licensed?',
        'Have you ever had a patent deposited abroad through Patent Cooperation Treaty (PCT)?',
        'Have any patent request been the result of interaction with the industry?'
]
data[yes_no_attrs] = (data[yes_no_attrs] == 'Yes')
display(pd.DataFrame([data[yes_no_attrs].sum(), data[yes_no_attrs].mean()]).T)

input_attrs.extend(yes_no_attrs)
```

Atributo gênero

```
data['Gender'] = (data['Gender'] == 'Female')
display(data['Gender'].sum(), data['Gender'].mean())

input_attrs.append('Gender')
```

Quadro D4 – Código-fonte do algoritmo de árvore de decisão com a aplicação da validação cruzada conforme a seção V.4.2.

```
target_attrs = [
	'Interaction in patenting process. Active or passive?',
	'Classification regarding professional orientation',
	'Nature of motivation',
	'Relationship between standards / personal values'
]
results = {}

for ta in target_attrs:
	display(data[ta].value_counts().to_frame())

	clf = sktree.DecisionTreeClassifier().fit(data[input_attrs], data[ta])
	display(pd.DataFrame([
		clf.feature_names_in_[clf.feature_importances_.argsort()],
		clf.feature_importances_[clf.feature_importances_.argsort()],
	]).T)

	cv_output = skms.cross_validate(
		sktree.DecisionTreeClassifier(),
		data[input_attrs],
		data[ta],
		cv = skms.RepeatedStratifiedKFold(n_splits=4, n_repeats=25),
		#cv = skms.LeaveOneOut(),
		#scoring = 'f1_macro'
	)
	results[ta] = cv_output['test_score']

	print('-'*111)


results = pd.DataFrame(results)
display(results.describe())

results.plot.box(vert=False)
```