



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

**Previsão da evasão acadêmica no ensino superior: O  
caso dos cursos de graduação presencial da  
Universidade de Brasília**

Edmilson Cosme da Silva

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Orientador

Prof. Dr. Sergio Antônio Andrade de Freitas

Brasília  
2023

Ficha catalográfica elaborada automaticamente,  
com os dados fornecidos pelo(a) autor(a)

CS586p Cosme da Silva, Edmilson  
Previsão da evasão acadêmica no ensino superior: O caso dos cursos de graduação presencial da Universidade de Brasília / Edmilson Cosme da Silva; orientador Sergio Antônio Andrade de Freitas. -- Brasília, 2023.  
136 p.

Dissertação(Mestrado Profissional em Computação Aplicada)  
-- Universidade de Brasília, 2023.

1. Evasão e Indicadores de Evasão. 2. Algoritmos de aprendizagem de máquina. 3. Previsão e predição.. I. Antônio Andrade de Freitas, Sergio , orient. II. Título.



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

**Previsão da evasão acadêmica no ensino superior: O  
caso dos cursos de graduação presencial da  
Universidade de Brasília**

Edmilson Cosme da Silva

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Prof. Dr. Sergio Antônio Andrade de Freitas (Orientador)  
FGA/UnB

Prof. Dr. Rafael Ferreira Leite de Mello  
Universidade Federal Rural de Pernambuco (UFRPE)

Profa. Dra. Andrea Felipe Cabello  
FACE/UNB

Prof. Dr. Gladston Luiz da Silva  
Coordenador do Programa de Pós-graduação em Computação Aplicada

Brasília, 03 de julho de 2023

# Dedicatória

O presente trabalho é dedicado à memória dos meus pais Rivaldo Cosme da Silva e Adalva Miguel da Silva, que partiram cedo demais, mas deixaram um legado de amor, dedicação e sabedoria. À minha esposa, Aurea Regina Candido Aguiar, minha cúmplice e parceira de vida, que me instiga imensamente a ser quem sou. E pôr fim ao meu filho, Matheus Albuquerque Cosme, meu maior orgulho e alegria. Você é a minha inspiração e o meu motivo para seguir em frente.

# Agradecimentos

Agradeço, em primeiro lugar, a Deus por ter-me concedido saúde, força e sabedoria para concretizar o meu sonho de mestrado.

À minha família, que sempre esteve presente, em especial aos meus irmãos, Edison, Edileuza, Cristóvão e Helenice, que sempre souberam tolerar perfeitamente a minha ausência.

Expresso minha profunda gratidão ao meu orientador, o Prof. Dr. Sergio Freitas, pelo seu conhecimento, sua disponibilidade, sua generosidade e seu entusiasmo. Sua orientação foi crucial para a realização deste trabalho. Me encorajando mesmo nos momentos em que tive que me aventurar em lugares que nunca imaginei estar. Tudo isso contribuiu para o meu crescimento pessoal e profissional.

Quero agradecer também a minha coorientadora Profa. Cristiane Soares, sua orientação foi essencial para a condução da revisão sistema da literatura.

Dedico também agradecimentos às colegas do grupo de pesquisa Amanda Emilly e Letícia Karla, com as quais foi possível realizar um trabalho em equipe, para planejar e conduzir uma revisão sistemática da literatura que cominou na escrita e publicação de um artigo forjado a cinco mãos.

Gostaria de expressar minha sincera gratidão aos ilustres membros da banca Prof. Dr. Rafael Ferreira e Profa. Dra. Andrea Cabello. As suas contribuições e orientações foram fundamentais para o desenvolvimento desse estudo.

Aos meus amigos e colegas da Secretária de Tecnologia da Informação (STI) e do Arquivo Central (ACE) que estiveram presentes durante este período, oferecendo apoio nos momentos mais importantes.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

# Resumo

Já há algum tempo, pesquisadores (inter)nacionais têm estudado a evasão nos cursos de ensino superior, classificando-a em dois tipos: estudantes que abandonam a universidade; e estudantes que abandonam o ensino superior. Ambas as situações causam danos à instituição, aos alunos e à sociedade em geral. A partir de 1995, com a criação da ANDIFES, os estudos começaram a se tornar mais frequentes no Brasil. Essa comissão elaborou relatórios que analisaram os índices de diplomação, retenção e evasão nos cursos de graduação nas Universidades Brasileiras. A evasão institucional foi um dos temas de estudo, sendo descrita como a saída do aluno de seu curso de origem, sem finalizar o curso. A Universidade de Brasília (UnB), tendo em vista a problemática que envolve a evasão do curso pelos estudantes, tem criado mecanismos que permitam aumentar a permanência dos estudantes nos cursos de graduação. O objetivo deste trabalho foi elaborar e avaliar um modelo de análise que permita utilizar os dados acadêmicos para previsão de evasão nos cursos de graduação presencial. A Revisão Sistemática da Literatura foi realizada com o propósito de identificar os fatores que impactam na evasão, a partir desses fatores, definir indicadores que possam ser extraídos dos sistemas acadêmicos da UnB. Ela também contribuiu na para selecionar algoritmos/ferramentas que deram suporte nas análises. O principal resultado da revisão sistemática foi a identificação de 29 fatores usados pelos pesquisadores, onde a pontuação média, gênero, notas do curso foram os mais utilizados. Sobre as ferramentas, a Regressão, a Árvore de Decisão e a Rede Neural foram os algoritmos mais usados. A partir desse resultado preliminar, foi criado o Modelo de Análise na Graduação (MAGRA), que utiliza os indicadores existentes nos sistemas acadêmicos da UnB em conjunto com ferramentas de aprendizagem de máquina para prever alunos com potencial de risco de evasão. A pesquisa se desenvolveu em um ambiente que contempla dois cenários. O primeiro foi desenvolvido na Faculdade do Gama (FGA), que serviu de protótipo para a criação do MAGRA, e o segundo na UnB. Nas etapas de testes, nas quais foram usadas somente as disciplinas cursadas pelos alunos, mostrou-se que o número de vezes que os alunos cursam uma disciplina pode ser um indicador da dificuldade desse aluno em cumprir a disciplina no tempo determinado. Empregando as novas variáveis, houve um aumento no número de modelos válidos e, conseqüentemente, um aumento nos

curios e turmas analisados, o que resultou em um número maior de previsões. A situação foi percebida ao conduzir os estudos na Faculdade do Gama e na Universidade de Brasília. Para melhorar a identificação precoce dos alunos com características de evasão, é necessário criar mecanismos de feedback por parte dos coordenadores de cursos, incluir novos sistemas, melhorar a qualidade dos dados e ajustar os parâmetros dos algoritmos.

**Palavras-chave:** Evasão, Indicadores de Evasão, Algoritmos de aprendizagem de máquina, Previsão, Predição

# Abstract

*For some time now, (inter)national researchers have been studying dropout rates in higher education courses, classifying it into two types: students who drop out of university; students dropping out of higher education. Both situations cause harm to the institution, students and society in general. From 1995, with the creation of ANDIFES, studies began to become more frequent in Brazil. This commission prepared reports that analyzed the graduation, retention and dropout rates in undergraduate courses in Brazilian Universities. Institutional dropout was one of the topics of study, being described as the exit of the student from his course of origin, without finishing the course. The Universidade de Brasília (UnB), considering the problem that involves the evasion of the course by students, has created mechanisms that allow increasing the permanence of students in undergraduate courses. The objective of this work was to elaborate and evaluate an analysis model that allows the use of academic data to predict dropout in undergraduate courses. The Systematic Review of the Literature was carried out with the purpose of identifying the factors that impact on dropout, from these factors, to define indicators that can be extracted from the academic systems of the UnB. She also contributed to the selection of algorithms/tools that supported the analyses. The main result of the systematic review was the identification of 29 factors used by the researchers, where the average score, gender, course grades were the most used. About the tools, the Regression, the Decision Tree and the Neural Network were the most used algorithms. From this preliminary result, the Undergraduate Analysis Model (MAGRA) was created, which uses the existing indicators in the academic systems of the UnB in conjunction with machine learning tools to predict students with potential risk of dropping out. The research was developed in an environment that contemplates two scenarios. The first was developed at the Faculdade do Gama, which served as a prototype for the creation of MAGRA, and the second at the UnB. In the stages of tests, in which only the subjects taken by the students were used, it was shown that the number of times that the students take a course can be an indicator of the difficulty of this student in fulfilling the discipline in the determined time. With the use of the new variables, there was an increase in the number of valid models and, consequently, an increase in the courses and classes analyzed, which resulted in a higher*

*number of predictions. The situation was perceived while conducting studies at the Faculdade do Gama and the Universidade de Brasília. To improve the early identification of students with dropout characteristics, it is necessary to create feedback mechanisms by course coordinators, include new systems, improve data quality and adjust algorithm parameters.*

**Keywords:** Drop Out, Drop Out indicators, Machine learning algorithms, Forecasting, Prediction

# Sumário

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introdução</b>   | <b>1</b>  |
| 1.1      | Contextualização . . . . .  | 1         |
| 1.2      | Motivação . . . . .   | 2         |
| 1.3      | Questão de Pesquisa . . . . .                                     | 3         |
| 1.4      | Objetivos Principal e Secundários . . . . .                       | 3         |
| 1.5      | Metodologia . . . . .   | 4         |
| 1.6      | Estrutura da Dissertação . . . . .                                | 4         |
| <b>2</b> | <b>Fundamentação Teórica</b>                                      | <b>6</b>  |
| 2.1      | Evasão . . . . .  | 6         |
| 2.2      | Retenção . . . . .  | 8         |
| 2.3      | Indicadores de Evasão e Retenção . . . . .                        | 11        |
| 2.3.1    | Fatores de e Evasão e Retenção . . . . .                          | 11        |
| 2.3.2    | Identificação dos Fatores . . . . .                               | 12        |
| 2.3.3    | Fatores de retenção e Ferramentas de Previsão . . . . .           | 16        |
| 2.3.4    | Lista final de indicadores . . . . .                              | 21        |
| 2.4      | Trabalhos Relacionados . . . . .                                  | 22        |
| <b>3</b> | <b>Contexto Organizacional</b>                                    | <b>29</b> |
| 3.1      | Universidade de Brasília . . . . .                                | 29        |
| 3.1.1    | Secretaria de Tecnologia da Informação (STI) . . . . .            | 30        |
| 3.1.2    | Decanato de Ensino de Graduação (DEG) . . . . .                   | 32        |
| 3.1.3    | Centro de Educação a Distância (CEAD) . . . . .                   | 33        |
| 3.2      | Legislação . . . . .  | 34        |
| 3.3      | Contexto da Pesquisa . . . . .                                    | 37        |
| <b>4</b> | <b>Mapeamento das Bases de Dados Educacionais da UnB</b>          | <b>38</b> |
| 4.1      | Sistemas Mapeados . . . . .                                       | 38        |
| 4.1.1    | Sistema de Integrado de Gestão de Atividades Acadêmicas . . . . . | 40        |
| 4.1.2    | Sistema de Graduação . . . . .                                    | 42        |

|          |   |           |
|----------|---|-----------|
| 4.1.3    | Ambientes de Ensino-aprendizagem On-line . . . . .          | 45        |
| <b>5</b> | <b>Modelo de Análise na Graduação - MAGRA</b>               | <b>46</b> |
| 5.1      | Estrutura do MAGRA . . . . .                                | 46        |
| 5.1.1    | Seleção das bases de dados, tabelas e atributos . . . . .   | 47        |
| 5.1.2    | Construção do ETL para a base de dados analítica . . . . .  | 48        |
| 5.1.3    | Algoritmos e modelos de aprendizagem . . . . .              | 50        |
| 5.1.4    | Avaliar a melhor solução . . . . .                          | 51        |
| <b>6</b> | <b>Testes e Análises de Dados</b>                           | <b>53</b> |
| 6.1      | 1º Cenário — Faculdade do Gama . . . . .                    | 53        |
| 6.1.1    | Construção da Matriz de Fatores . . . . .                   | 54        |
| 6.1.2    | Construção dos Modelos . . . . .                            | 55        |
| 6.1.3    | Avaliação dos Modelos . . . . .                             | 55        |
| 6.1.4    | Análises nas Previsões . . . . .                            | 57        |
| 6.1.5    | Inclusão de novos fatores . . . . .                         | 59        |
| 6.2      | 2º Cenário — Universidade de Brasília . . . . .             | 62        |
| 6.2.1    | Construção da Matriz de Fatores da UnB . . . . .            | 63        |
| 6.2.2    | Construção dos Modelos da UnB . . . . .                     | 64        |
| 6.2.3    | Avaliação dos Modelos UnB . . . . .                         | 64        |
| 6.2.4    | Análises das Previsões UnB . . . . .                        | 64        |
| 6.2.5    | Inclusão de novos fatores UnB . . . . .                     | 67        |
| 6.2.6    | Análises das Previsões com novas variáveis UnB . . . . .    | 67        |
| 6.2.7    | Implicações Práticas . . . . .                              | 69        |
| <b>7</b> | <b>Conclusão</b>  | <b>71</b> |
| 7.1      | Trabalhos futuros: . . . . .                                | 73        |
|          | <b>Referências</b>  | <b>74</b> |
|          | <b>Apêndice</b>   | <b>80</b> |
| <b>A</b> | <b>Revisão Sistemática da Literatura</b>                    | <b>81</b> |
| A.1      | Planejamento da Revisão . . . . .                           | 81        |
| A.1.1    | Especificação das Questões de pesquisa para a RSL . . . . . | 82        |
| A.1.2    | Desenvolvimento do Protocolo . . . . .                      | 82        |
| A.1.3    | Validar o Protocolo . . . . .                               | 85        |
| A.2      | Condução da Revisão . . . . .                               | 85        |
| A.2.1    | Identificação das pesquisas . . . . .                       | 85        |

|          |  |            |
|----------|--|------------|
| A.2.2    | Seleção dos estudos . . . . .                  | 86         |
| A.2.3    | Avaliar a qualidade . . . . .                  | 87         |
| A.2.4    | Extração de dados . . . . .                    | 87         |
| A.2.5    | Sintetize dos dados . . . . .                  | 87         |
| <b>B</b> | <b>Ambientes Virtuais de Aprendizagem</b>      | <b>94</b>  |
| B.1      | Aprender UnB — Moodle . . . . .                | 94         |
| B.2      | Microsoft Teams . . . . .                      | 94         |
| B.3      | Disponibilidade dos Dados . . . . .            | 96         |
| <b>C</b> | <b>Fatores e Atributos dos Bancos de Dados</b> | <b>98</b>  |
| <b>D</b> | <b>Consulta SQL e Tipos de Dados</b>           | <b>101</b> |
|          | <b>Anexo</b>                                   | <b>106</b> |
| <b>I</b> | <b>Legislação</b>                              | <b>106</b> |

# Lista de Figuras

|      |   |    |
|------|---|----|
| 2.1  | Fatores Demográficos - Fonte: Silva et al. [1]. . . . .   | 14 |
| 2.2  | Fatores Acadêmicos - Fonte: Silva et al. [1]. . . . .   | 14 |
| 2.3  | Fatores de Aprendizagem - Fonte: Silva et al. [1]. . . . .  | 15 |
| 2.4  | Algoritmos utilizados nos artigos. . . . .  | 18 |
| 2.5  | Exemplo do gráfico da regressão logística. Fonte: Hosmer et al. [2]. . . . .                                      | 18 |
| 2.6  | Exemplo de árvore de decisão. Fonte: Rokach e Maimon [3]. . . . .   | 19 |
| 2.7  | Estrutura básica da rede neural artificial. Fonte: Aggarwal [4]. . . . .  | 20 |
| 2.8  | Visão geral dos indicadores - Fonte: Elaborado pelo autor. . . . .  | 22 |
| 2.9  | Metodologia de estudo: utilizando abordagens qualitativas e quantitativas<br>- Fonte: Freitas et al. [5]. . . . . | 23 |
| 2.10 | Modelo conceitual de interligação das variáveis preditoras (Adaptado de<br>Adejo e Conolly [6]) . . . . .         | 25 |
| 2.11 | Framework para prever o desempenho dos estudantes — Fonte: Adejo e<br>Conolly [6] . . . . .                       | 26 |
| 2.12 | Diagrama da estrutura do algoritmo de empilhamento — Fonte: Adejo e<br>Conolly [6] . . . . .                      | 27 |
| 3.1  | Organograma da Secretaria da Tecnologia da Informação. . . . .  | 32 |
| 4.1  | Interface do SIGAA - Menu de graduação. . . . .   | 40 |
| 4.2  | Interface do SIGAA - Dados do Aluno. . . . .  | 41 |
| 4.3  | Mapeamento modelo de dados e fatores no SIGAA. . . . .  | 42 |
| 4.4  | Distribuição do percentual de alunos por sexo no SIGAA. . . . .   | 42 |
| 4.5  | Distribuição do percentual da forma de ingresso de alunos ativos - SIGAA. . . . .                                 | 43 |
| 4.6  | Distribuição do total de cursos ativos por período no SIGAA. . . . .  | 43 |
| 4.7  | Modelo de dados SIGRA. . . . .  | 43 |
| 4.8  | Distribuição de alunos por sexo no SIGRA. . . . .   | 44 |
| 4.9  | Distribuição da forma de ingresso no SIGRA. . . . .   | 44 |
| 4.10 | Distribuição da forma de saída registrada no SIGRA. . . . .   | 45 |

|      |   |    |
|------|---|----|
| 5.1  | Modelo para predição da Evasão. . . . .   | 47 |
| 5.2  | Área de construção do fluxo de dados no Pentaho. . . . .  | 49 |
| 5.3  | Diagrama do fluxo de dados do MAGRA. . . . .  | 50 |
| 6.1  | Matriz de Fatores com as disciplinas e situação do aluno. . . . .                                 | 54 |
| 6.2  | Total de previsões geral por modelo na FGA. . . . .   | 58 |
| 6.3  | Total de previsões por curso e modelo na FGA. . . . .   | 58 |
| 6.4  | Listagem de matrículas previstas por algoritmos. . . . .  | 59 |
| 6.5  | Novos fatores incluídos. . . . .  | 60 |
| 6.6  | Total de previsões geral por modelo na FGA com novos fatores. . . . .                             | 61 |
| 6.7  | Total de previsões por curso e modelo na FGA com novos fatores. . . . .                           | 61 |
| 6.8  | Total de previsões por forma de entrada na FGA. . . . .   | 62 |
| 6.9  | Listagem de previsão de alunos da FGA com novas variáveis por algoritmos. . . . .                 | 63 |
| 6.10 | Listagem de modelos válidos por coorte. . . . .   | 65 |
| 6.11 | Visão Geral das Previsões por modelo na UnB. . . . .  | 66 |
| 6.12 | Visão Geral das Previsões por curso e modelo na UnB. . . . .                                      | 66 |
| 6.13 | Visão Geral das Previsões por modelo na UnB com as novas variáveis. . . . .                       | 68 |
| 6.14 | Total de previsões por forma de ingresso na UnB. . . . .  | 68 |
| 6.15 | Visão Geral das Previsões por curso e modelo na UnB com novas variáveis. . . . .                  | 69 |
| A.1  | Visão geral do processo de revisão sistemática . . . . .  | 82 |
| A.2  | Definição da pontuação para as questões de qualidade dos estudos na ferramenta Parsifal . . . . . | 84 |
| A.3  | Percentual da distribuição de artigos identificados por base de pesquisa. . . . .                 | 86 |
| A.4  | Número de artigos selecionados e aceitos por base . . . . .                                       | 86 |
| A.5  | Total de artigos por país. . . . .  | 88 |
| B.1  | Tabelas selecionadas do modelo do Moodle. . . . .   | 95 |
| B.2  | Exemplo da consulta no Moodle. . . . .  | 95 |
| B.3  | Tela da funcionalidade Insights do Teams. . . . .   | 96 |

# Lista de Tabelas

|     |   |    |
|-----|---|----|
| 2.1 | Fatores identificados na literatura - Fonte: Silva et al. [1]. . . . .            | 13 |
| 2.2 | Algoritmos identificados na literatura. . . . .                                   | 17 |
| 2.3 | Ferramentas identificadas na literatura. . . . .                                  | 21 |
| 3.1 | Sistemas acadêmicos mantidos pela STI. . . . .                                    | 31 |
| 3.2 | PDI - Melhorar os resultados da avaliação institucional. . . . .                  | 35 |
| 4.1 | Indicação do sistema fonte dos fatores mapeados na revisão da literatura. . . . . | 39 |
| 6.1 | Total de alunos por cursos e coortes analisados na fase de avaliação. . . . .     | 56 |
| 6.2 | Forma de ingresso presente no conjunto de dados da UnB. . . . .                   | 64 |
| A.1 | Fatores citados pelos artigos. . . . .  | 89 |
| A.2 | Algoritmos citados pelos artigos. . . . .   | 90 |
| A.3 | Softwares citados pelos artigos. . . . .  | 90 |
| A.4 | Artigos selecionados na busca pelos fatores. . . . .                              | 91 |
| A.5 | Artigos selecionados na busca pelos fatores - Continuação. . . . .                | 92 |
| A.6 | Artigos selecionados na busca pelos fatores - Continuação. . . . .                | 93 |
| C.1 | Mapeamento de Fatores, Tabelas e Atributos . . . . .                              | 98 |

# Lista de Abreviaturas e Siglas

**ABI** Área Básica de Ingresso.

**ANDIFES** Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras.

**CEAD** Centro de Educação a Distância.

**CFE** Conselho Federal de Educação.

**CNE** Conselho Nacional de Educação.

**DEG** Decanato de Ensino de Graduação.

**EDM** Mineração de Dados Educacionais.

**ENEM** Exame Nacional do Ensino Médio.

**ETL** Extract Transform-Load.

**FGA** Faculdade do Gama.

**HICSS** 56<sup>a</sup> Conferência Internacional do Havaí sobre Ciências do Sistema.

**MAGRA** Modelo de Análise na Graduação.

**MAP** Modelo de Aprendizagem Analítico.

**PAS** Programa de Avaliação Seriada.

**PDI** Plano de Desenvolvimento Institucional.

**PPPI** Projeto Político Pedagógico Institucional.

**RSL** Revisão Sistemática da Literatura.

**SIGAA** Sistema Integrado de Gestão das Atividades Acadêmicas.

**SIGRA** Sistema de Graduação.

**STEM** The education of science, technology, engineering, and mathematics.

**STI** Secretaria de Tecnologia da Informação.

**TIC** Tecnologias de Comunicação e Informação.

**UnB** Universidade de Brasília.

# Capítulo 1

## Introdução

Neste capítulo são apresentados o contexto desta dissertação, descreve a motivação que leva a questão de pesquisa. Apresenta o objetivo da pesquisa e por fim, descreve a estrutura deste documento.

### 1.1 Contextualização

As transformações causadas na sociedade nos últimos anos estão exigindo mudanças na postura das pessoas que buscam uma melhor qualificação no mercado de trabalho. Desta forma, as universidades têm um papel importante na qualificação dos estudantes que procuram uma formação superior. Esforços devem ser feitos pela instituição de ensino para manter o aluno no curso que ele escolheu. Para Silva Filho et al. [7], o número de alunos que deixa a universidade é um problema internacional que afeta o resultado dos sistemas educacionais. As causas têm sido analisadas por pesquisadores a um bom tempo [7, 8, 9, 10, 11, 12, 13].

Segundo os autores Lenning et al. [14], a evasão há muito tempo é termo familiar no ensino superior. Instituições de elite relacionavam o abandono a uma consequência da manutenção das condições acadêmicas competitivas das quais depende sua reputação. Isso gera sequelas para os alunos e as instituições.

Para Lenning et al. [14], as repercussões para o aluno, estão relacionadas ao desperdício de tempo, energia e dinheiro que podem até mesmo desmotivá-lo a procurar outra instituição.

Conforme DesJardins et al. [15], o estudante que deixa a faculdade antes da conclusão da graduação, impõe custos a sociedade, a instituição e ao estudante. A sociedade perde a partir do momento que não pode contar com um indivíduo melhor preparado intelectual e culturalmente. Economicamente é um profissional que está perdendo a oportunidade de ter uma renda melhor e conseqüentemente ter uma melhor qualidade de vida. Por fim,

as instituições sofrem perdas relacionadas ao financiamento fornecido ao ensino superior, aonde grande parte é baseado em matrículas.

DesJardins et al. [15], acrescenta que quando as matrículas são opostamente afetadas pelo desgaste dos estudantes, as instituições podem incorrer em redução de financiamento desproporcional à perda de estudantes. Além disso, as instituições muitas vezes são cobradas pelas agências de supervisão e pelo público pelas altas taxas de abandono que sinalizam um fracasso da instituição.

Segundo Silva Filho et al. [7], as perdas dos estudantes que iniciam e não terminam seus cursos causam desperdício social, acadêmico e econômico. O setor público é afetado com base nos recursos, os quais são investidos e não tem o devido retorno. No setor privado, se resume em uma importante perda de receita.

Uma das iniciativas pioneiras foi realizada pela Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras (ANDIFES) [16]. Essa comissão reuniu um conjunto significativo de dados sobre o desempenho das universidades públicas brasileiras sobre os índices de diplomação, retenção e evasão dos estudantes, com isso foi possível uniformizar a coleta e tratamento dos dados. O que permitiu que as instituições participantes entendessem o cenário em que estavam inseridas, possibilitando assim uma melhor definição as políticas institucionais.

## 1.2 Motivação

A Secretaria de Tecnologia da Informação (STI) tem um papel relevante nesse processo dado a oportunidade de trabalhar com os dados que promovem os fatores e indicadores que sinalizam as condições para haja a evasão. Isso permite a criação de um ambiente gerador de conhecimento com um valor agregado para os diversos contextos no meio acadêmico.

Desta forma, é de vital importância que a STI disponibilize ferramentas para a comunidade acadêmica, em especial aos coordenadores de cursos e professores, que permitam a visualização prévia dos indicadores de evasão.

A Universidade de Brasília, ciente da problemática que envolve a evasão do curso pelos estudantes, tem criado mecanismos que permitam aumentar a permanência dos estudantes nos cursos de graduação. Dentre esses mecanismos se encontram a Monitoria e Tutoria, ambas de caráter educacional. Outras iniciativas que visam dar apoio a indivíduos pertencentes as famílias menos abastadas que possam ter problema em continuar no ensino superior, como a Bolsa MEC, criada pela portaria do MEC n.º 389 de maio de 2013. Essa portaria pretende: I — viabilizar a permanência, no curso de graduação, de estudantes em situação de vulnerabilidade socioeconômica, em especial os indígenas e quilombolas; II — reduzir custos de manutenção de vagas ociosas em decorrência de

evasão estudantil; e III — promover a democratização do acesso ao ensino superior, por meio da adoção de ações complementares de promoção do desempenho acadêmico.

Outra iniciativa que pode ser vista na UnB é o incentivo ao desenvolvimento das Tecnologias de Comunicação e Informação (TIC). Com o crescimento da demanda por cursos à distância e presenças remotamente, novos processos tem surgido demandando novas formas de organização acadêmica presencial e a distância. O Projeto Político Pedagógico Institucional (PPPI) da UnB, descreve que esses processos têm uma especificidade relacionados as TIC aplicados aos contextos educacionais.

Ainda segundo o PPPI, as ferramentas de tecnologia da informação e comunicação configuram estratégias metodológicas que atendem aos princípios fundamentais de ensino, pesquisa e extensão. Tendo como uma de suas diretrizes a promoção de atualização constante da infraestrutura para o uso eficiente e efetivo das TIC na UnB.

Neste cenário a UnB oferece 130 cursos presenciais, sendo 30 noturnos, que geram uma massa de informação que pode ser rastreada para explorar o comportamento do aluno com propósito de agregar valor ao ambiente acadêmico.

Diante disso, ter acesso aos dados que proporcionam indicadores do abandono na graduação, possibilita a descoberta de informações que podem auxiliar coordenadores e professores a diminuição das taxas de evasão.

### **1.3 Questão de Pesquisa**

Como utilizar os indicadores presentes nos sistemas acadêmicos institucionais na identificação proativa da evasão em cursos de graduação presenciais na UnB?

### **1.4 Objetivos Principal e Secundários**

O objetivo desse trabalho é realizar a previsão de evasão nos cursos de graduação da Universidade de Brasília (UnB), utilizando informações provenientes dos dados acadêmicos para a construção de modelos de aprendizagem de máquina que sustentem o processo de predição.

Dando suporte ao objetivo principal, foram definidos os objetivos específicos:

- Fazer uma Revisão Sistemática da Literatura (RSL) visando identificar os fatores que impactam a evasão;
- A partir dos fatores identificados definir quais indicadores estão presentes nos sistemas acadêmicos da UnB;

- Identificar os métodos de aprendizagem de máquina utilizados no processo de previsão de evasão.
- Implementar um sistema capaz de avaliar os indicadores existentes na UnB, que consiga prever quais alunos têm características de evasão.
- Avaliar o modelo proposto, por meio dos resultados do sistema implementado, e compará-lo com os trabalhos relacionados.

## 1.5 Metodologia

A pesquisa tem o propósito verificar os indicadores acadêmicos que impactam na evasão no ensino superior na Universidade de Brasília (UnB). Desta forma, foi classificada como aplicada e Segundo Gil et al. [17], esse tipo de pesquisa é direcionado à aquisição de conhecimento com vistas a aplicação numa situação específica.

O caráter exploratório da pesquisa é observado na busca de fatores que tornem mais explícitos os motivos que levam os alunos a apresentar características de evasão. Para explorar esses fatores, uma Revisão Sistemática da Literatura (RSL) foi realizada.

A pesquisa está sendo delineada como estudo de caso. Segundo Runeson et al. [18], o estudo de caso em engenharia de software é um inquérito empírico que se baseia em múltiplas fontes de evidência para investigar um caso de um fenômeno contemporâneo em um contexto real.

Os fatores de evasão identificados serão confrontados com os dados registrados nos sistemas acadêmicos da UnB, com propósito de torná-los indicadores passíveis de serem usados em análises quantitativas.

No decorrer do trabalho foi desenvolvido a estrutura para realizar o procedimento de coleta, tratamento e análise dos dados. O principal propósito dessa estrutura é servir de base para aplicabilidade dos conceitos e ferramentas identificados na pesquisa.

A avaliação comparativa dos resultados considera os trabalhos relacionados.

## 1.6 Estrutura da Dissertação

Este trabalho está organizado em 7 capítulos, com este incluso que compreende o contexto, motivação, questão de pesquisa, objetivos e metodologia adotada.

No Capítulo 2 — Fundamentação Teórica. Apresenta os conceitos teóricos sobre retenção e evasão, com base nas informações extraídas da Revisão Sistemática da Literatura (RSL). Descreve dos fatores de Retenção e Evasão relatados nos artigos, relaciona os algoritmos e ferramentas usadas no processo de previsão.

No Capítulo 3 — Contexto Organizacional. Descreve as unidades organizacionais envolvidas no escopo do trabalho. Apresenta um resumo das leis, normas e decretos que de alguma forma, direta ou indireta, contribuem como instrumento para a melhoria do ensino de graduação.

No Capítulo 4 — Mapeamento das Bases Educacionais. O propósito deste capítulo é descrever o procedimento realizado durante a análise nas bases de dados dos sistemas acadêmicos envolvidos neste estudo e buscar uma correlação entre os fatores de evasão identificados durante a revisão sistemática da literatura.

No Capítulo 5 — Modelo de Análise na Graduação — MAGRA. Este capítulo apresenta um modelo analítico para análises da evasão nos cursos de graduação presenciais da Universidade de Brasília (UnB).

No Capítulo 6 — Testes e Análises de Dados. O propósito deste capítulo é descrever os procedimentos executados durante a etapa de testes e análises dos dados.

No Capítulo 7 — Conclusão, apresentam-se as conclusões deste trabalho e trabalhos futuros.

# Capítulo 2

## Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica necessária ao desenvolvimento do presente trabalho. Os conceitos sobre a evasão e retenção são apresentados, bem como o desenvolvimento de uma revisão sistemática da literatura para explorar os fatores e indicadores relacionados a retenção no ensino superior.

### 2.1 Evasão

A pesquisa sobre a evasão nos cursos de ensino superior tem sido alvo de estudos por pesquisadores já há algum tempo [8, 9, 10, 11, 12].

O relatório de Vicent Tinto em 1975 [8], classifica a evasão em dois tipos de estudantes, aqueles abandonam a universidade e aqueles que abandonam o ensino superior.

O primeiro considera a pessoa que abandona a universidade onde está inscrito, trazendo para a esta instituição, dentre outras coisas, o uso ineficiente dos recursos, já escassos, para a instituição e a ocupação de uma vaga que poderia ser preenchida por outra pessoa.

O segundo trata da evasão para as pessoas desistem da universidade por não conseguirem o diploma em nenhuma outra instituição. Neste caso, o tratamento é mais amplo, considera a transferência de indivíduos entre instituições. Tinto [8] realça a impotência da criação de uma gama variada de programas educacionais e institucionais atendendo a diversos tipos de alunos.

O trabalho de John Bean em 1979 [9], buscava informações sobre as causas da evasão no ensino superior, desenvolveu um modelo para investigar determinantes desse fenômeno. Compara a rotatividade de alunos com a rotatividade nas organizações de trabalho, e busca semelhanças nas razões que levam os estudantes a deixar a instituição de ensino e as que levam os funcionários deixar as empresas [9].

Para Rumberger e Lim [12], uma vez que os estudantes se matriculam, dimensões sociais e acadêmicas influenciam a permanência desse estudante:

- A dimensão social que lida com a integração social dos estudantes com a instituição e com o valor da escolaridade. Acontece na interação do corpo docente e outros ambientes informais;
- A dimensão acadêmica que lida com a integração ou engajamento acadêmico dos estudantes em aprendizagem. Pode acontecer no sistema formal de aulas.

Rumberger e Lim [12], reforçam que a permanência na instituição depende a integralização em algum grau nas duas dimensões. A saída antecipada do estudante da instituição de ensino constitui uma situação problemática tanto para o aluno quanto para instituição.

Atif et al. [19], destacam que os problemas que afetam os alunos envolvem a carreira e seu potencial de renda no futuro. Para a instituição, as perdas estão ligadas à reputação, a renda e a oportunidade de contribuir para a vida do estudante. A força de trabalho perdida afeta diretamente a sociedade, deixando de contar com pessoas preparadas para o mercado global.

De acordo com Veloso e Almeida [20], os estudos começaram a ficar mais frequentes no Brasil, a partir de 1995, com a criação da Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras (ANDIFES) et al. [16]. Esse relatório reúne um conjunto de dados importantes sobre o desempenho das universidades públicas brasileiras. Esses dados contemplam os índices de diplomação, retenção e evasão dos estudantes em seus cursos de graduação. A comissão faz uma distinção do termo evasão em tipos:

- **evasão de curso:** quando o estudante se desliga do curso superior em situações diversas tais como: abandono (deixa de matricular-se), desistência (oficial), transferência ou mudança de curso, exclusão por norma institucional;
- **evasão da instituição:** quando o estudante se desliga da instituição na qual está matriculado;
- **evasão do sistema:** quando o estudante abandona temporária ou definitivamente o ensino superior.

Nesse relatório, a comissão adota como objeto de estudo a evasão institucional, onde é descrita como a saída definitiva do aluno de seu curso de origem, sem concluí-lo.

Outra definição pode ser vista no trabalho de Cislighi [21], onde objetivo era construir um modelo de gestão do conhecimento em uma instituição de ensino brasileira, a evasão foi considerada o fenômeno no qual um estudante ingressa em um determinado curso, não integraliza o currículo e, conseqüente, não é diplomado.

Entender o fenômeno da evasão tem sido importante para as instituições, contudo, conforme recomenda o estudo feito pela ANDIFES et al. [16], outros fatores devem ser avaliados em conjunto, como na diplomação e retenção.

As questões relacionadas a diplomação estão fora do escopo desse trabalho. A pesquisa fixa sobre o conceito da evasão, porém a próxima seção descreve a retenção por entender que é um processo que, em muitos casos, pode preceder a evasão.

## 2.2 Retenção

Essa seção descreve os conceitos e definições adotados por autores (inter)nacionais sobre o fenômeno da retenção nas instituições de ensino superior.

O estudo realizado por Lenning et al. [14], esclarece os conceitos sobre retenção em uma composição conceitual unificadora abordando vários aspectos. Os autores descrevem que a retenção pode ser definida como ocorrendo a partir do momento que os estudantes completam, continuam ou retomam seus estudos. Outra faceta exemplificada pelos autores e definida como caso tradicional, é a conclusão do curso no tempo para o qual o programa foi projetado, ou seja, um curso de bacharelado de quatro anos concluído nesse período. Dentre as definições discutidas pelos autores e grupos de pesquisa de educadores, há uma definição positiva e geral para a retenção: é o sucesso em alcançar alguma meta ou objetivo.

Conforme Aguiar et al. [22], em 2014, o entendimento sobre retenção mudou ao longo do tempo, esforços para resolver a questão estão onipresentes no ensino superior atualmente. O autor aponta o crescimento do assunto nos últimos anos e a existência de indícios claros da complexidade envolvida em ajudar o aluno a ter sucesso no seu curso.

Os pesquisadores Raju e Schumaker [23], avaliaram outros estudos que consideraram a retenção de um ano para o outro, porém no seu trabalho eles definiram que a retenção/sucesso do estudante é alcançada a partir do momento da chegada dele na graduação, que dura seis anos a partir da sua matrícula.

Atif et al. [19], fizeram um estudo, em uma instituição europeia, para desenvolver um modelo de previsão do sucesso acadêmico. Utilizou dados demográficos combinados com os dados dos sistemas acadêmicos e de gerenciamento dos cursos dos estudantes. Para os autores, a retenção trata da inscrição e da conclusão bem-sucedida dos estudos por um estudante.

Rintala et. al [24], Finlândia, apresentam um detalhamento mais abrangente sobre a retenção. Os autores dividem a retenção em quatro tipos:

- **Institucional:** é a medida da proporção de alunos que permanecem matriculados na mesma instituição de ano a ano.

- **De sistema:** mais concentrada no aluno, usa a persistência do sistema como medida. Um estudante que deixa uma instituição para frequentar outra não é considerado um não-persistente, mas como um persistente.
- **Dentro de uma área principal ou disciplina:** tem uma visão mais limitada do tópico ao ver a retenção em uma área principal de estudo, disciplina ou um departamento específico;
- **Dentro de um curso:** menor unidade de análise. É medida com a conclusão do curso.

No estudo de Pérez et al. [25], para modelar um sistema de previsão de retenção para os alunos do primeiro ano na Universidade Bernardo O'Higgins (Santiago do Chile), foram avaliadas variáveis participantes do processo de permanência com sucesso do estudante na instituição.

Para Pérez et al. [26], o estudante chileno é levado a desistir do curso a partir da combinação de várias variáveis de diversos tipos. Para os autores, o estudo dessas variáveis pode melhorar a retenção desses alunos nos cursos escolhidos.

De acordo com Pereira [27], as pesquisas nacionais empregam o termo retenção, quando estão relacionadas ao ensino superior.

Conforme Pereira e Carneiro [13], muitos estudos, tanto nacionais quanto internacionais, a retenção tem papel secundário, tendo em vista que a maioria tem como foco a evasão. Silva Garcia et al. [28] esclarece que, muitas vezes, estes problemas são abordados como fenômenos similares ou como causa um do outro.

Para Pereira et al. [29], veem razão em considerar que a retenção e evasão apresentam estreita relação entre si. Ambos ocorrem no ambiente universitário envolvendo o mesmo tipo de atores (estudantes, professores, familiares, colegas e outros envolvidos). Sendo assim, os autores entendem que os modelos aplicados nos estudos de evasão podem ser aplicados aos estudos de retenção observando as restrições de cada teoria.

No Brasil, o relatório elaborado pela ANDIFES et al. [16] define retenção como sendo a permanência nos cursos para além do tempo máximo da integralização curricular. O termo retido é descrito como a condição em que o aluno, apesar de esgotado o prazo máximo de integralização curricular fixado pelo Conselho Federal de Educação (CFE), atual Conselho Nacional de Educação (CNE), ainda não concluiu o curso, mantendo-se, entretanto, matriculado na universidade. Isso pode ser visto na pesquisa de Nascimento e Rios [30], onde foi delineado um panorama das causas de evasão e retenção nos cursos de engenharia. A retenção foi definida como a ultrapassagem ou superação do tempo de permanência do aluno no curso para além do previsto na grade curricular.

No estudo de Corrêa e Noronha [31], avaliando a existência de retenção nos cursos de Administração de uma universidade pública, e para Cislighi [21], onde foi proposto a construção e de um modelo de sistema de gestão do conhecimento baseado em um *framework* projetado para a melhoria da permanência de alunos no ensino superior, a retenção é tratada como permanência prolongada. Esses pesquisadores entendem que a permanência prolongada acontece quando o tempo de conclusão curso universitário, ultrapassa o tempo planejado pelo currículo ou o projeto pedagógico.

Para a Ney [32], o conceito de retenção pode ser ampliado. Sendo necessário acrescentar aos períodos letivos suficientes para a conclusão do curso o período permitido para realizar o trancamento. Desta forma, o pesquisador define que a retenção é a situação em que o aluno permanece matriculado no curso mesmo após o tempo suficiente para concluí-lo. Por suficiente, o autor explica que é o tempo de conclusão do curso, definido pela matriz curricular, acrescido do tempo de trancamento.

Com base nestes estudos internacionais é possível identificar uma conotação positiva no entendimento do termo retenção, sendo aplicado com o sentido de sucesso ao alcançar algum objetivo.

O significado do termo retenção, adotado pelos norte-americanos e europeus, é semelhante ao adotado pelos autores latino-americanos, retenção é a permanência na universidade, conforme anteriormente observado por Pereira [27].

Com base nas pesquisas nacionais apresentadas é possível perceber uma conotação negativa para o termo retenção. As definições usadas pela ANDIFES [16], Nascimento e Rios [30] estão atreladas ao conceito de extrapolação do tempo definido pela grade curricular.

Parte dos modelos teóricos construídos nas últimas cinco décadas foram desenvolvidos para analisar o fenômeno do abandono, no entanto, sua estreita relação com a retenção de estudante pavimentou a aplicação deles para estudos de retenção, mediante observações das características e restrições de cada teoria.

Este trabalho adota a seguinte definição de retenção ou forma sinônima permanência prolongada:

- **Retenção** é a condição que o estudante demanda um tempo maior do que o previsto na matriz curricular para a integralização da carga horária do curso.

Este trabalho a partir de agora senta-se no estudo da evasão, no entanto, a sua ligação estreita com a retenção norteia a aplicação dos mesmos fatores/indicadores para a retenção, analisando as características e limitações de cada teoria.

## 2.3 Indicadores de Evasão e Retenção

A partir da análise de Atif et al. [19], é possível perceber os motivos que levam o estudante a sair do ensino superior é resultado de uma combinação de fatores, com um grau de complexidade e interconectados que se desenvolvem ao longo do tempo de permanência.

Avaliando esses fatores preditores em um formato quantificável, entendendo que eles também podem ser agrupados, optou-se em empregar o termo indicador para representar esse contexto. Isso possibilita a mensuração objetiva e uma melhor representação sistemática no contexto educacional.

Neste contexto, com o propósito de melhorar o entendimento sobre os indicadores que podem causar a evasão, foi realizada uma Revisão Sistemática da Literatura (RSL) para identificar fatores que contribuem para a evasão e retenção no ensino superior. Nesta mesma Revisão Sistemática da Literatura (RSL) foram coletadas informações que ajudaram a identificar ferramentas e algoritmos usados para análises dos dados e desenvolvimento de modelos de predição. O procedimento de planejamento de condução da RSL está descrito no Apêndice A.

### 2.3.1 Fatores de Evasão e Retenção

Esta seção responde à primeira questão da revisão sistemática (QP1): quais fatores estão associados à previsão de retenção e/ou evasão no ensino superior?

Os trabalhos selecionados na RSL foram avaliados em busca de informações que contribuem para identificação dos fatores envolvidos no processo de retenção e evasão na educação.

As pesquisas realizadas por Esmat e Pitts [33] e Tucker e McKnight [34] preconizam que os fatores que levam a retenção ou evasão do ensino superior estão ligados a falta conhecimento base adquirido antes da graduação.

No caso do artigo de Atieh et al. [35], o objetivo foi avaliar os hábitos dos estudantes dos cursos de tecnologia tendo como base as habilidades de estudos superficial e profunda: na superficial o aluno tenta decorar novos assuntos sem associá-los; na profunda o aluno visa associar informações com conhecimento previamente adquiridos.

Cohen e Kelly [36], entendem que o desempenho pode vir associando a uma disciplina. Eles exploraram a relação dos alunos matriculados nos cursos de Educação da Ciência, Tecnologia, Engenharia e Matemática, do inglês *The education of science, technology, engineering, and mathematics (STEM)*, e o desempenho desses alunos nas disciplinas de matemática. Apenas metade dos alunos que cursam pré-cálculo na faculdade avançaram para o cálculo I e destes que fizeram Cálculo I apenas 40% se matricularam em Cálculo

II. O estudo ainda mostrou que mais de 40% dos estudantes com notas baixas em Cálculo I tramitem a mensagem que não conseguem lidar com trabalhos mais avançados.

Cano e Leonard [37], avaliam que os fatores sociodemográficos desempenham um papel importante como preditores do sucesso acadêmico.

Para Darabi et al. [38], a inclusão do contexto socioeconômico no processo de admissão melhora o processo de avaliação dos alunos.

Pelo grande volume de informações que estão sendo registradas nos sistemas educacionais, os autores Sadati, Libre [39] e ElAtia [40] defendem que estes sistemas estão se tornando fonte crescente de informações úteis para prever o comportamento dos alunos.

Os autores Raju e Schumacker [23], consideraram a média de notas do ensino médio, o GPA, como tendo uma correlação significativa com a persistência universitária.

### 2.3.2 Identificação dos Fatores

A presente seção responde à segunda pergunta da revisão sistemática (QP2): quais são os fatores usados para prever a evasão do ensino superior?

A identificação e classificação dos fatores foi realizada com objetivo de definir o contexto no qual cada fator é empregado.

Um total de 29 fatores foram identificados e apresentados no artigo, escrito pelo grupo de pesquisa que executou a RSL. O artigo foi publicado nos anais da 56<sup>a</sup> Conferência Internacional do Havaí sobre Ciências do Sistema (HICSS), em 2023 [1].

A Tabela 2.1 traz a lista desses fatores, sendo os 10 mais citados: pontuação média, gênero, notas do curso, curso, idade, etnia, bolsa de estudos, código postal, estudo regular, horário do curso e tipo de entrada na universidade.

Para Atif et al. [19], a evasão é calculada com base em taxas definidas considerando variáveis demográficas, acadêmicas, motivacionais, institucionais e de interação. Elas consideram também a personalidade, valor dos estudantes e aspirações iniciais dos estudantes. Os pesquisadores descrevem que essa variedade de variáveis torna o cálculo das taxas complexo.

Considerando as características de cada um dos fatores relacionados neste estudo, foi feita uma categorização desses fatores em: demográfico; acadêmico e de aprendizagem.

Os fatores demográficos totalizam 11 e estão dispostos na Figura 2.1. No caso dos fatores acadêmicos também com total de 11 e descritos na Figura 2.2. Para os de aprendizagem foram identificados 7 fatores conforme mostra a Figura 2.3

A pontuação média dos alunos foi o fator mais utilizado pelos trabalhos que avaliam casos de evasão e retenção no ensino superior. Para Darabi et al. [38] e Al-Sheeb et al. [41], a pontuação média do curso foi usada para prever a pontuação dos candidatos às vagas em faculdades de engenharia. Os estudos de Cohen e Kelly [36], Al-Sheeb et al. [41]

Tabela 2.1: Fatores identificados na literatura - Fonte: Silva et al. [1].

| <b>Fator</b>               | <b>Código do Artigo referente a Tabela A.4</b>   | <b>Total</b> |
|----------------------------|--|--------------|
| pontuação média            | A-2, A-9, A-14, A-16, A-18, A-19, A-20, A-22, A-23, A-24, A-25, A-27, A-29, A-30, A-31, A-33, A-34, A-36, A-37, A-38, A-39, A-41, A-43, A-44, A-45, A-51, A-55 | 27           |
| gênero                     | A-2, A-5, A-6, A-8, A-9, A-11, A-14, A-15, A-16, A-18, A-20, A-22, A-23, A-25, A-26, A-37, A-38, A-40, A-45, A-47, A-48, A-50, A-55                            | 23           |
| notas do curso             | A-2, A-5, A-11, A-12, A-16, A-17, A-18, A-19, A-22, A-23, A-33, A-34, A-37, A-39, A-45, A-51, A-57, A-58   | 18           |
| curso                      | A-1, A-9, A-18, A-22, A-25, A-27, A-30, A-31, A-33, A-37, A-40, A-44, A-45, A-46, A-48, A-50, A-51   | 17           |
| idade                      | A-2, A-8, A-9, A-11, A-14, A-15, A-17, A-19, A-20, A-22, A-23, A-25, A-48, A-50  | 14           |
| etnia                      | A-2, A-5, A-13, A-15, A-18, A-26, A-29, A-37, A-38, A-45, A-46, A-48, A-55   | 13           |
| bolsa de estudos           | A-1, A-2, A-6, A-9, A-13, A-17, A-22, A-26, A-29, A-39, A-47, A-48   | 12           |
| código postal              | A-2, A-6, A-9, A-14, A-16, A-19, A-26, A-37, A-47, A-54  | 10           |
| estudo regular             | A-5, A-9, A-10, A-13, A-23, A-25, A-44, A-58   | 8            |
| horário do curso           | A-4, A-13, A-17, A-36, A-38, A-44, A-48, A-51  | 8            |
| tipo de entrada            | A-16, A-17, A-19, A-22, A-23, A-50, A-52, A-54   | 8            |
| emprego                    | A-2, A-6, A-8, A-11, A-14, A-26, A-47  | 7            |
| matérias aprovadas         | A-4, A-13, A-18, A-22, A-27, A-35, A-49  | 7            |
| nível educacional dos pais | A-1, A-2, A-6, A-14, A-15, A-41  | 6            |
| acesso a fórum             | A-10, A-11, A-12, A-37, A-58   | 5            |
| acesso ao sistema          | A-12, A-36, A-37, A-56   | 5            |
| resolução de exercícios    | A-10, A-12, A-37, A-58   | 4            |
| acesso a vídeo             | A-10, A-21, A-58   | 3            |
| conhecimento prévio        | A-11, A-19, A-23   | 3            |
| estado civil               | A-9, A-11, A-48  | 3            |
| estrangeiro                | A-17, A-22, A-45   | 3            |
| matérias reprovadas        | A-13, A-17, A-19   | 3            |
| necessidades especiais     | A-36, A-41, A-52   | 3            |
| trancamento                | A-13, A-24, A-35   | 3            |
| curso matriculado          | A-12, A-13   | 2            |
| dados de pesquisa          | A-14, A-39   | 2            |
| filhos                     | A-8, A-11  | 2            |
| apresentação tardia        | A-10   | 1            |
| estágio                    | A-13   | 1            |

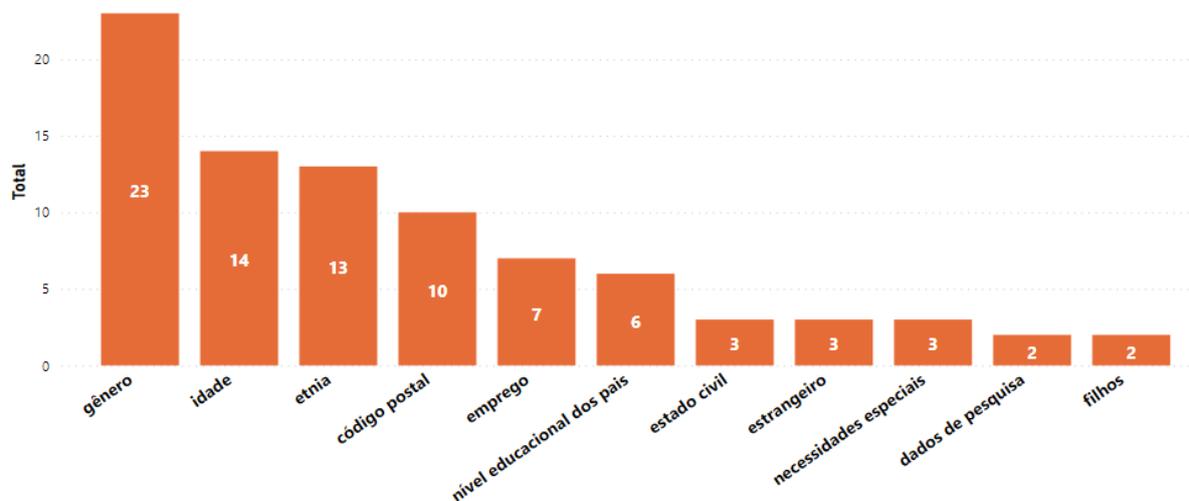


Figura 2.1: Fatores Demográficos - Fonte: Silva et al. [1].

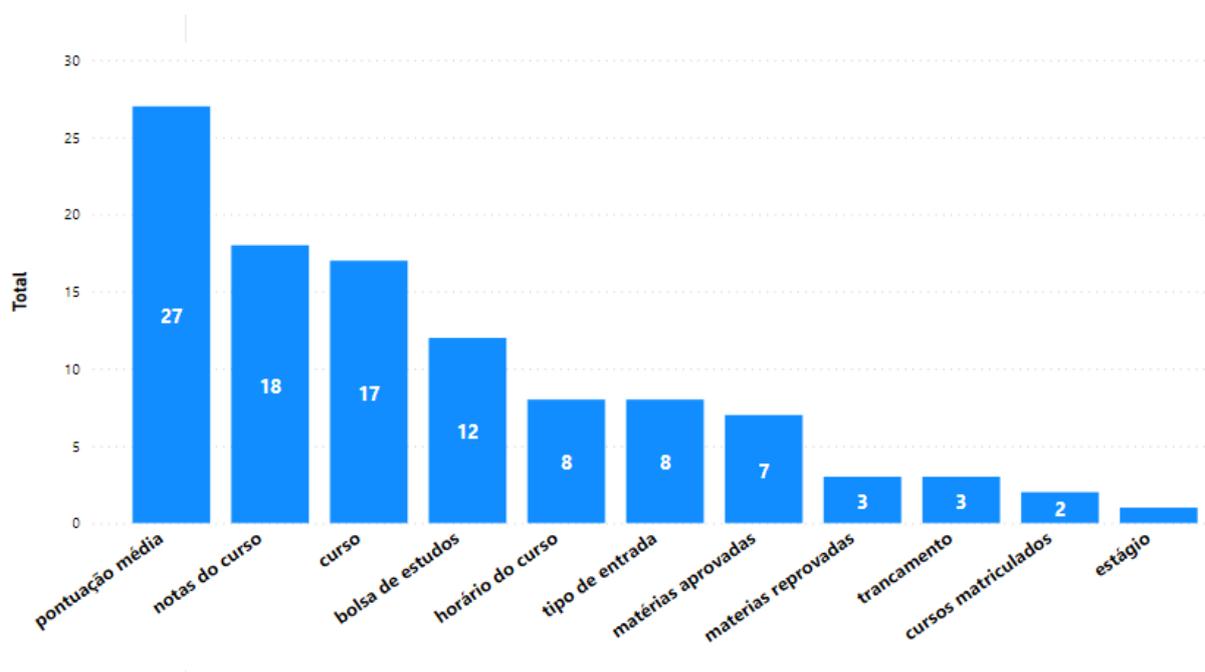


Figura 2.2: Fatores Acadêmicos - Fonte: Silva et al. [1].

avaliam a retenção de alunos em cursos de exatas tomando como base a pontuação média em matemática e as notas médias do primeiro semestre.

Os pesquisadores Frischenschlager et al. [42], avaliaram o comportamento dos alunos em uma Universidade na Áustria e identificaram que os alunos do sexo masculino, fluentes no alemão e bom desempenho no ensino médio tinham melhores índices de sucesso acadêmico.



Figura 2.3: Fatores de Aprendizagem - Fonte: Silva et al. [1].

Aguiar et al. [22], avaliando o uso de portfólios eletrônicos, perceberam que características dos alunos se correlacionam, podendo ajudar a melhorar a precisão da previsão da evasão. Nas análises de correlação, o gênero apresentou um ganho de informação em relação a outros fatores.

Para Raju e Schumacker [23], o gênero e etnia são fatores regularmente usados como preditores importantes de retenção que leva a graduação.

As notas dos testes padronizados, como o SAT<sup>1</sup>, são usadas prever o desempenho dos alunos [43, 44]. No artigo de Reed et al. [45], a taxa de sucesso dos calouros apresentou uma correlação quando as notas do SAT eram acima de 1000 pontos.

Os resultados da pesquisa de Tharp [46], mostraram que os estudantes matriculados em cursos de dois anos tinham melhores índices de persistência do que os estudantes de bacharelado. As horas do primeiro semestre e a média de pontos do primeiro semestre, apresentaram um valor significativo em sua capacidade de prever a evasão escolar.

Beck e Milligan [47], avaliaram grau de compromisso institucional que os alunos apresentavam em relação à instituição. Foram avaliadas questões como: lealdade, confiança na escolha da universidade e o quanto o aluno estava integrado ou satisfeito com as condições de ensino apresentadas pela universidade. As análises de correlação mostraram que os estudantes mais velhos apresentavam valores do compromisso institucional mais elevado.

Para identificar fatores relacionados ao sucesso dos estudantes de medicina, Frischenschlager et al. [42], avaliaram que a maturidade, aliado com outros fatores, apresentou

<sup>1</sup>*Scholastic Aptitude Test* (SAT): <https://www.studentprogress.org/gre/scholastic-aptitude-test-sat/>

influência significativa.

O local de residência do aluno foi um dos fatores que desempenhou um papel importante na pesquisa de Campbell e Mislevy [48]. As mulheres que não residem na mesma localidade da instituição de ensino, têm maior probabilidade de serem transferidas em relação à possibilidade de dar continuidade no curso na instituição.

Esta seção descreveu alguns dos fatores encontrados na revisão sistemática. Para fazer o tratamento desses fatores são necessárias técnicas que facilitem a coleta e análise informações. A próxima seção detalha as ferramentas usadas pelos trabalhos analisados.

### **2.3.3 Fatores de retenção e Ferramentas de Previsão**

Esta seção responde à questão da revisão sistemática (QP3): como os fatores são utilizados no processo de previsão?

A abundância de dados muitas vezes extrapola a capacidade humana de análise desses dados, sendo necessário recorrer a ferramentas que auxiliem nesse processo. As ferramentas identificadas nos trabalhos atuam no contexto de uso de diferentes tipos de algoritmos (modelos) para a previsão e softwares suportam a construção desses algoritmos com o objeto de facilitar a exploração dos dados.

As seções seguintes descrevem os algoritmos e softwares para o suporte a construção e análise desses modelos. Foram identificados 13 algoritmos na RSL, conforme Tabela 2.2.

Segundo Eubanks et al. [49], o ensino superior tem a exigência de informações mais precisas para a tomada de decisão. Com o grande volume de dados sendo armazenados nos sistemas educacionais é preciso usar de técnicas sofisticadas para a coleta e síntese desses dados. Pensando nisso, pesquisadores tem mostrado interesse em usar a mineração de dados para identificar e investigar questões relacionadas ao ambiente educacional. Baker et al. [50], descrevem que a partir desse contexto surge a área de pesquisa conhecida como Mineração de Dados Educacionais.

Conforme indicado por Raju e Schumacker [23], a mineração de dados é a extração não trivial de informações implícitas, previamente desconhecidas, e potencialmente úteis a partir de dados para transformar em informação útil para o ensino superior.

Para Tekin [44], a mineração de dados é considerada com um aspecto interdisciplinar da ciência da computação, busca padrões baseados em computação em grandes conjuntos de dados, envolvendo vários métodos de inteligência artificial, aprendizagem de máquina, estatística e sistemas de banco de dados.

A Figura 2.4 dá uma visão geral dos algoritmos usados nos trabalhos. Para uma melhor visualização não foram considerados no gráfico os artigos que não reportaram a utilização de algoritmos de aprendizagem de máquina e/ou similares.

Tabela 2.2: Algoritmos identificados na literatura.

| <b>Algoritmos</b>               | <b>Código do Artigo referente a Tabela A.4</b>   | <b>Total</b> |
|---------------------------------|--|--------------|
| <i>Regression</i>               | A-1, A-3, A-4, A-5, A-6, A-8, A-10, A-11, A-13, A-14, A-15, A-16, A-18, A-20, A-23, A-24, A-25, A-26, A-30, A-32, A-33, A-37, A-38, A-43, A-44, A-45, A-46, A-48, A-51, A-52, A-54, A-55, A-57 | 33           |
| <i>Decision trees</i>           | A-1, A-7, A-8, A-9, A-11, A-12, A-17, A-22, A-23, A-38, A-41, A-49, A-53   | 13           |
| <i>Neural Networks</i>          | A-9, A-22, A-34, A-37, A-38, A-44, A-45, A-49, A-53  | 9            |
| <i>Random Forest</i>            | A-9, A-10, A-11, A-22, A-23, A-37, A-50  | 7            |
| <i>Naive Bayes</i>              | A-11, A-27, A-37, A-43, A-53   | 5            |
| <i>Support Vector Machines</i>  | A-10, A-23, A-34, A-49   | 4            |
| <i>Bayesian networks</i>        | A-11, A-19   | 2            |
| <i>C4.5</i>                     | A-7, A-12  | 2            |
| <i>K-NN</i>                     | A-11, A-53   | 2            |
| <i>Boosting and AdaBoost</i>    | A-49   | 1            |
| <i>Cluster analysis</i>         | A-6  | 1            |
| <i>Extreme Learning Machine</i> | A-34   | 1            |
| <i>Gradient Boosted Tree</i>    | A-22   | 1            |
| não reportado                   | A-2, A-21, A-28, A-29, A-31, A-35, A-36, A-39, A-40, A-42, A-47, A-56, A-58  | 13           |

Nos tópicos seguintes são detalhados os três algoritmos mais citados na literatura que compõe a RSL.

## Regressão

A regressão é um componente importante na análise de dados. Seu objetivo é descrever a relação entre uma variável de resposta (dependente) e uma ou mais variáveis explicativas (independente). O modelo de regressão linear é o modelo comum de modelagem para as análises onde a variável de resposta é contínua. Para estudos onde a variável de resposta é dicotômica, a regressão logística é a mais indicada (Hosmer et al.) [2].

Conforme Hosmer et al. [2], os métodos empregados em uma análise utilizando regressão logística seguem, mais ou menos, os mesmos princípios gerais usados na regressão linear. A Figura 2.5 apresenta um exemplo resultando da modelo regressão logística, avaliando a porcentagem de sujeitos com problema no coração (CHD) em cada grupo de idade (AGE).

Pesquisas realizadas por Dey e Astin [51], na década de 90, conduziram estudos comparativos entre as técnicas logísticas e probit em relação à regressão linear na análise da retenção no ensino superior. Tharp [46], realizando um estudo longitudinal sobre evasão,

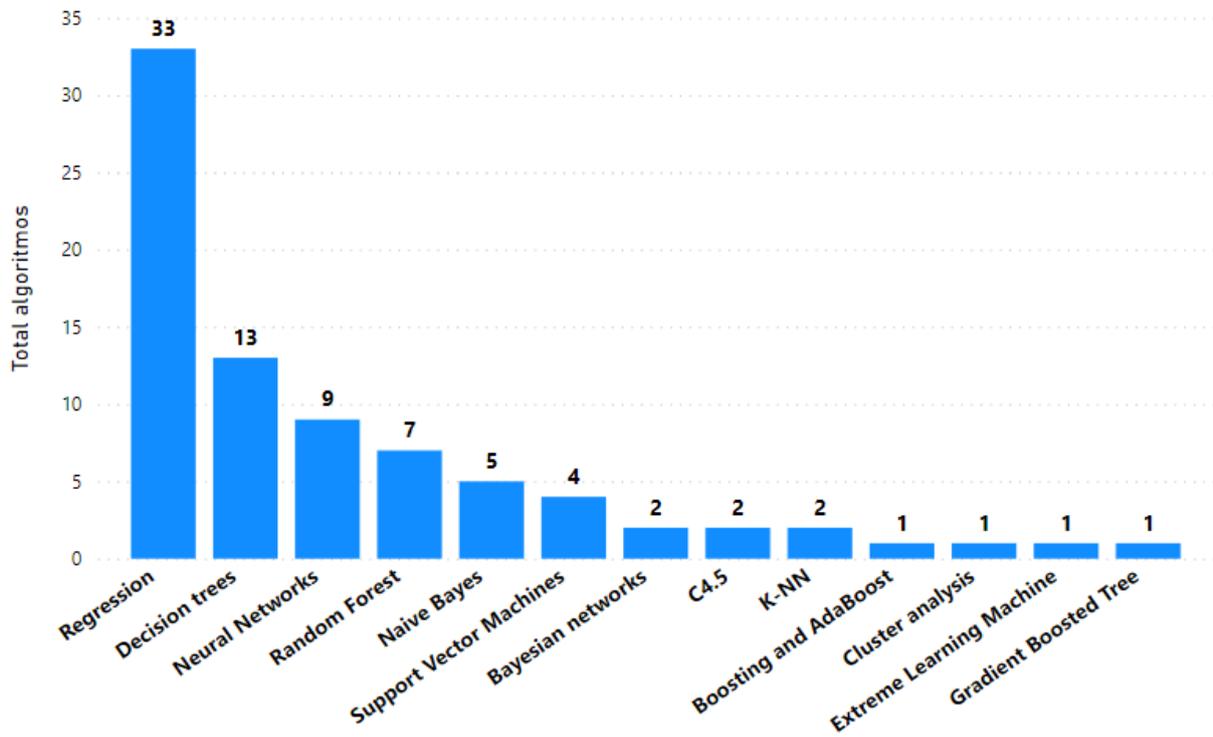


Figura 2.4: Algoritmos utilizados nos artigos.

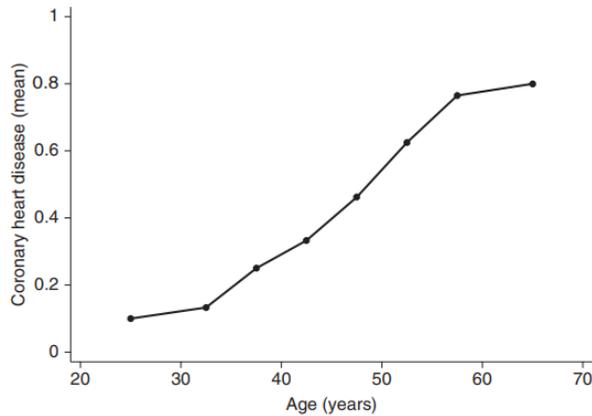


Figura 2.5: Exemplo do gráfico da regressão logística. Fonte: Hosmer et al. [2].

definiu que a técnica de regressão múltipla foi necessária para avaliar a importância individual das variáveis independentes, assim como a capacidade combinada de todas as variáveis.

Com base no entendimento da condição binária da situação da retenção e evasão, Reed et al. [45], considera o uso do método estatístico regressão logística mais apropriado para

esses estudos educacionais.

Publicações recentes mostram que técnicas modernas de regressão são usadas nas análises do desempenho acadêmico. Esmat e Pitts [33], usaram a regressão para examinar a relação entre o desempenho em cursos de educação geral e o desempenho em cursos de Ciência do Exercício, na universidade Kennesaw State University (KSU), Estados Unidos. Al-Sheeb et al. [41], tinham como um dos seus objetivos examinar a eficiência de combinar uma técnica clássica de seleção de variáveis com a Regressão Ridge. Para Litzler e Young [52], Campbell e Mislavy [48], alguns padrões de comportamento não podem ser analisados por técnicas tradicionais e optaram pelo uso da Regressão Logística Multinomial.

Segundo Raju e Schumacker [23], regressão logística e árvore de decisão são ótimos para identificação de variáveis importantes. As médias das notas do ensino médio foram classificadas como variáveis significativas para um conjunto de dados pré-universitário e universitário.

## Árvores de Decisão

Segundo Rokach e Maimon [3], árvore de decisão é um classificador composto por um conjunto de nós formando uma árvore enraizada, ou seja, é uma árvore dirigida com um nó chamado “raiz” que não tem bordas de entrada. Os outros nós têm bordas de saída sendo chamados nós internos. As folhas são chamadas terminais e representa o valor alvo mais apropriado.

A Figura 2.6 mostra um exemplo de árvore de decisão, que descreve se um cliente responderá ou não uma mala direta. Os círculos são os nós internos e as folhas são representados pelos triângulos.

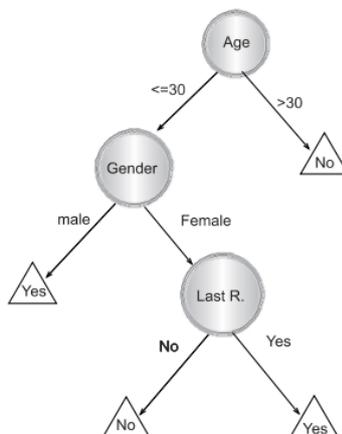


Figura 2.6: Exemplo de árvore de decisão. Fonte: Rokach e Maimon [3].

O algoritmo árvores de decisão foi o segundo algoritmo mais usado nos artigos. Para Figueroa-Cañas e Sancho-Vinuesa [53], o objetivo principal para a escolha da Árvore de decisão foi a possibilidade de ter um classificador com bom desempenho, que fosse interpretado facilmente. Para Cano e Leonard [37], a preferência pelos modelos baseados em árvores de decisão vem do fornecimento uma justificativa e explicação do raciocínio por trás da decisão, tornando o processo decisório compreensível para os seres humanos.

O estudo de Gustian e Hundayani [54], realizou um processo de classificação com a árvore de decisão (C4.5), para identificar alunos com características de evasão. Os resultados obtidos mostram um desempenho de acurácia nos dados de teste de 97.67%, precisão de 96.5% e o percentual de verdadeiros positivos de evasão de 93,33%.

## Redes Neurais

Para Aggarwal [4], as redes neurais artificiais são técnicas de aprendizagem de máquinas que simulam o mecanismo de aprendizagem em organismos biológicos. As unidades computacionais são chamadas neurônios, são conectados mutualmente por meio de pesos, que fazem o mesmo papel da sinapse nos organismos biológicos. Os pesos afetam a função computada no neurônio. Os dados de treinamento são os estímulos externos que dão *feedback* para a correção dos pesos na rede neural. A correção dos pesos influencia na probabilidade da classificação correta em comparação a classificação fornecida pelos dados de treinamento. A Figura 2.7 mostra a estrutura básica da rede neural artificial.

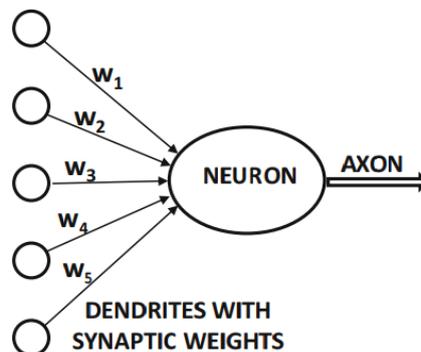


Figura 2.7: Estrutura básica da rede neural artificial. Fonte: Aggarwal [4].

Segundo Tekin [44], as redes neurais são métodos matemáticos com inspiração biológica capazes de modelar funções não lineares. São compostas por neurônios artificiais conectados, que processam a informação por meio de uma abordagem conexionista. Podem ser usando para classificação quanto para regressão. Para Reed et al. [45], as redes neurais produzem resultados de previsão/classificação similares ou melhores comparando

Tabela 2.3: Ferramentas identificadas na literatura.

| Software                 | Código do Artigo referente a Tabela A.4   | Total |
|--------------------------|---|-------|
| IBM SPSS                 | A-1, A-3, A-5, A-9, A-14, A-15, A-20, A-25, A-28, A-32, A-35, A-36, A-42, A-44, A-46, A-58  | 16    |
| R                        | A-8, A-12, A-20, A-23, A-53, A-54   | 6     |
| Weka                     | A-11, A-41  | 2     |
| JCLAL                    | A-11  | 1     |
| Python                   | A-4   | 1     |
| Python/Scikit            | A-39  | 1     |
| SAP Predictive Analytics | A-8   | 1     |
| SAS                      | A-38  | 1     |
| Não reportado            | A-2, A-6, A-7, A-10, A-13, A-16, A-17, A-18, A-19, A-21, A-22, A-24, A-26, A-27, A-29, A-30, A-31, A-33, A-34, A-37, A-40, A-43, A-45, A-47, A-48, A-49, A-50, A-51, A-52, A-55, A-56, A-57 | 32    |

com a regressão logística. Contudo, os modelos de redes neurais necessitam de uma mostra grande de treinamento, o que pode inviabilizar estudos com um número limitado de dados.

### Ferramentas de suporte aos algoritmos

Nos artigos da RSL foram identificadas 8 ferramentas de suporte ao desenvolvimento dos algoritmos. A Tabela 2.3 mostra o total de ferramentas usadas e os artigos que referir-se a essas ferramentas. O IBM SPSS e o R foram o dois que apresentaram uma preferência de uso pelos artigos. O IBM SPSS<sup>2</sup> é um software da IBM com um conjunto de ferramentas para análise de dados, com um custo de uso. O R<sup>3</sup> é uma linguagem de programação, de código aberto, que proporciona um ambiente para desenvolvimento de ferramentas computacionais que envolvam conceitos estatísticos e apresentação de resultados na forma gráfica (Silva et al.) [55].

#### 2.3.4 Lista final de indicadores

Esta seção apresenta os 29 indicadores categorizados e relacionados ao sistema educacional que dá suporte ao indicador identificado. A Figura 2.8 mostra a distribuição dos indicadores e o mapeamento da origem desses indicadores em relação aos sistemas da UnB.

A maioria dos indicadores são sustentados pelo sistema SIGAA, estes foram classificados em demográficos e acadêmicos, nas cores azul e amarelo, respectivamente. Indicadores

<sup>2</sup>IBM <https://www.ibm.com/br-pt/analytics/spss-statistics-software>

<sup>3</sup>R <https://cran.r-project.org/>

demográficos como estado civil, nível escolar dos pais e emprego não foram encontrados nos sistemas em operação na UnB.

Os indicadores de aprendizagem, na cor rosa, estão relacionados aos Ambientes Virtuais de Aprendizagem (AVA), como, por exemplo, o Aprender 3.

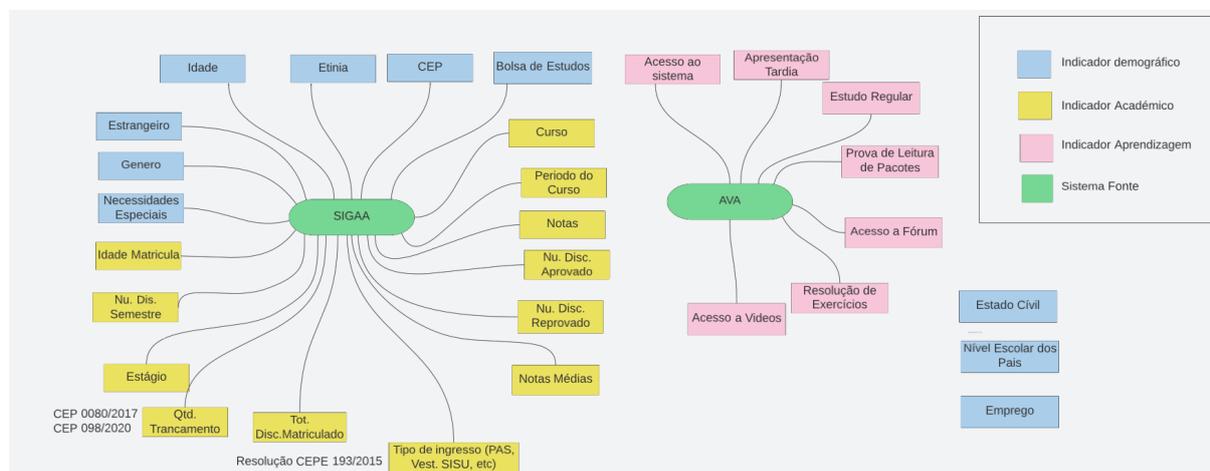


Figura 2.8: Visão geral dos indicadores - Fonte: Elaborado pelo autor.

## 2.4 Trabalhos Relacionados

O estudo realizado na Universidade de Curtin por Freitas et al. [5], reuniu um conjunto de mais de 51 mil registros de alunos visando explorar como aumentar a permanência dos estudantes. Esse grande volume de informações foi submetido a várias metodologias de tratamento de dados para garantir que esse conjunto de registros pudesse ser submetido a um conjunto de hipóteses inter-relacionadas para enfrentar uma série de ideias e preconceitos sobre as influências na retenção de estudantes.

Os autores descrevem que a pesquisa estabeleceu a criação um modelo para agrupar estudantes vulneráveis que provavelmente deixarão a universidade tão cedo quanto o primeiro semestre de estudos. Esse modelo analítico de aprendizagem (MAP) teve o objetivo de prover a melhoria da experiência do estudante, diminuindo a evasão e fornecer uma estrutura baseada em evidências para criação de cursos educacionais e apoio personalizado ao estudante.

Os autores reforçam o entendimento que esse modelo foi usado na criação de um sistema de painel de controle para admissão de tutores e de serviços de apoio ao estudante. A Figura 2.9 apresenta a estrutura do modelo usando a abordagem qualitativa e quantitativa proposta pelo autor.

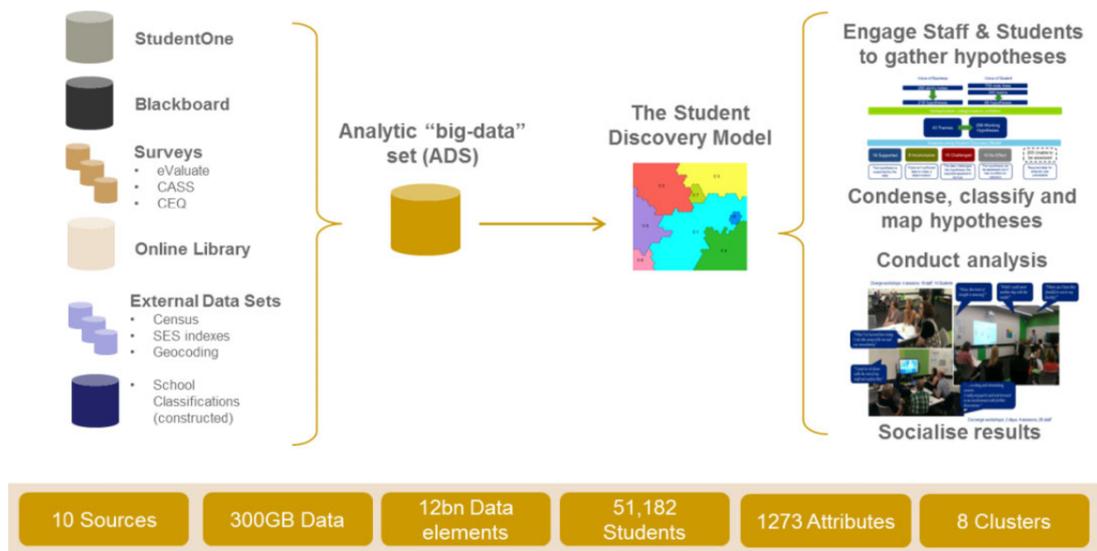


Figura 2.9: Metodologia de estudo: utilizando abordagens qualitativas e quantitativas - Fonte: Freitas et al. [5].

O principal objetivo desse estudo foi reduzir as taxas de atrito estudantil e aumentar as taxas de graduação ano a ano até 90% em 2017. Estudos foram feitos sobre a evasão destinado identificar as principais causas de abandono. Como parte da estratégia do estudo, foi implementado um conjunto de dados altamente granular para análise e geração relatório de evasão. Essa estrutura foi chamada de Modelo de Descoberta de Estudantes (MDE). Esse modelo usa um algoritmo de rede neural semi-supervisionado para e colocar os estudantes em grupos de perfil comportamental.

O conjunto de dados, desenvolvido para analisar o comportamento dos alunos, disponibilizou 273 atributos usados para treinamento do algoritmo e colocar os alunos em um mapa auto-organizado. Esse conjunto de dados foi extraído de 10 fontes de informação. O sistema de gerenciamento de aprendizado (*BlackBoard*) forneceu dados sobre o uso dos trabalhos e matérias do curso por parte dos estudantes, o sistema de gerenciamento do estudante (*Student One*) forneceu dados de inscrição e resultados acadêmicos.

Freitas et al. [5], incluíram na pesquisa outras fontes de informações como, por exemplo, dados socioeconômicos e de geolocalização, objetivando avaliar hipóteses relacionados a distância da residência até o campus e se isso impacta no tempo de estudo como medida de equilíbrio entre trabalho-vida-escola.

Após a execução do algoritmo, foram identificados grupos que foram nomeados pelos interessados no projeto em: primeiro ano de experiência; grupo em risco; estudantes internacionais; e estudantes regular. Esses grupos foram mantidos durante os testes de hipóteses e na análise dos fatores e atributos associados à evasão.

Para as hipóteses, o autor selecionou 5 grupos, sendo eles: educação prévia; formação cultural; envolvimento *on-line*; mentalidade do estudante e semelhança com o resto da coorte. Após os testes foram avaliadas as hipóteses: Se um estudante internacional poderia afetar negativamente o atrito. O estudo indicou que os estudantes internacionais tinham taxa de graduação maior que os locais. O maior uso de material *on-line* e presença no local, reduziu o atrito. Alunos de escolas privadas estão mais bem preparados para estudar e terão maiores taxas de aprovação não teve efeito no modelo. As taxas de graduação de alunos de escolas públicas e privadas foram quase idênticas. Alunos satisfeitos com o seu desempenho acadêmico estão menos inclinados a ter atrito. Hipótese apoiada pelo estudo. Estudantes com idade próxima à idade média da coorte tem maior probabilidade de retenção. Hipótese apoiada pelo estudo.

Com base neste estudo foi desenvolvido um Modelo Analítico de Aprendizagem (LAM) constando de 9 princípios que podem ser utilizados por outra instituição:

1. Desenvolver uma estratégia analítica de aprendizagem.
2. Compromisso de criar uma infraestrutura para uma grande integração de dados
3. Serviço centrado no aluno.
4. Olhar dinâmico para a jornada de aprendizagem dos estudantes (por exemplo, marketing, admissões, recrutamento, matrícula, retenção, formatura e emprego).
5. Modelo adaptativo de comportamento do usuário.
6. Ligando a análise da aprendizagem em um contexto dinâmico mais amplo.
7. Formação de hipóteses qualitativas com base na multidão.
8. Rigorosa visão de ética e adesão aos mais altos padrões de procedimentos éticos.
9. Processos de revisão externa e interna e de validação cruzada.

O autor concluí que os resultados foram significativos, confirmando a construção de um conjunto de indicadores, podendo ser utilizado para fornecer um painel de controle simples para tutores, serviços de admissão e de apoio ao estudante. Isso possibilitaria a implementação de serviços de resposta adaptativa para apoiar estudantes vulneráveis.

O processo de análise baseado em múltiplas fontes de dados traz mais força ao processo de identificação dos alunos em dificuldades, e o autor acrescenta que a e realização de análises com diferentes algoritmos pode melhorar os resultados encontrados na pesquisa.

Outro estudo feito por Adejo e Conolly [6], no Reino Unido, traz dados disponibilizados pelo *Higher Educational Statistical Agency* (HESA), onde taxa de abandono escolar nos últimos cinco anos tem aumentado especialmente para os estudantes no primeiro ano.

Técnicas de mineração de dados têm sido usadas em diferentes áreas profissionais: finanças, marketing, saúde, engenharia, governo, segurança [6]. Na educação não poderia ser diferente. Pesquisas com foco na busca de padrões e relacionamento significativos, permite a extração de conhecimento e a valorização dos dados. A partir desses dados encontrados nas bases de dados ensino superior é possível realizar a descoberta de padrões estudantis, visando a previsão desempenho acadêmico. Esse processo é conhecido como Mineração de Dados Educacionais, do inglês *Educational Data Mining* (EDM).

Os autores descrevem que esses procedimentos são usados para a melhoria do setor educacional, contudo a eficiência e a precisão do desempenho ainda são insuficientes para garantir a identificação precoce dos alunos. Para melhorar essa eficiência, o autor recomenda inclusão de mais variáveis, sendo possível até o uso de dados fora das bases de dados da universidade. Esses fatores externos são dependentes dos estudantes e essas informações precisam de pesquisas que envolvam o aluno para serem coletados.

Desta forma, os autores acrescentam que esses fatores que afetam e determinam o desempenho do estudante não estão isolados, estão interligados, inter-relacionados e interdependentes. A Figura 2.10 ilustra o modelo conceitual de interligação das variáveis preditoras.

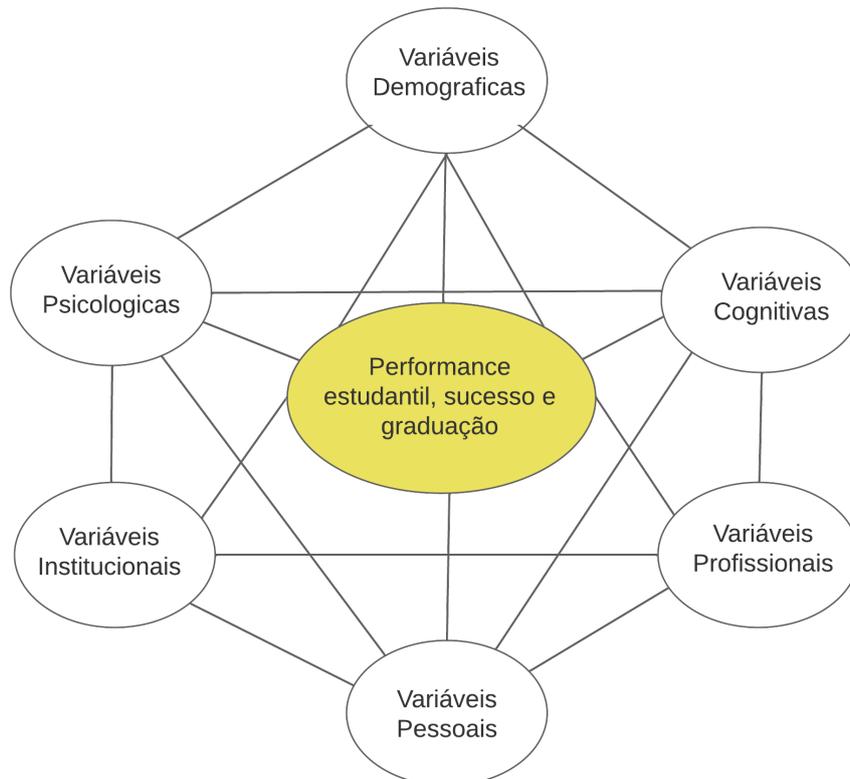


Figura 2.10: Modelo conceitual de interligação das variáveis preditoras (Adaptado de Adejo e Conolly [6])

Para o alcance dos resultados esperado no processo de predição, Adejo e Connolly [56], propõem uma estratégia baseada um framework que seja abrangente e holístico em sua abordagem. A Figura 2.11 mostra o framework que compreende três camadas:

1. A camada de interface com o usuário — pode ser chamada de camada de visualização, contempla as interfaces gráficas de usuário;
2. A camada do sistema de banco de dados — fornece acesso aos diferentes bancos de dados disponíveis no repositório de instituições de ensino superior de onde ocorre a abstração de dados para análise posterior e
3. A camada de execução ou sistema especializado — consiste em diferentes unidades para modelagem, avaliação e recomendação de decisão.

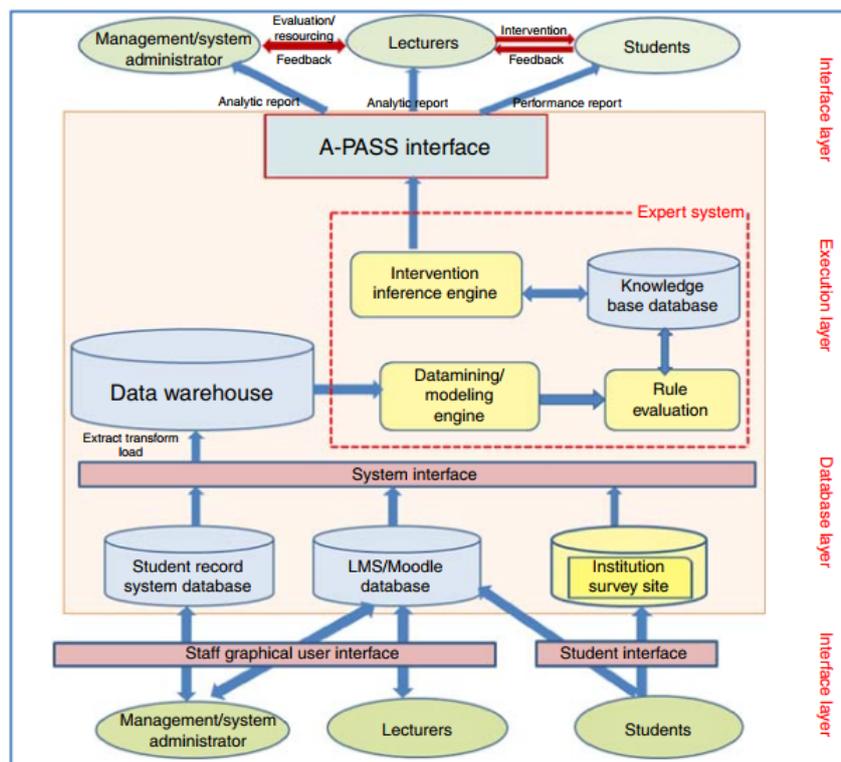


Figura 2.11: Framework para prever o desempenho dos estudantes — Fonte: Adejo e Conolly [6]

Adejo e Conolly[56], entendem que o baixo desempenho nas pesquisas de predição é gerado também pelo uso singular de classificadores e sugere o uso de um conjunto de algoritmos de previsão de desempenho. Para servir de base de comparação para o estudo, os autores utilizam três classificadores:

1. ANN — algoritmo de rede neural — tem boa e rápida descoberta de dados, especialmente em um grande conjunto de dados.

2. SVN — Desenvolvido para problemas de previsão e reconhecimento de padrões, assim como para análise de mapeamento de funções lineares. Ele constrói um hiperplano (classes) em um espaço de dimensões que pode ser usado para classificação.
3. DT — Estrutura de árvore de decisão, assume que existe apenas uma variável independente, chamada também variável alvo, e múltiplas variáveis dependentes. Por se intuitiva é uma das abordagens mais utilizadas na análise de dados.

A abordagem sugerida pelo autor é usar um conjunto de Classificadores (*Classifiers ensemble*). Esse algoritmo de aprendizado supervisionado combina um conjunto de classificadores em meta classificadores, realizando um processo de votação ou votação ponderada de sua previsão para a previsão final. Segundo o autor existem três abordagens, normalmente usadas na construção de conjuntos de classificadores: ensacar (*bagging*), impulsionar (*boosting*) e empilhar (*stacking*).

Todas as abordagens têm suas vantagens, mas para o autor o algoritmo de empilhamento (*stacking*) é um conjunto de algoritmos de aprendizagem que combinam vários modelos de diferentes formas. Funciona através da construção de múltiplos classificadores com previsão intermediária para um meta-classificador que define a previsão final. Por trabalhar com diferentes classes de algoritmos e apresentar uma capacidade de combinar os pontos fortes de diferentes classificadores heterogêneos sobre seus pontos fracos, foi utilizado no estudo. A Figura 2.12 apresenta o diagrama da estrutura do algoritmo de empilhamento.

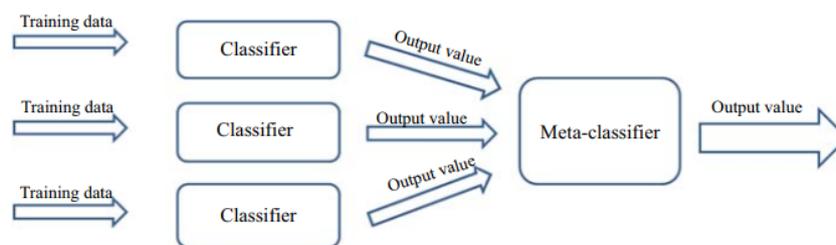


Figura 2.12: Diagrama da estrutura do algoritmo de empilhamento — Fonte: Adejo e Conolly [6]

No estudo os autores utilizaram três base de dados:

- A primeira reunia informações que não constavam nas bases da universidade, sendo coletadas a partir de um questionário. Após o pré-processamento das informações as variáveis ficaram definidas desta forma: situação econômica dos pais, horas de trabalho e pressão, qualificação de entrada, hora média de estudo, apoio familiar, satisfação do curso, impacto tecnológico, estilo de aprendizagem, situação de saúde,

primeiro ano na universidade, adaptação, apoio universitário e conhecimento prévio do curso;

- A segunda com dados do Sistema de Registro dos Estudantes onde foram selecionadas 8 variáveis: idade, sexo, etnia, localização do CEP, campus de estudo, qualificação de entrada no modo de estudo e deficiência;
- A terceira do Sistema de Gerenciamento de Aprendizagem (Moodle), onde foram selecionadas 5 variáveis: tempo total de *logon* do aluno, o número de recursos visualizados, o número de tentativas de *quiz* do módulo submetido, o número de módulos do curso do fórum visualizado e o número de módulos de discussão lidos ou visualizados.

Os resultados apresentados no estudo, mostraram que o modelo usando um conjunto híbrido de algoritmos e com as três bases de dados apresentou os melhores resultados. Na verificação de desempenho os resultados apresentados para erro médio quadrático (RMSE) foi de 0.396, a precisão de 79.62%, *recall* de 75.86%, F-score de 77.69% e o erro de classificação de 18.33%. Outro resultado relevante apresentado no estudo identifica que as atividades no Moodle fazem uma grande diferença no desempenho dos modelos, especialmente quando combinadas com outras fontes de dados.

# Capítulo 3

## Contexto Organizacional

Este capítulo apresenta um panorama da Universidade de Brasília, descreve a atuação da Secretaria de Tecnologia da Informação enquanto órgão complementar, no apoio as unidades da instituição. Apresenta o Decanto de Ensino de Graduação, o seu papel na criação e desenvolvimento de políticas educacionais. Descreve o trabalho do Centro de Educação a Distância, levando a educação de forma remota ao aluno. Finalizando, apresenta algumas leis criadas para combater direta ou indiretamente a evasão e retenção no ensino superior.

### 3.1 Universidade de Brasília

A Universidade de Brasília (UnB), sediada em Brasília, capital do Distrito Federal, é uma instituição pública federal de ensino superior que completa, em 2022, 60 anos de existência. Faz parte da sua missão e ser inovadora e inclusiva, comprometida na sua essência com o ensino, pesquisa e extensão. Visa qualificar pessoas para conseguirem transformar a sociedade via uma gestão eficaz e uma melhor qualidade de vida [57].

Avaliando os dados disponibilizados no Anuário Estatístico 2020 [58], a UnB consta com 134 cursos de graduação ativos (curso com ingressantes no ano base). No ano de 2019 ingressaram mais de 9 mil alunos pelo processo de seleção primário Enem, PAS, Vestibular, Transferência obrigatória, Transferência Facultativa, Programas Especiais e Convênio (PEC).

Desse modo, é possível perceber que há um número grande de estudantes que se matriculam na universidade. O papel da UnB passa pelo desafio de manter o aluno engajado no curso até a sua graduação.

Dados do Plano de Desenvolvimento Institucional (PDI) [59], mostram que o percentual de discentes que permaneceram na UnB após o tempo padrão de seu curso — Taxa de Retenção — foi de 50,16%. Para os percentuais de alunos matriculados com classificação

de evadido, desligado, ou transferido externo — Taxa de Evasão — o valor foi de 24,54% e no caso do aluno cotista a taxa ficou em 16,84% no ano de 2016.

A universidade conta com as ações do Decanato de Ensino de Graduação (DEG) que visam melhorar a qualidade o ensino. Outras ações são feitas a partir do registro dos anseios e desafios da Universidade no PDI e no Projeto Político Pedagógico Institucional (PPPI) [60]. Isso faz parte de um conjunto de estratégias que visam reduzir as taxas de e evasão em cursos de graduação.

### **3.1.1 Secretaria de Tecnologia da Informação (STI)**

A Secretaria de Tecnologia da Informação (STI) (antigo Centro de Processamento de Dados CPD) foi criado em 1991, e hoje sua missão é viabilizar soluções de tecnologia da informação que promovam a disponibilidade, integridade, confiabilidade e autenticidade das informações dos ativos relacionados aos sistemas informatizados da Universidade de Brasília (UnB). Seus principais objetivos são: de realizar o apoio ao desenvolvimento do ensino, da pesquisa e da extensão nas atividades de processamento de dados; promover o compartilhamento de recursos computacionais entre a comunidade acadêmica da UnB e as redes de pesquisa (inter)nacionais; desenvolver, implantar e manter sistemas em mainframe e em microcomputadores; e planejar e coordenar a implementação de serviços no âmbito do tratamento de eletrônico de informações [61].

A Estrutura Organizacional da STI é formada pelo Direção, Vice direção e o Conselho Consultivo: a Direção planeja e coordena a supervisão da atividade de tecnologia da informação e comunicação e segurança; a Vice Direção tem como principal competência substituir o Diretor nas falta e impedimentos, exercendo atribuições delegadas pelo Diretor e assessorando os demais setores interno da STI; O Conselho Consultivo assessora a STI na determinação de políticas, diretrizes e metas e também na tomada de decisões internas.

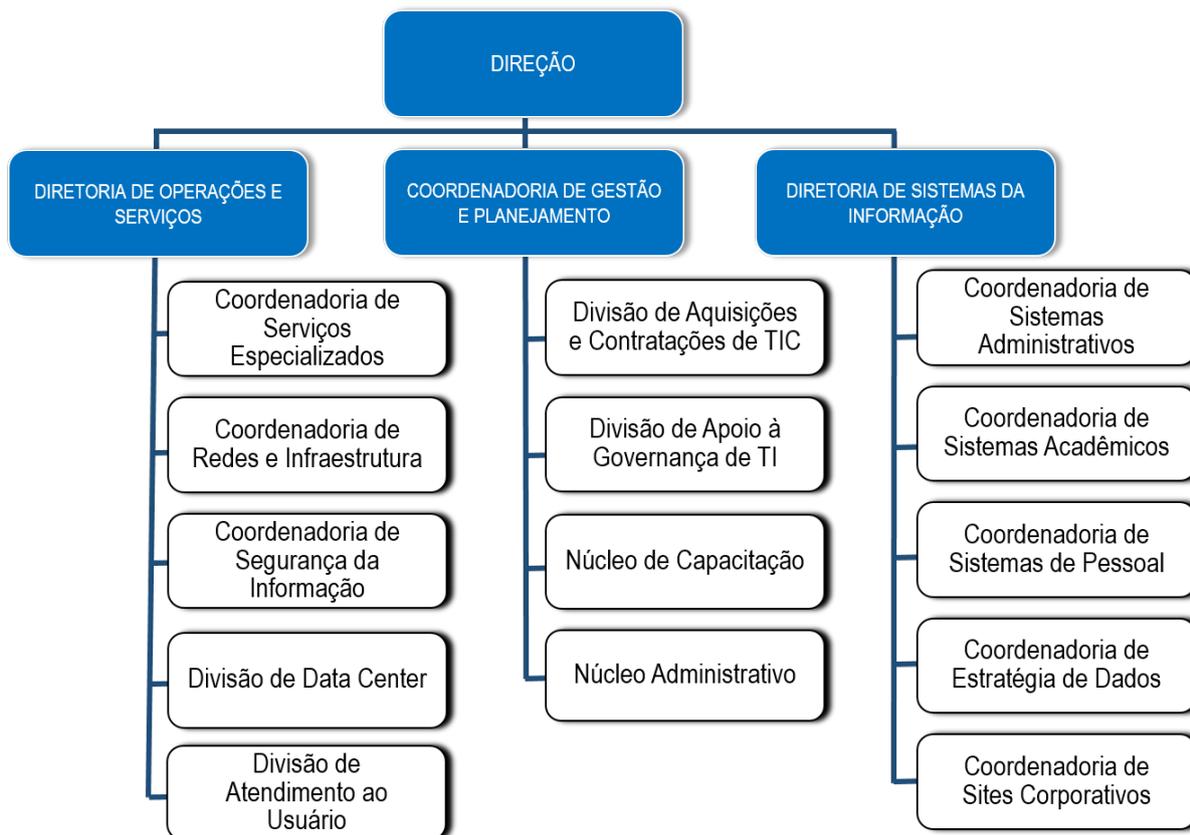
Outras três unidades fazem parte do organograma, A Diretoria de Operações e Serviços (DOS), a Coordenação de Gestão e Planejamento (CGESP) e a Diretoria de Sistemas da Informação (DSI) conforme pode ser visto na Figura 3.1. A Diretoria de Sistema de Informação (DSI) tem sob a sua gestão as Coordenações de Sistemas Acadêmicos, Administrativo, Pessoal, Sites e a Estratégia de Dados.

A Coordenações de Sistemas Acadêmicos é a unidade responsável pela sustentação dos sistemas acadêmicos. As atividades executadas estão ligadas ao desenvolvimento de sistemas, prestação de suporte ao usuário, realização de manutenções corretivas e evolutivas nos sistemas de informação acadêmicos. A Tabela 3.1 apresenta uma lista dos sistemas mantidos por essa coordenação.

Dentre os muitos sistemas mantidos por essa coordenação está o Sistema Integrado de Gestão das Atividades Acadêmicas (SIGAA). O SIGAA é um sistema que faz parte

Tabela 3.1: Sistemas acadêmicos mantidos pela STI.

| <b>Sigla - Denominação</b>  | <b>Órgão Responsável</b> |
|---|--------------------------|
| Acadêmico - Portal Acadêmico  | SAA / DEG                |
| Aluno - Portal do Aluno   | SAA / DEG / DPG          |
| CONDOC - Consulta de Docentes   | DEG / DPG / DGP          |
| Diploma - Sistema de Emissão de Diploma Digital da UnB                              | SAA / GRE                |
| DDUnB - Documento Digital   | SAA / VRT                |
| EditalDPG - Sistema de Auxílio a Participação em Eventos                            | DPG                      |
| EEditalTese - Sistema de Registro de Teses e Dissertações para Premiação            | DPG                      |
| MatWeb - Matrícula Web  | SAA / DEG / DPG          |
| SADD - Sistema de Acompanhamento de Desempenho Docente                              | VRT / DGP                |
| SAE - Sistema de Assistência Estudantil   | DDS                      |
| SIDIP - Sistema de Diplomas   | SAA                      |
| SIED - Sistema de Ensino à Distância  | DEG / CEAD               |
| SIEHE - Sistema de Emissão de Histórico Escolar                                     | SAA / DEG / DPG          |
| SIEX - Sistema de Extensão  | SAA / DEX                |
| SIEX - Sistema de Extensão e Pós-graduação (Lato-sensu)                             | DPG                      |
| SIGAA - Sistema Integrado de Gestão das Atividades Acadêmicas                       | SAA / DEG                |
| SIGRA - Sistema de Graduação  | SAA / DEG                |
| SigraWeb - Sistema de Lançamento de Menções dos alunos de Graduação e Pós-Graduação | SAA / DEG / DPG          |
| SSIPIIC - Sistema de Inscrição no Programa de Iniciação Científica (ex-SIBOL)       | DPG                      |
| SIPPOS - Sistema de Pós-Graduação   | SAA / DPG                |
| Sistema de Avaliação de Disciplinas por Discente                                    | DEG                      |
| DocAluno - Sistema Gerenciador dos Documentos dos Aluno                             | SAA/DEG                  |
| Sistemas de Inscrição Pós-graduação   | DPG                      |
| Sistema de Inscrição para Cursos de Ensino à Distância                              | CEAD/DEG                 |



Aprovada pelo Ato da Reitoria 0522/2020  
Processo SEI: 23106.107441/2019-77

Figura 3.1: Organograma da Secretaria da Tecnologia da Informação.

de um pacote de ferramentas on-line (SIG-UnB), adotado para unificar e modernizar os vários sistemas internos de gestão da Universidade de Brasília [62]. O SIGAA na sua configuração completa foi implementado para informatizar procedimentos da área de graduação, pós-graduação, pesquisa, extensão e assuntos comunitários. Atualmente está disponível para a área graduação e extensão na UnB [63].

### 3.1.2 Decanato de Ensino de Graduação (DEG)

O Decanato de Ensino de Graduação (DEG), cuida das políticas e processos institucionais relacionadas aos cursos de graduação e tem como missão promover e desenvolver a educação superior, ao nível de graduação, democrática, inclusiva e inovadora, comprometida com a formação de profissionais qualificados, éticos e socialmente referenciados.

Para o DEG [64], o SIGAA participa como ferramenta de simplificação das rotinas acadêmicas, concentra em um único software os procedimentos e políticas institucionais realizados pelo DEG para os cursos de graduação. O principal objetivo da adoção do

SIGAA foi o de simplificar rotinas acadêmicas, várias atividades além do acesso a diversas informações passaram a ser efetivadas com um único login e senha, sem necessidade de acesso a múltiplas plataformas. Por meio do Sistema — que congrega temas de interesse de professores (as) e estudantes — é possível, por exemplo, a emissão de documentos (Atestado de Matrícula, Declaração de Vínculo e Histórico Escolar para discentes de graduação) e a edição do perfil público dos(as) docentes, além de emissão de declaração de disciplinas ministradas.

### 3.1.3 Centro de Educação a Distância (CEAD)

O Centro de Educação a Distância da Universidade de Brasília (CEAD) desde 1979 vem facilitando o acesso à educação e cultura. Sua principal tarefa é desenvolver e viabilizar ações educativas a distância em várias áreas do conhecimento. Dentre os programas em que atua, os que mais se destacam são a Universidade Aberta (UAB) e a Aprendizagem para o 3º Milênio (A3M).

O Centro de Educação a Distância (CEAD) tem como unidade principal no seu organograma a Direção, no suporte a Direção está a Assessoria da Direção e a Coordenação Universidade Aberta do Brasil (UAB). A Coordenação de Tecnologia da Informação e da Comunicação da CEAD (CTIC), atua em ações que visam do desenvolvimento de tecnologias e metodologias inovadoras e na operacionalização dos Ambientes Virtuais de Aprendizagem (AVA) da UnB [65].

A UnB oferece à sua comunidade acadêmica dois ambientes virtuais de aprendizagem Moodle e o Microsoft Teams [66]:

- Moodle um sistema de gerenciamento de aprendizagem de código aberto, denominado na universidade como Aprender. Suas funcionalidades flexíveis permitem aos educadores criarem espaços online para o desenvolvimento das atividades educacionais. Atualmente a infraestrutura do sistema Aprender é mantido pela GigaCandanga, instituição de Ciência e Tecnologia, sem fins lucrativos, constituída em 2018. Uma das soluções fornecidas pela GigaCandanga é a hospedagem e customização da plataforma do Moodle em um serviço em nuvem [67].
- Microsoft Teams é parte da suíte de aplicativos Microsoft Office 365 Educacional e encontra-se disponível para todos dos alunos regulares da UnB, bem como para o seu corpo técnico-administrativo e professores mediante acordo de cooperação entre a UnB e Microsoft. A unidade responsável pela administração e configuração de perfis e a Coordenadoria de Serviços Especializados da STI.

## 3.2 Legislação

Com o crescimento do ensino superior, mecanismos foram criados para manter nas instituições alunos de família menos privilegiadas.

O Decreto N 7234, de 19 de julho de 2010, dispõe sobre o Programa Nacional de Assistência Estudantil (PNES). O artigo 4 parágrafo único, descreve que as ações de assistência estudantil devem considerar a necessidade de viabilizar a igualdade de oportunidades, contribuir para a melhoria do desempenho acadêmico e agir, preventivamente, nas situações de retenção e evasão decorrentes da insuficiência de condições financeiras.

Decreto n.º 6.096, DE 24 de abril de 2007, institui o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais — REUNI.

Art. 1.º Fica instituído o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais — REUNI, visando criar condições para a ampliação do acesso e permanência na educação superior, no nível de graduação, pelo melhor aproveitamento da estrutura física e de recursos humanos existentes nas universidades federais.

Art. 2.º O Programa terá as seguintes diretrizes: I — redução das taxas de evasão, ocupação de vagas ociosas e aumento de vagas de ingresso, especialmente no período noturno.

Segundo os dados do Relatório do primeiro ano do REUNI [68], o aumento de novas vagas trouxe um número maior de alunos, muitos deles de escolas públicas e com condições socioeconômicas não favorável a permanência na universidade. A partir da instituição da Portaria Normativa n.º 39, de 12 de dezembro de 2007, o Programa Nacional de Assistência Estudantil (PNAES), esses estudantes passaram a contar com auxílio para alimentação, transporte, moradia, apoio pedagógico, inclusão digital, assistência à saúde, cultura, esportes e serviços de creche. Possibilitando assim, melhores condições para que esses alunos cursassem a universidade federal e aumentar as taxas de diplomação e diminuir a evasão e repetência.

O Plano de Desenvolvimento Institucional 2018 – 2022 (DPI) e o documento que vem orientando a Universidade entre os anos de 2018 e 2022, nos seus anseios e desafios a partir do contexto em que a Universidade está presente. Nele são descritos os objetivos, indicadores e metas das unidades que compõem os eixos temáticos. O Decanto de Ensino de Graduação (DEG), que pertence ao eixo das Áreas Institucionais, coordena e avalia o ensino de graduação da UnB e propõe modificações para aprimorá-lo. Dentre essas ações foram descritos os objetivos e metas para combater a evasão e altas taxas de retenção conforme descrito na Tabela 3.2:

A Lei n.º 9.394, de 20 de dezembro de 1996, estabelece as diretrizes e bases da educação nacional. O título II trata dos princípios e dos fins da educacional nacional:

Tabela 3.2: PDI - Melhorar os resultados da avaliação institucional.

| <b>Objetivo</b>                                   | <b>Indicador</b>   | <b>Meta</b>   |
|---|--|---|
| Melhorar os resultados da avaliação institucional | Percentual de matrículas com status de Evadido, Desligado ou Transferido Externo - Taxa de Evasão <b>Histórico do Indicador:</b> 2016: 24,54%                            | 2018: 24%<br>2019: 23%<br>2020: 22%<br>2021: 21%<br>2022: 20% |
|   | Percentual de matrículas de cotistas com status de Evadido, Desligado ou Transferido Externo - Taxa de Evasão <b>Histórico do Indicador:</b> 2016: 16,84%                | 2018: 16%<br>2019: 16%<br>2020: 15%<br>2021: 15%<br>2022: 15% |
|   | Percentual de discente que permaneceram na UnB após o tempo padrão de seu curso - Taxa de Retenção <b>Histórico do Indicador:</b> 2016: 50,16% 2015: 56,10% 2014: 60,27% | 2018: 50%<br>2019: 50%<br>2020: 48%<br>2021: 48%<br>2022: 46% |

- Art. 2.º — A educação, dever da família e do Estado, inspirada nos princípios de liberdade e nos ideais de solidariedade humana, tem por finalidade o pleno desenvolvimento do educando, seu preparo para o exercício da cidadania e sua qualificação para o trabalho.
- Art. 3.º — O ensino será ministrado com base nos seguintes princípios:
  - I — Igualdade de condições para o acesso e permanência na escola;
- Art. 4.º — O dever do Estado com educação escolar pública será efetivado mediante a garantia de:
  - I — Igualdade de condições para o acesso e permanência na escola;

Instrução Normativa da Câmara de Ensino de Graduação N.º 001/2012, estabelece procedimentos para tratar de processos de discente em risco de desligamento, reintegração e orientação acadêmica.

- Art. 3º — Cabe à SAA e ao DEG a capacitação de servidores e docentes sobre as normas e procedimentos referentes às questões acadêmicas dos Cursos de Graduação, particularmente, IRA (Índice de Rendimento Acadêmico), TGM (trancamento geral de matrícula), mudança de curso, duplo curso, dupla habilitação, fase probatória (risco de desligamento), desligamento e reintegração.

- Art. 4º — Cabe à DAIA, com apoio da Comissão de Acompanhamento e Orientação (CAO): I — propor ações para orientação de docentes, coordenadores e discentes que promovam a permanência de discentes nos Cursos de Graduação; II — avaliar estatísticas de evasão e rendimento de discentes de graduação; III — indicar ações de acompanhamento das diversas situações que levem ao afastamento de discentes; IV — estabelecer ações de integração e recuperação acadêmicas de discentes;

A Resolução do Conselho de Ensino, Pesquisa e Extensão N. 10/2007, cria o Programa de Tutoria Especial (PTE), normaliza o apoio acadêmico a estudantes com necessidades especiais e dá outras providências. Este programa consiste no apoio acadêmico a estudantes da UnB com necessidade educacional especial, realizado por estudante da Universidade, sob a supervisão do professor da respectiva disciplina e da sua Unidade Acadêmica.

Principais objetivos:

- Propiciar uma formação acadêmica mais ampla e aprofundada ao aluno universitário;
- Ampliar a participação do aluno nas atividades da Universidade;
- Incentivar no aluno universitário, o interesse pela dedicação à docência e à pesquisa;
- Despertar vocações acadêmicas;
- Possibilitar maior integração dos segmentos na universidade.

A Resolução do Conselho de Ensino, Pesquisa e Extensão N. 10/2007, cria o Programa de Tutoria Especial (PTE), normaliza o apoio acadêmico a estudantes com necessidades especiais e dá outras providências. Este programa consiste no apoio acadêmico a estudantes da UnB com necessidade educacional especial, realizado por estudante da Universidade, sob a supervisão do professor da respectiva disciplina e da sua Unidade Acadêmica.

O Projeto Político Pedagógico institucional (PPPI) da UnB, revisado em 2018 para adequação as leis vigentes, traz diversas diretrizes que orientam o ensino, pesquisa e extensão. Em relação aos aspectos didáticos-pedagógicos e as diferentes formas de organizações, tem como parte de um grupo de orientações fundamentais a implantação de mecanismos que reduzam a evasão em cursos de graduação. O PPPI também aponta a evasão como um dos indicadores relevantes para a avaliação do ensino e pesquisa.

### 3.3 Contexto da Pesquisa

A Universidade tem um papel importante no direcionamento da vida de um estudante, muitas vezes, nesta etapa o universitário identifica se está no rumo certo com relação à profissão desejada e até mesmo interesses pessoais.

Neste processo, um grande volume de dados está sendo coletados pelas instituições de ensino, dados esses que se analisados podem ajudar a identificar alunos com problemas na condução do curso.

Desta forma, proceder de forma científica usando técnicas de Mineração de Dados Educacionais (EDM) com o objetivo de encontrar e entregar padrões uteis nestas informações pode ajudar na diminuição das taxas de evasão da Universidade de Brasília.

# Capítulo 4

## Mapeamento das Bases de Dados Educacionais da UnB

O propósito deste capítulo é descrever o procedimento realizado durante a análise nas bases de dados dos sistemas acadêmicos da UnB envolvidos neste estudo e buscar uma correlação entre os fatores de evasão identificados durante a revisão sistemática da literatura.

### 4.1 Sistemas Mapeados

A UnB disponibiliza aos usuários três ambientes de ensino para o apoio ao desenvolvimento das atividades acadêmicas para os cursos da graduação.

O Sistema Integrado de Gestão das Atividades Acadêmicas (SIGAA), que oferece suporte aos procedimentos acadêmicos e que veio em substituição ao Sistema de Graduação (SIGRA). Durante o processo de substituição, todos dos dados dos alunos ativos foram migrados para o SIGAA. Hoje o SIGRA é usado para obtenção de dados históricos.

Para dar suporte ao aprendizado remoto, a UnB disponibiliza dois ambientes: o sistema Aprender 3 (baseado a plataforma do Moodle) e o Microsoft Teams (pertencente ao conjunto de aplicativos Office 365).

Com base em uma análise visual das funcionalidades dos sistemas envolvidos foi possível gerar a Tabela 4.1 que mostra o sistema fonte dos fatores relacionados na revisão.

As seções seguintes detalham os procedimentos realizados que definem a inclusão ou exclusão dos sistemas avaliados neste trabalho.

Tabela 4.1: Indicação do sistema fonte dos fatores mapeados na revisão da literatura.

| <b>Fator</b>               | <b>SIGAA/SIGRA</b> | <b>Aprender</b> | <b>Teams</b> |
|----------------------------|--------------------|-----------------|--------------|
| pontuação média            | x                  | x               | x            |
| gênero                     | x                  |                 |              |
| notas do curso             | x                  | x               | x            |
| curso                      | x                  |                 |              |
| idade                      | x                  |                 |              |
| etnia                      |                    |                 |              |
| bolsa de estudos           | x                  |                 |              |
| código postal              | x                  |                 |              |
| estudo regular             |                    |                 |              |
| horário do curso           | x                  | x               | x            |
| tipo de entrada            | x                  |                 |              |
| emprego                    |                    |                 |              |
| matérias aprovadas         | x                  | x               | x            |
| nível educacional dos pais |                    |                 |              |
| acesso a fórum             |                    | x               | x            |
| acesso ao sistema          |                    | x               | x            |
| resolução de exercícios    |                    | x               | x            |
| acesso a vídeo             |                    | x               | x            |
| conhecimento prévio        | x                  |                 |              |
| estado civil               | x                  |                 |              |
| estrangeiro                | x                  |                 |              |
| matérias reprovadas        | x                  |                 |              |
| necessidades especiais     | x                  |                 |              |
| trancamento                | x                  |                 |              |
| curso matriculados         | x                  |                 |              |
| dados de pesquisa          |                    |                 |              |
| filhos                     |                    |                 |              |
| apresentação tardia        |                    | x               | x            |
| estágio                    | x                  |                 |              |

### 4.1.1 Sistema de Integrado de Gestão de Atividades Acadêmicas

O Sistema Integrado de Gestão das Atividades Acadêmicas (SIGAA) faz parte de um conjunto de sistemas desenvolvidos no âmbito da UFRN que, direta ou indiretamente, viabilizam a operacionalização e a otimização dos processos de gestão acadêmica, administrativa e de pessoal da Universidade de Brasília [69]. O com a implantação do SIGAA, a gestão dos cursos de graduação passou a ser executada nele em substituição ao Sistema de Graduação (SIGRA).

As informações acadêmicas dos alunos de graduação estão centralizadas no SIGAA. Nele, estão registrados dados que vão desde o momento da matrícula do aluno até a sua saída da instituição. A Figura 4.1 mostra algumas das funcionalidades disponíveis no SIGAA. Essas funcionalidades fizeram parte da análise visual, citada no início desse capítulo, realizada no sistema. Esse procedimento serviu para a identificação dos fatores e/ou indicadores selecionados para este trabalho.



Figura 4.1: Interface do SIGAA - Menu de graduação.

Para uma melhor identificação dos dados registrados no sistema, a Figura 4.2 mostra um detalhamento dos dados da tela de consultar dados do aluno.

Os fatores etnia, emprego, nível de escolaridade dos pais, dados de pesquisa, e filhos não foram encontrados neste sistema, conforme o descrito na Tabela 4.1.

No procedimento de análise da estrutura do modelo de dados do SIGAA foram selecionadas 26 tabelas e delas foram mapeados 36 atributos que trabalhados de forma individual ou combinados atendem aos fatores identificados na revisão sistemática, sendo eles: pontuação média, gênero, notas do curso, curso, idade, bolsa de estudos, código postal (CEP), horário do curso, tipo de entrada na universidade, matérias aprovadas, estado

| DADOS DO ALUNO  |  |
|---|--|
| <b>Matrícula:</b>   | [REDACTED]                                       |
| <b>Nome Social:</b>   | [REDACTED]                                       |
| <b>Status:</b>  | ATIVO  |
| <b>Forma de Ingresso:</b>                                     | VESTIBULAR                                       |
| <b>Cota de Inscrição:</b>                                     |  |
| <b>Cota de Ingresso:</b>                                      | Universal  |
| <b>Nota Final:</b>  | 0.0  |
| <b>Ano/Período de Ingresso:</b>                               | 2014.2   |
| <b>Curso:</b>   | FARMÁCIA/FCS - Bacharel - Presencial - D - 20132 |
| <b>Estrutura Curricular:</b>                                  | 7692/-2 - 2013.2                                 |
| <b>Data de Cadastro:</b>                                      | 24/04/2020                                       |
| <b>Cadastrado Por:</b>  |  |
| <b>DADOS PESSOAIS</b>   |  |
| <b>Sexo:</b>  | M  |
| <b>Data de Nascimento:</b>                                    | 12/09/1985                                       |
| <b>Raça:</b>  | Testes   |
| <b>Nacionalidade:</b>   | Brasil   |
| <b>Nome do Pai:</b>   | Informação Anonimizada                           |
| <b>Nome da Mãe:</b>   | Informação Anonimizada                           |
| <b>Possui tipos de Necessidades Educacionais Específicas:</b> | Não  |
| <b>Escola de conclusão do Ensino Médio:</b>                   | Testes   |
| <b>Ano de Conclusão:</b>                                      | 1999   |
| <b>Estado Civil:</b>  | SOLTEIRO   |
| <b>Naturalidade:</b>  | NÃO INFORMADO / Pará                             |
| <b>Tipo Sanguíneo:</b>  | XX+  |
| <b>País:</b>  | BRASIL   |
| <b>Tipo da Escola de Ensino Médio:</b>                        | Privada  |
| <b>DOCUMENTOS</b>   |  |
| <b>CPF:</b>   | 977.422.100-10                                   |
| <b>RG:</b>  | 99999  |
| <b>Título de Eleitor:</b>                                     | 999999999999                                     |
| <b>Certificado Militar:</b>                                   | 999999   |
| <b>Série:</b>   | X  |
| <b>Passaporte:</b>  |  |
| <b>Órgão de Expedição:</b>                                    | XXXX   |
| <b>Data de Expedição:</b>                                     |  |
| <b>UF:</b>  | PA   |
| <b>Zona:</b>  | 999  |
| <b>Seção:</b>   | 9999   |
| <b>UF:</b>  | Não Informado                                    |
| <b>Data de Expedição:</b>                                     |  |
| <b>Categoria:</b>   | 99   |
| <b>Órgão:</b>   |  |
| <b>ENDEREÇO</b>   |  |
| <b>Endereço:</b>  | OUTROS Conteúdo protegido Conteúdo protegido     |
| <b>Número:</b>  | 999  |
| <b>Bairro:</b>  | Bairro   |
| <b>Unidade Federativa:</b>                                    | Distrito Federal                                 |
| <b>CEP:</b>   | 59000-000  |
| <b>Complemento:</b>   |  |
| <b>Município:</b>   | BRASILIA   |
| <b>CONTATOS</b>   |  |
| <b>Telefone:</b>  | (61) 9999-9999                                   |
| <b>Celular:</b>   | (61) 9999-9999                                   |
| <b>E-Mail:</b>  | desenv@si.ufrn.br                                |
| <b>DADOS BANCÁRIOS</b>  |  |
| <b>Banco:</b>   |  |
| <b>Conta:</b>   |  |

Figura 4.2: Interface do SIGAA - Dados do Aluno.

civil, estrangeiro, matérias reprovadas, necessidades especiais e trancamento disciplina. A Figura 4.3 mostra as tabelas usadas para realizar o mapeamento dos atributos para atender aos fatores selecionados na revisão sistemática.

O SIGAA entrou em produção no primeiro semestre de 2020, e em maio de 2022 contabilizava mais de 69 mil matrículas de alunos, sendo boa parte desses registros migrados do SIGRA, que será descrito na Seção 4.1.2.

Dessas 69 mil matrículas registras, as mulheres representam 52%, conforme pode ser visto na Figura 4.4.

Tomando como base os alunos em situação de ativo, estes chegam a mais 64% do total geral de alunos. Para esses alunos, o ingresso na universidade pelo Programa de Avaliação Seriada compreende 36%, o Vestibular representa 24%, o Exame Nacional do Ensino Médio (ENEM) com 16% e o Sistema de Seleção Simplificada (SISU) com 8%.

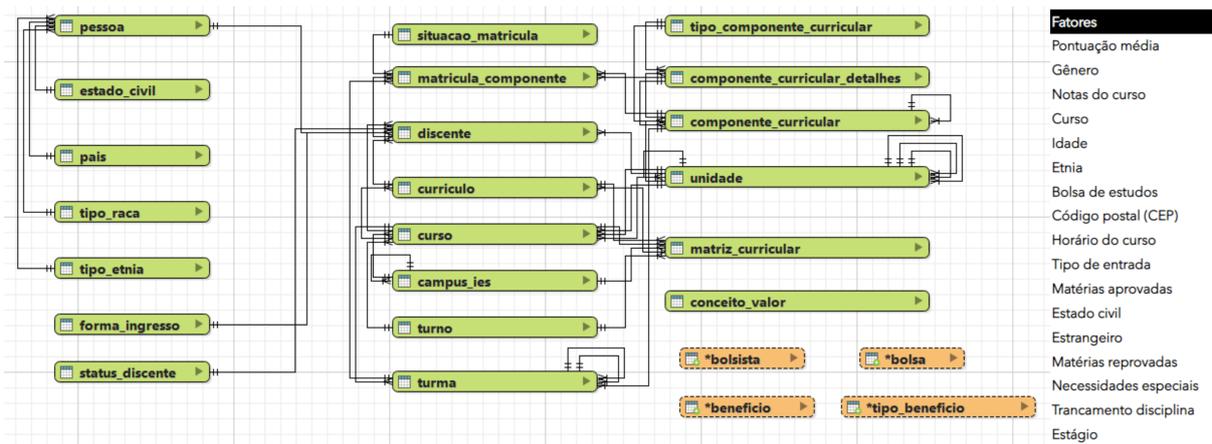


Figura 4.3: Mapeamento modelo de dados e fatores no SIGAA.

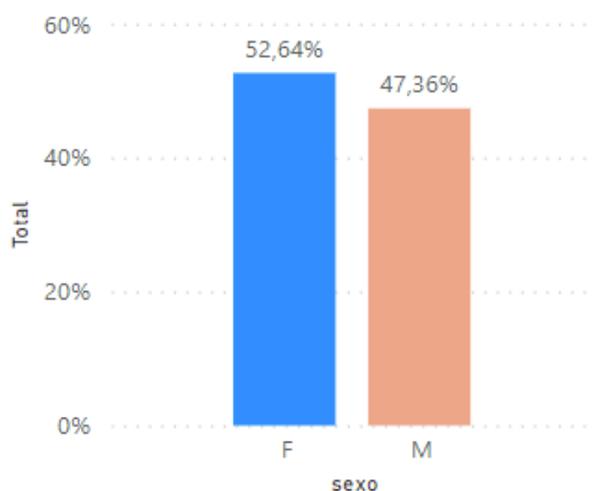


Figura 4.4: Distribuição do percentual de alunos por sexo no SIGAA.

Um panorama com os percentuais dos tipos de ingresso na universidade é apresentado na Figura 4.5.

Os cursos ativos no SIGAA somam 281, sendo 126 no período diurno, 31 no noturno e 124 não apresentam informação do turno conforme descrito na Figura 4.6.

#### 4.1.2 Sistema de Graduação

O Sistema de Graduação (SIGRA) foi usado na gestão acadêmica nos últimos 10 anos, sendo substituído pelo SIGAA. Os dados do SIGRA atendem às demandas das unidades acadêmicas que necessitam de informações de carácter histórico e fornecem também dados para pesquisas na universidade. A Figura 4.7 mostra as tabelas usadas neste estudo.

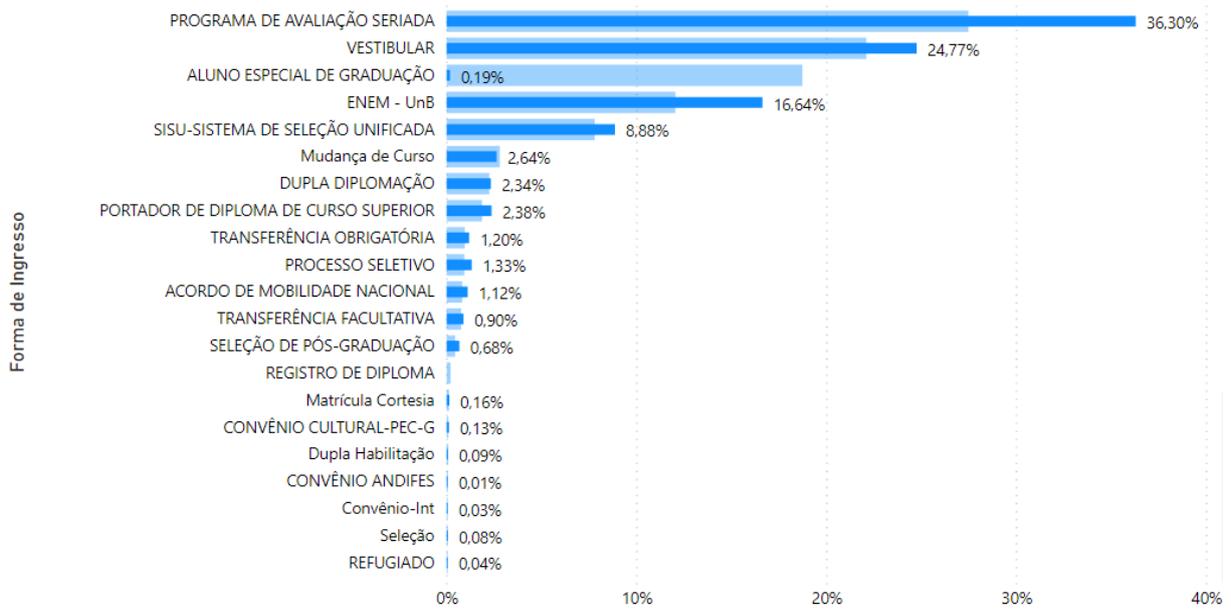


Figura 4.5: Distribuição do percentual da forma de ingresso de alunos ativos - SIGAA.

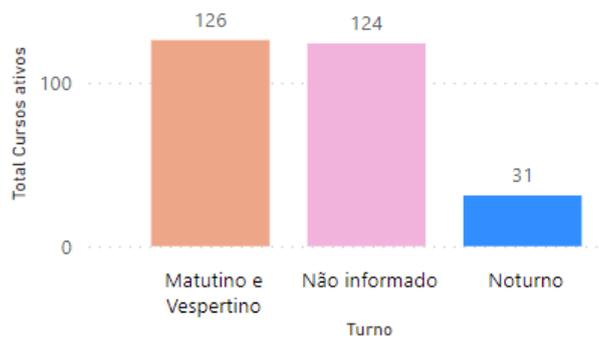


Figura 4.6: Distribuição do total de cursos ativos por período no SIGAA.



Figura 4.7: Modelo de dados SIGAA.

Na Figura 4.8 e apresentada a distribuição por sexo. A base de dados soma um total de 161 mil registros de matrículas, onde 62% são do sexo feminino, 32% do sexo masculino e quase 6% não apresentam valor neste campo.

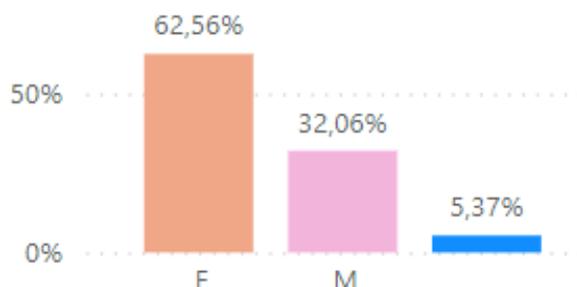


Figura 4.8: Distribuição de alunos por sexo no SIGRA.

A forma de ingresso pelo Vestibular chega a mais de 67%, o Programa de Avaliação Seriada compreende 13% e o Sistema de Seleção Unificada apresenta 4%. Para os dados referente a forma de saída da Universidade, a Formatura representa 49% e Desligamento por não Cumprir Condição e Abandono são 16,40% e 16,67% respectivamente. A Figura 4.9, mostra as nove primeiras formas de ingresso registrada no SIGRA.

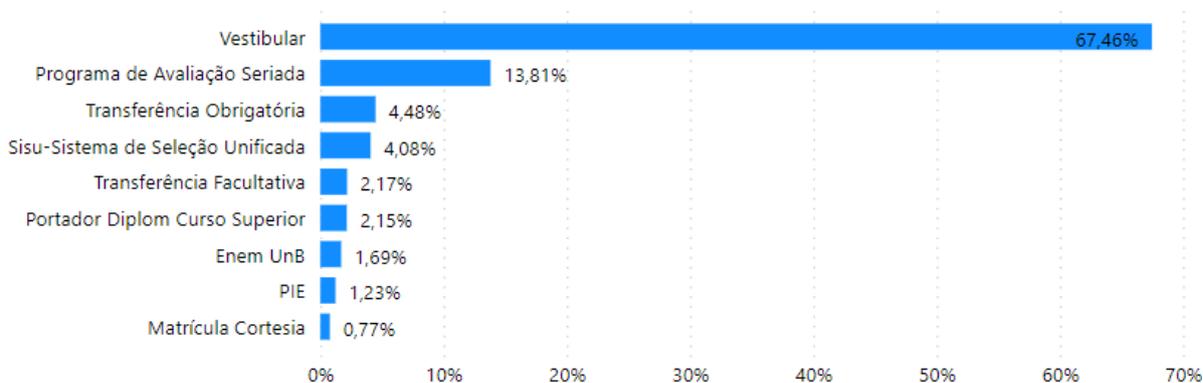


Figura 4.9: Distribuição da forma de ingresso no SIGRA.

Avaliando a forma de saída Universidade, é possível identificar dois itens que chamam a atenção. O Desligamento por Abandono e a Reprovação 3 vezes na mesma disciplina, representam 16,37% e 1,81% respectivamente, do total de alunos que deixam a instituição, conforme pode ser visto na Figura 4.10.

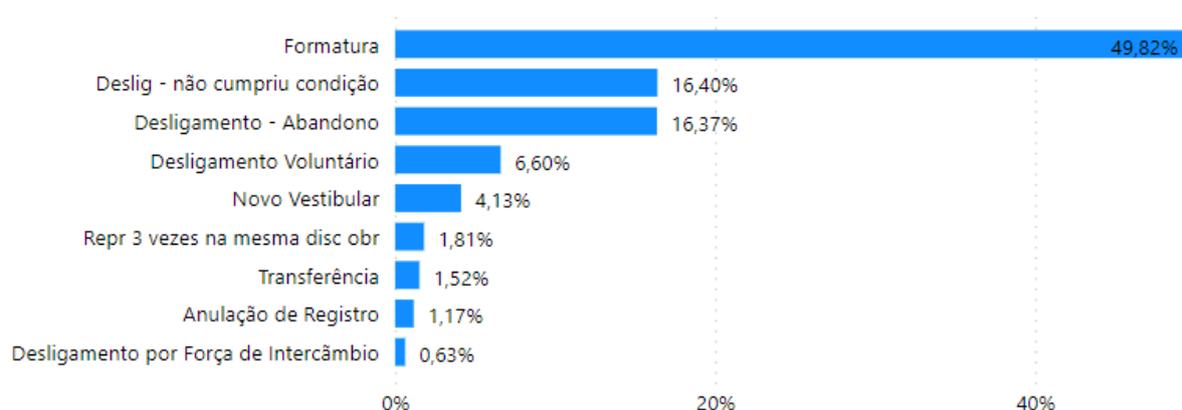


Figura 4.10: Distribuição da forma de saída registrada no SIGRA.

### 4.1.3 Ambientes de Ensino-aprendizagem On-line

A Universidade de Brasília dispõe de dois ensino-aprendizagem on-line. O Aprender 3, que utiliza a plataforma Moodle, e o Microsoft Teams, que faz parte da suíte de aplicativos Microsoft Office 365 Educacional.

O mapeamento desses sistemas está disponível no Apêndice B, mas, devido à indisponibilidade dos dados, esses sistemas não foram considerados no presente trabalho.

Após o mapeamento das bases de dados que podem ser usadas, iniciou-se a extração dos dados selecionados. Esses dados, após serem extraídos, transformados e carregados em um novo ambiente, chamado de Base Analítica, constituem um dos elementos do modelo de análise da evasão desenvolvido e que será detalhado no próximo capítulo.

# Capítulo 5

## Modelo de Análise na Graduação - MAGRA

Considerando a definição de evasão como fenômeno no qual um estudante ingressa em um determinado curso, não integraliza o currículo e, conseqüente, não é diplomado. Considerando o impacto negativo que esse fenômeno traz para o aluno e a instituição, este capítulo apresenta um modelo analítico para análises da evasão nos cursos de graduação presenciais da Universidade de Brasília (UnB).

### 5.1 Estrutura do MAGRA

O Modelo de Análise na Graduação (MAGRA) utiliza os indicadores existentes nos sistemas acadêmicos da UnB em conjunto com ferramentas de aprendizagem de máquina para prever alunos com potencial de risco de evasão.

Os indicadores, algoritmos e ferramentas identificados durante a revisão sistemática da literatura norteiam a estrutura desenvolvida para as análises de evasão. Os indicadores selecionados compõem a estrutura de variáveis submetidas aos modelos de predição.

O modelo está dividido em cinco partes conforme a Figura 5.1:

1. Seleção das bases de dados, tabelas e atributos
2. Construção do ETL para a base de dados analítica;
3. Algoritmos e modelos de aprendizagem;
4. Avaliar a melhor solução;
5. Produção de resultados e relatórios.

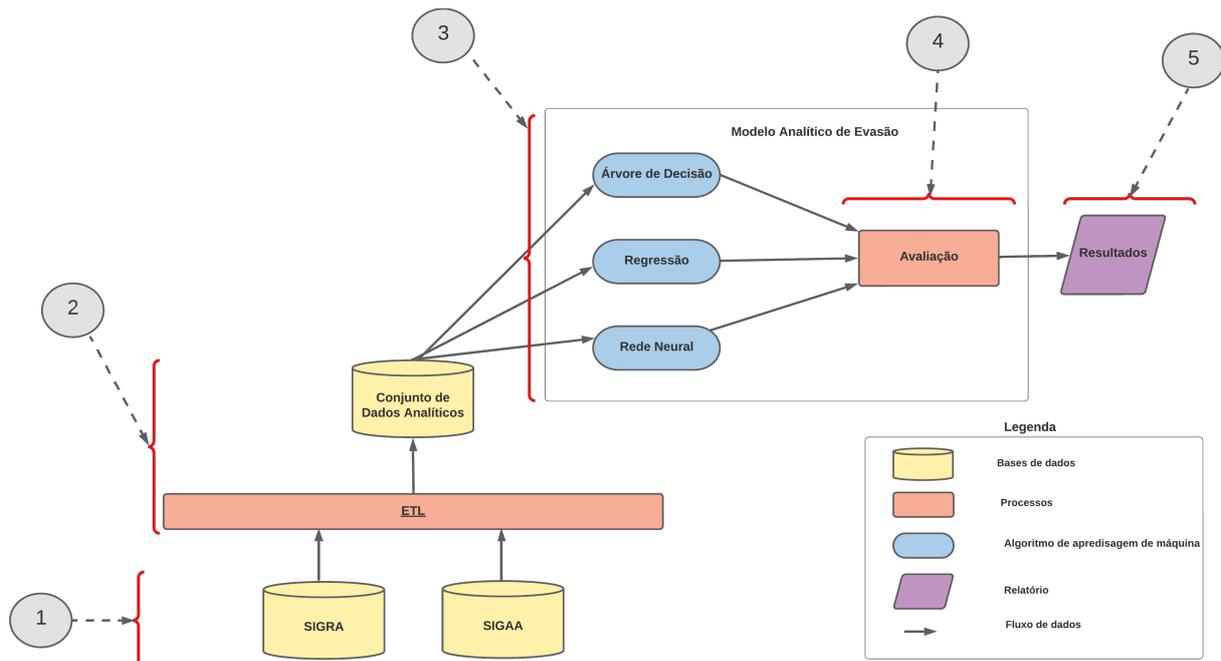


Figura 5.1: Modelo para predição da Evasão.

### 5.1.1 Seleção das bases de dados, tabelas e atributos

O escopo do MAGRA abrange os dados contemplados no SIGAA e no SIGRA, eles estão acessíveis em ambientes de desenvolvimento com informações pessoais dos alunos já anonimizadas pela STI. O acesso está disponível a partir de autenticação via credências de acesso disponibilizadas aos analistas da STI. O SIGAA (sistema atual em produção da UnB), e o SIGRA (sistema legado) usam banco de dados distintos. O primeiro usa PostgreSQL<sup>1</sup> e o segundo usa o SQLServer<sup>2</sup>. A gestão desses dois bancos de dados é feita pela Coordenadoria de Estratégia de Dados (CED) da Diretoria de Sistemas de Informação (DSI) na Secretaria de Tecnologia da Informação (STI).

A construção do *script* para a seleção dos dados do SIGAA envolveu 21 tabelas e 55 atributos. Os atributos selecionados nesta etapa tinham um foco principal, atender aos fatores selecionados na revisão sistemática, mas alguns deles também foram selecionados para validação de informações importantes na construção do MAGRA. O *script* em linguagem SQL executada no banco de dados do SIGAA e a tabela com o tipo de cada atributo foi fornecido no Apêndice D.

<sup>1</sup>PostgreSQL <https://www.postgresql.org/>

<sup>2</sup>SQLServer <https://www.microsoft.com/pt-br/sql-server/>

### 5.1.2 Construção do ETL para a base de dados analítica

Por serem sistemas diferentes e com modelagem e banco de dados distintos, a utilização conjunta dessas duas bases traz uma complexidade que deve ser considerada. Para Kimball e Caserta [70], um sistema de extração, transformação e carga, do inglês *Extract Transform-Load (ETL)*, extrai dados dos sistemas fontes, reforça padrões de qualidade, consistência e conforma os dados para que fontes separadas possam ser usadas juntas, entregando os dados em um formato pronto para desenvolvedores e usuários finais possam tomar suas decisões.

Conforme Kimball e Caserta [70], o ETL deve ser mais que um canal para obter dados, especialmente deve se preocupar em:

- Remove erros e corrige dados ausentes
- Fornece medidas documentadas de confiança nos dados
- Captura o fluxo de dados transacionais para a guarda segura
- Ajuste nos dados de múltiplas fontes para serem usados juntos
- Estruturar dados para serem utilizados por ferramentas do usuário final

Com objetivo de automatizar e facilitar o processo de ETL, o Pentaho Data Integration <sup>3</sup> foi escolhido como ferramenta de apoio para os procedimentos de extração, transformação e carga dos dados. Essa ferramenta foi escolhida por ser a ferramenta usada pelos analistas responsáveis pela gestão dos bancos de dados na STI.

O Pentaho é um software que permite a construção de um fluxo de extração, transformação e gravação dos dados usando componentes que podem ser arrastados e soltos na área de trabalho do Pentaho. Esses componentes permitem uma configuração e parametrização flexível ao realizar a conexão com diversos tipos de fontes de dados. Os componentes disponíveis na ferramenta também permitem realizar tarefas para tratamento dos dados que estão sendo manipulados.

As consultas extraídas dos bancos de dados acadêmicos foram gravadas em um banco de dados PostgreSQL. Esse banco foi chamado de Base Analítica, onde sua a função é servir de área de estágio para limpeza e tratamento dos dados para a etapa de construção dos modelos de predição do MAGRA. A Figura 5.2 mostra a estrutura do fluxo de extração, transformação e carga (ETL) realizado no Pentaho.

Ao finalizar o processo de ETL, foram geradas duas tabelas, uma contendo os dados do SIGAA e outra contendo os dados do SIGRA. Contudo, foi necessário unificar as

---

<sup>3</sup>HITACHI [https://help.hitachivantara.com/Documentation/Pentaho/7.1/OD0/Pentaho\\_Data\\_Integration](https://help.hitachivantara.com/Documentation/Pentaho/7.1/OD0/Pentaho_Data_Integration)

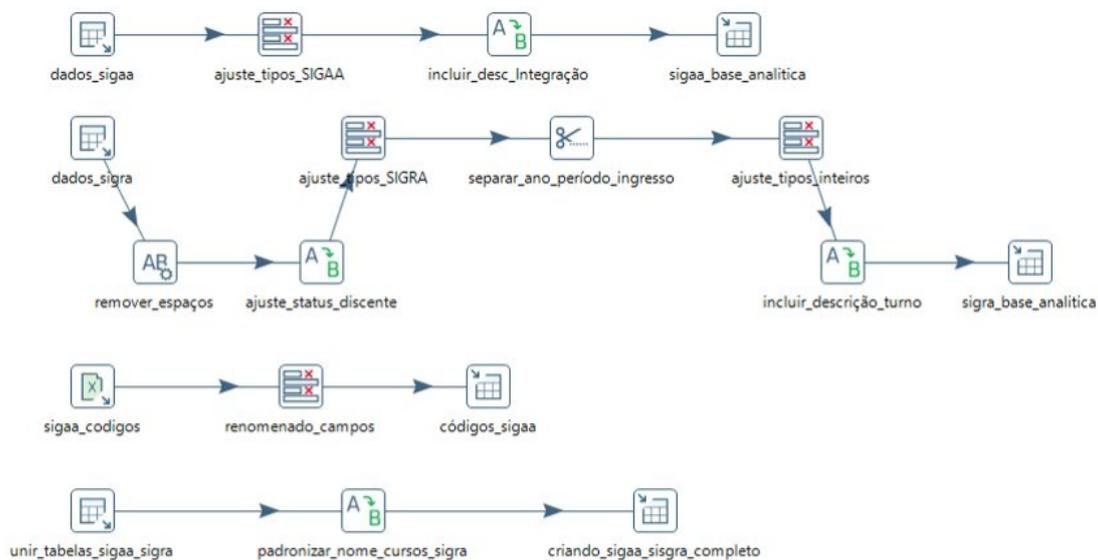


Figura 5.2: Área de construção do fluxo de dados no Pentaho.

informações dos dois bancos de dados. Para isso foi realizado um procedimento de junção das bases, onde o objetivo foi obter um único conjunto de dados onde fosse possível realizar resgate de uma coorte de alunos completa.

Nesse conjunto de dados é possível selecionar os alunos do SIGAA de uma determinada coorte, mais os alunos do SIGRA dessa mesma coorte que já concluíram, trancaram ou abandonaram a Universidade. Isso permitiu compor as turmas em um único conjunto de informações para a etapa de treinamento dos modelos.

No SIGAA apenas um atributo define a coorte por curso. No caso do SIGRA, três atributos tiveram que ser considerados nesse processo, sendo eles: um para definir o curso, outro para a data de início do curso e a data de início do currículo do curso em questão.

Uma vez definido esses atributos, um procedimento foi escrito no banco para executar essa operação. Ao final da execução da rotina foi possível unir as duas bases de dados realizando uma simples consulta de seleção e união das bases. A Figura 5.3 mostra o fluxo de dados onde o resultado é uma única base de dados composta por informações dos dois sistemas envolvidos.

Outra ferramenta utilizada nesta fase foi o Microsoft Power BI. Ela permite conectar em várias fontes de dados, tratar e gerar a visualização gráfica dos dados [71]. Os dados extraídos e carregados na Base de Analítica foram tratados para gerar as análises descritivas e geração dos relatórios.

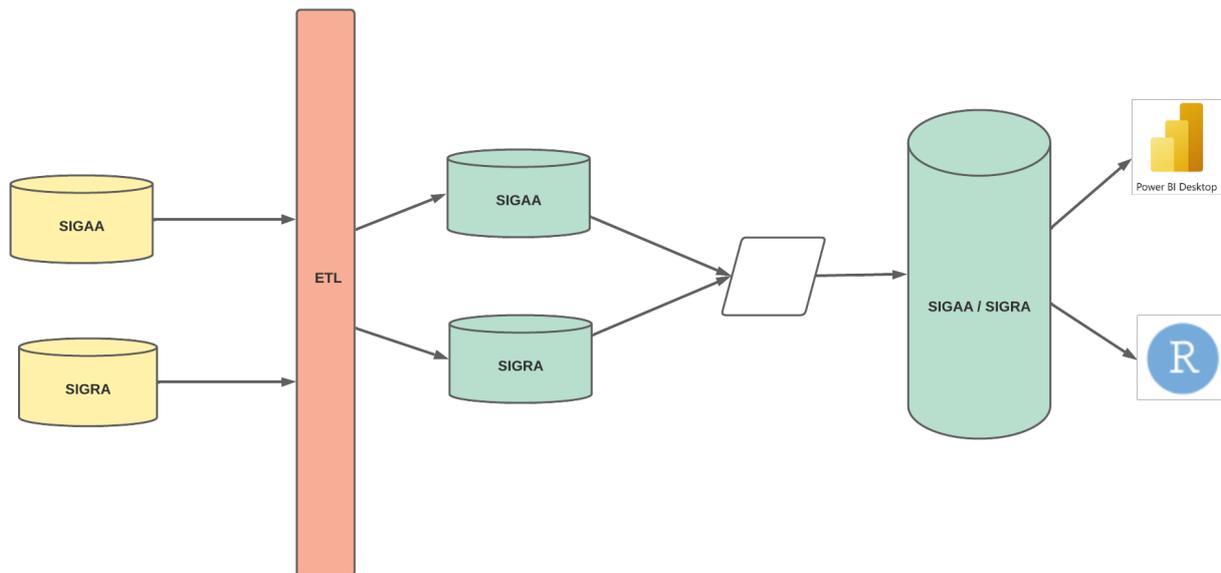


Figura 5.3: Diagrama do fluxo de dados do MAGRA.

### 5.1.3 Algoritmos e modelos de aprendizagem

Com a execução da revisão sistemática da literatura foi possível identificar os três algoritmos mais utilizados nos artigos que estudam as causas de retenção, evasão no ensino superior: a Regressão, Árvores de Decisão e Redes Neurais.

Outra informação importante nesta etapa, diz respeito às ferramentas usadas para dar suporte ao processo de preparação dos algoritmos e geração dos modelos. Duas ferramentas ficaram em destaque nos artigos pesquisados: O IBM SPSS e linguagem R. O IBM SPSS<sup>4</sup> e um software da IBM com um conjunto de ferramentas para validar suposições, analisar o desempenho passado e prever tendências. Por ser uma ferramenta com um custo de uso e não será adotada no trabalho.

A linguagem R<sup>5</sup> por sua vez é tem código aberto, podendo ser executada na linha de comando ou por meio do RStudio Desktop<sup>6</sup>, na versão com licença de código aberto. Para a execução dos algoritmos pretendidos neste trabalho é necessário executar ou instalar os pacotes específicos de cada algoritmo no RStudio.

O algoritmo de regressão (RegLog) foi implementado com o pacote GLM<sup>7</sup>, acrônimo para *Fitting Generalized Linear Models*. Esse pacote já é nativo do R e o modelo gerado por ele permite fazer análises de correlação de variáveis e construção de modelos de classificação.

<sup>4</sup>IBM <https://www.ibm.com/br-pt/analytics/spss-statistics-software>

<sup>5</sup>R <https://cran.rstudio.com/>

<sup>6</sup>RStudio <https://www.rstudio.com/products/rstudio/download/>

<sup>7</sup>R <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/glm>

Para o algoritmo que usa árvore de decisão foram usados três pacotes que implementam modelos com essas características:

- *Recursive Partitioning and Regression Trees* (RPART) permite trabalhar com a árvore e decisão em duas abordagens: uma para classificação e outra para problemas de regressão [55]. Neste trabalho foi usado na classificação.
- C5O <sup>8</sup>, são modelos de árvore de classificação de ajuste ou modelos baseados em regras usando o algoritmo C5.0 da Quinlan.
- Ranger (RF) <sup>9</sup>, O Ranger é uma implementação otimizada das florestas aleatórias (Breiman 2001) ou partição recursiva, particularmente adequada para classificação e regressão em dados de alta dimensão.

Por fim, o algoritmo de Rede Neural (RN) pertence ao pacote H2O <sup>10</sup>, plataforma escalável de aprendizagem de máquinas de código aberto que oferece implementações paralelas de muitos algoritmos de aprendizagem de máquinas supervisionados e não supervisionados.

Uma vez definidos os algoritmos, a próxima etapa é avaliar a qualidade dos modelos gerados com base nos algoritmos descritos. Essa etapa será descrita na próxima seção.

#### 5.1.4 Avaliar a melhor solução

Conforme Silva et al. [55], uma medida comumente usada na avaliação de classificadores e acurácia ou taxa de classificações corretas. Ainda segundo o autor, existe outra forma de avaliar a acurácia de um classificador. Avaliar em termos do erro de generalização e uma função de perda binária, sendo interpretada como a probabilidade de ocorrer uma classificação correta.

Outra métrica muito usada é o *F1 Score*, definida como a média harmônica entre a precisão do inglês *precision* e da revocação do inglês *recall* [72]. A função tem a seguinte fórmula:

$$F_1 \text{ score} = \frac{2.TP}{2.TP+FP+FN} = 2 \cdot \frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}}$$

O valor do *F1 Score* obtido, etapa de treinamento dos modelos, é avaliado para verificar se é igual ou maior que 70%. Atendendo a essa condição o modelo é selecionado para a etapa de teste no processo de previsão da evasão.

---

<sup>8</sup>C5O <http://127.0.0.1:12389/library/C50/html/C5.0.html>

<sup>9</sup>Ranger <http://127.0.0.1:22767/library/ranger/html/ranger.html>

<sup>10</sup>H2O <http://127.0.0.1:12389/library/h2o/html/h2o-package.html>

Os modelos selecionados na fase de treinamento serão testados com um conjunto de informações dos alunos ativos do SIGAA. O conjunto de previsões resultante dessa etapa contém os alunos com previsão de evasão. Pode acontecer do aluno ser previsto em situação de evasão em mais de um modelo, isso foi chamado de Força de Previsão do Modelo de Análise na Graduação (MAGRA)

Essa força de previsão serve como um sinalizador para os coordenadores de cursos. Previsões onde alunos aparecem em 5 modelos pode funcionar como um indicador de prioridade para os coordenadores. É importante realçar também que, isso não limita a análise nas outras previsões, funciona como um alerta. E mesmo aquelas previsões realizadas por apenas um algoritmo devem ser consideradas.

# Capítulo 6

## Testes e Análises de Dados

O propósito deste capítulo é descrever os procedimentos executados durante a etapa de testes e análises dos dados.

O primeiro cenário serviu para o desenvolvimento de um protótipo, usando os dados dos alunos do campus da Faculdade do Gama e depois foi ampliando em um segundo cenário com os alunos dos outros campus da Universidade de Brasília.

### 6.1 1º Cenário — Faculdade do Gama

A Faculdade do Gama oferece para os alunos da graduação 5 cursos, sendo eles Engenharia Aeroespacial, Engenharia Automotiva, Engenharia de Energia, Engenharia de Software e Engenharia Eletrônica. Quando o aluno ingressa na Faculdade ele não optará imediatamente por um dos cursos citados anteriormente, o ingresso ocorrerá no curso Engenharia, onde o aluno permanece por três períodos letivos recebendo o conteúdo de formação básica. Passado esse período o aluno fará a opção pelo curso específico.

As análises nesta etapa do trabalho não incluíram o curso de Engenharia que é da Área Básica de Ingresso (ABI), onde as disciplinas são comuns a todos os alunos.

O processo de análise consistiu na seleção das disciplinas obrigatórias cursadas pelos alunos de cada curso da Faculdade do Gama, elevando em consideração a quantidade de vezes que o aluno cursa com sucesso ou não uma determinada disciplina. Neste processo, o objetivo é encontrar um padrão que identifique os alunos com características de evasão. Nesta primeira análise atendemos a três dos fatores identificados na revisão sistemática: curso; disciplinas cursadas com aprovação e disciplinas que foram cursadas mais de uma vez.

### 6.1.1 Construção da Matriz de Fatores

A preparação do conjunto de dados para os modelos é uma etapa importante no processo de predição, pois ela é a chave mestra para a condução do processo. Para o trabalho em questão, foi necessário construir a Matriz de Fatores, sendo composta nesta etapa do trabalho somente por disciplinas cursadas pelos alunos.

Essa matriz contém linhas indexadas com o identificador dos alunos de um determinado curso e colunas com identificadores das disciplinas curadas. Essas disciplinas são chamadas variáveis independentes [55] e estão baseadas no currículo do curso que está sendo analisado.

Ao fazer o cruzamento entre coluna e linha, será encontrado um dígito que indica se a disciplina foi cursada e o número de vezes. A última coluna da matriz é reservada para receber a situação do aluno, que assume dois estados, “FORMADO” ou “EVADIDO”. A variável “SITUACAO” é denominada variável dependente ou de classe a ser prevista no processo de predição [55]. A Figura 6.1 mostra um exemplo da Matriz de Fatores com 20 alunos, com o número de vezes que a disciplina foi cursada e a situação de cada um. Nesse grupo, temos 4 estudantes em situação de evasão.

|        | FGA0086 | FGA0069 | MAT0025 | MAT0026 | MAT0027 | MAT0028 | MAT0030 | situacao |
|--------|---------|---------|---------|---------|---------|---------|---------|----------|
| 114911 | 0       | 0       | 0       | 2       | 2       | 0       | 1       | FORMADO  |
| 133230 | 0       | 0       | 0       | 2       | 2       | 0       | 2       | FORMADO  |
| 414037 | 0       | 0       | 0       | 1       | 1       | 0       | 1       | FORMADO  |
| 419545 | 0       | 0       | 0       | 1       | 1       | 0       | 1       | FORMADO  |
| 445579 | 0       | 0       | 0       | 1       | 1       | 0       | 2       | EVADIDO  |
| 445743 | 0       | 0       | 0       | 1       | 1       | 0       | 1       | FORMADO  |
| 445767 | 0       | 0       | 0       | 1       | 1       | 0       | 1       | FORMADO  |
| 445843 | 0       | 1       | 0       | 1       | 1       | 0       | 2       | FORMADO  |
| 445965 | 1       | 0       | 0       | 1       | 2       | 0       | 1       | FORMADO  |
| 446038 | 0       | 0       | 0       | 1       | 1       | 0       | 1       | FORMADO  |
| 446076 | 0       | 0       | 0       | 2       | 1       | 0       | 2       | FORMADO  |
| 446081 | 0       | 0       | 0       | 1       | 1       | 0       | 2       | EVADIDO  |
| 446320 | 0       | 1       | 0       | 1       | 1       | 0       | 1       | FORMADO  |
| 446545 | 0       | 0       | 0       | 1       | 2       | 0       | 2       | FORMADO  |
| 446739 | 0       | 0       | 0       | 2       | 2       | 0       | 2       | FORMADO  |
| 446743 | 0       | 0       | 0       | 2       | 4       | 0       | 2       | EVADIDO  |
| 446825 | 0       | 0       | 0       | 1       | 1       | 0       | 1       | FORMADO  |
| 447011 | 0       | 0       | 0       | 2       | 3       | 2       | 2       | EVADIDO  |
| 447036 | 0       | 0       | 0       | 1       | 1       | 0       | 1       | FORMADO  |
| 447110 | 0       | 0       | 0       | 1       | 3       | 0       | 1       | FORMADO  |

Figura 6.1: Matriz de Fatores com as disciplinas e situação do aluno.

### 6.1.2 Construção dos Modelos

Este trabalho utilizou 5 algoritmos, sendo três baseados em Árvores de Decisão: C5.0; o Random Forest e o RPART. Os outros algoritmos são: Regressão Logística e Rede Neural. A linguagem de programação utilizada para a criação dos modelos foi o R na plataforma de desenvolvimento R Studio.

No processo de criação dos modelos, o treinamento consistiu em fornecer aos algoritmos a Matriz de Fatores, que, nesta etapa, continha as disciplinas cursadas e a situação do aluno. Todos os dados que compõem a matriz foram separados por curso, currículo, ano e período de entrada na Universidade, formando as coortes das turmas.

Uma vez gerado os modelos, o procedimento de teste consistiu em fornecer a novamente a mesma Matriz de Fatores, contendo as disciplinas cursadas pelos alunos ativos, suprimindo a variável dependente “SITUACAO”. Para avaliar precisão do modelo foi usando a Métrica F1 Score.

### 6.1.3 Avaliação dos Modelos

No processo de avaliação dos modelos com base no F1 Score, dois comportamentos foram identificados:

- modelos que não apresentaram resultado, pois não conseguiram gerar nenhuma previsão com base nos dados fornecidos. Esse comportamento ocorre quando o modelo não consegue ter informação suficiente para determinar um padrão nos dados;
- modelos com valores baixos de F1 Score, caracterizando modelos com baixa precisão de previsão.

Os modelos que estavam na situação apresentada foram excluídos da etapa de previsão. Desta forma, todos os modelos com valor de nulo, 0 ou apresentaram previsões F1 Score abaixo de 70% foram descartados.

O total de alunos no conjunto de dados da FGA somou 1.815 alunos, porém após a eliminação das turmas que não foram passíveis de formar modelos válidos esse número passou para 786, representando 43% dos alunos que participaram do processo de treinamento dos modelos. Estando estes distribuídos em 47 coortes de alunos conforme apresentados na Tabela 6.1.

Tabela 6.1: Total de alunos por cursos e coortes analisados na fase de avaliação.

| <b>Currículo</b> | <b>Código</b> | <b>Curso</b>            | <b>Ano/Período</b> | <b>Total</b> |
|------------------|---------------|-------------------------|--------------------|--------------|
| 2012.1           | 60836/-2      | ENGENHARIA AEROESPACIAL | 2011/2             | 6            |
| 2012.1           | 60836/-2      | ENGENHARIA AEROESPACIAL | 2012/2             | 25           |
| 2012.1           | 60836/-2      | ENGENHARIA AEROESPACIAL | 2013/2             | 13           |
| 2012.1           | 60836/-2      | ENGENHARIA AEROESPACIAL | 2014/2             | 17           |
| 2012.1           | 60836/-2      | ENGENHARIA AEROESPACIAL | 2015/2             | 23           |
| 2012.1           | 60836/-2      | ENGENHARIA AEROESPACIAL | 2016/2             | 15           |
| 2015.2           | 6297/-2       | ENGENHARIA AUTOMOTIVA   | 2015/2             | 11           |
| 2015.2           | 6297/-2       | ENGENHARIA AUTOMOTIVA   | 2016/2             | 7            |
| 2015.2           | 6297/-2       | ENGENHARIA AUTOMOTIVA   | 2017/2             | 4            |
| 2008.2           | 6297/-5       | ENGENHARIA AUTOMOTIVA   | 2009/1             | 9            |
| 2008.2           | 6297/-5       | ENGENHARIA AUTOMOTIVA   | 2010/1             | 7            |
| 2008.2           | 6297/-5       | ENGENHARIA AUTOMOTIVA   | 2011/1             | 15           |
| 2008.2           | 6297/-5       | ENGENHARIA AUTOMOTIVA   | 2012/1             | 22           |
| 2008.2           | 6297/-5       | ENGENHARIA AUTOMOTIVA   | 2013/1             | 13           |
| 2008.2           | 6297/-5       | ENGENHARIA AUTOMOTIVA   | 2014/1             | 2            |
| 2018.1           | 6289/1        | ENGENHARIA DE ENERGIA   | 2018/1             | 2            |
| 2016.2           | 6289/-2       | ENGENHARIA DE ENERGIA   | 2013/2             | 2            |
| 2016.2           | 6289/-2       | ENGENHARIA DE ENERGIA   | 2015/1             | 7            |
| 2016.2           | 6289/-2       | ENGENHARIA DE ENERGIA   | 2017/1             | 3            |
| 2008.2           | 6289/-4       | ENGENHARIA DE ENERGIA   | 2008/2             | 20           |
| 2008.2           | 6289/-4       | ENGENHARIA DE ENERGIA   | 2009/2             | 23           |
| 2008.2           | 6289/-4       | ENGENHARIA DE ENERGIA   | 2010/2             | 30           |
| 2008.2           | 6289/-4       | ENGENHARIA DE ENERGIA   | 2011/2             | 42           |
| 2008.2           | 6289/-4       | ENGENHARIA DE ENERGIA   | 2012/2             | 28           |
| 2008.2           | 6289/-4       | ENGENHARIA DE ENERGIA   | 2013/2             | 5            |
| 2008.2           | 6289/-4       | ENGENHARIA DE ENERGIA   | 2014/2             | 6            |
| 2008.2           | 6289/-4       | ENGENHARIA DE ENERGIA   | 2015/2             | 5            |
| 2017.1           | 6360/1        | ENGENHARIA DE SOFTWARE  | 2015/2             | 4            |
| 2017.1           | 6360/1        | ENGENHARIA DE SOFTWARE  | 2016/2             | 12           |
| 2017.1           | 6360/1        | ENGENHARIA DE SOFTWARE  | 2017/2             | 22           |
| 2008.2           | 6360/-2       | ENGENHARIA DE SOFTWARE  | 2009/1             | 17           |

Continua...

| <b>Currículo</b> | <b>Código</b> | <b>Curso</b>           | <b>Ano/Período</b> | <b>Total</b> |
|------------------|---------------|------------------------|--------------------|--------------|
| 2008.2           | 6360/-2       | ENGENHARIA DE SOFTWARE | 2010/1             | 26           |
| 2008.2           | 6360/-2       | ENGENHARIA DE SOFTWARE | 2011/1             | 36           |
| 2008.2           | 6360/-2       | ENGENHARIA DE SOFTWARE | 2012/1             | 21           |
| 2008.2           | 6360/-2       | ENGENHARIA DE SOFTWARE | 2013/1             | 34           |
| 2008.2           | 6360/-2       | ENGENHARIA DE SOFTWARE | 2014/1             | 39           |
| 2008.2           | 6360/-2       | ENGENHARIA DE SOFTWARE | 2015/1             | 46           |
| 2008.2           | 6360/-2       | ENGENHARIA DE SOFTWARE | 2016/1             | 30           |
| 2020.1           | 6131/1        | ENGENHARIA ELETRÔNICA  | 2014/1             | 3            |
| 2016.1           | 6131/-2       | ENGENHARIA ELETRÔNICA  | 2017/1             | 2            |
| 2008.2           | 6131/-4       | ENGENHARIA ELETRÔNICA  | 2009/1             | 10           |
| 2008.2           | 6131/-4       | ENGENHARIA ELETRÔNICA  | 2010/1             | 14           |
| 2008.2           | 6131/-4       | ENGENHARIA ELETRÔNICA  | 2011/1             | 27           |
| 2008.2           | 6131/-4       | ENGENHARIA ELETRÔNICA  | 2012/1             | 23           |
| 2008.2           | 6131/-4       | ENGENHARIA ELETRÔNICA  | 2013/1             | 37           |
| 2008.2           | 6131/-4       | ENGENHARIA ELETRÔNICA  | 2014/1             | 14           |
| 2008.2           | 6131/-4       | ENGENHARIA ELETRÔNICA  | 2015/1             | 7            |
|                  |               |                        |                    | 786          |

Fonte: O autor

Após definir os modelos, uma nova rotina foi criada para realizar as previsões. Para cada modelo gerado, foi fornecida a matriz de disciplinas dos alunos ativos, sendo esse conjunto de dados compatível com a coorte do modelo em questão. Logo após o término desse processo, uma lista com os alunos com a previsão de evasão foi criada e sua análise foi feita na seção seguinte.

#### **6.1.4 Análises nas Previsões**

O número de previsões totalizou 1145. Ao excluir os alunos que estão em situação de evasão em mais de um modelo, temos 444 alunos com previsão de evasão. O Random Forest (RF) registrou o maior número de alunos, sendo responsável por 24,19% das previsões, seguido pelo C5 com 20,26% e Rede Neural com 20%, como pode ser observado na Figura 6.2.

Ao avaliar as previsões por curso, a Engenharia de Software foi responsável por 53,10% do total geral de previsões, sendo seguido pela Engenharia Aeroespacial com 18,77%. A Figura 6.3 mostra a distribuição alunos previstos por curso e modelo na FGA.

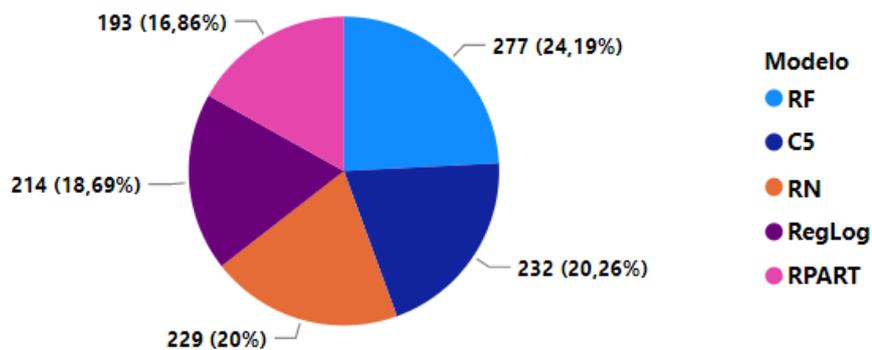


Figura 6.2: Total de previsões geral por modelo na FGA.

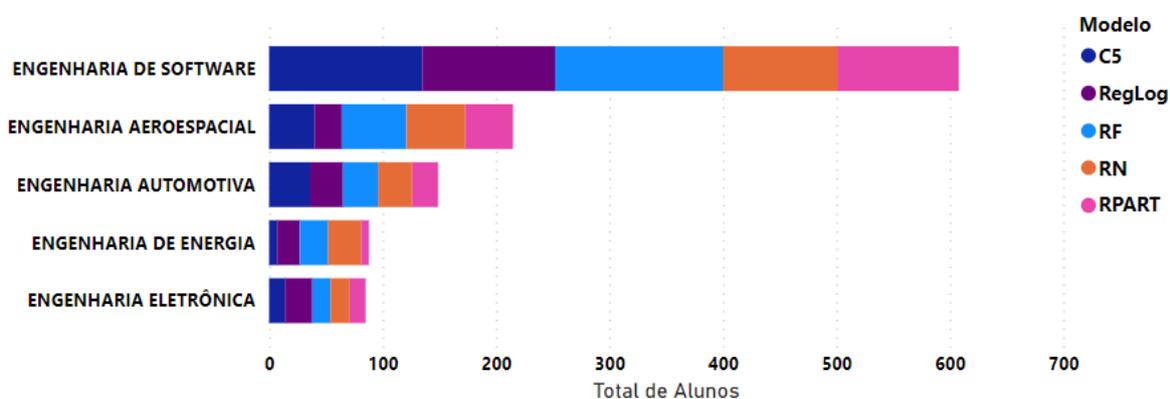


Figura 6.3: Total de previsões por curso e modelo na FGA.

Para analisar os alunos com previsão de evasão, considerou-se a força da previsão, baseada na quantidade de vezes que determinado aluno aparece com previsão de evasão nos modelos. Dessa maneira, o aluno que aparecer nos 5 modelos terá uma força na previsão maior, servindo como um ponto de partida para a análise pelos coordenadores dos cursos. Isso não significa que o aluno previsto num único modelo não deve ser avaliado. Todos os alunos que estão na lista de previsões devem ser considerados.

A Figura 6.4 apresenta uma parte da lista de matrículas identificada pelos modelos, indicando em qual algoritmo o aluno foi identificado com previsão de evasão. Na primeira coluna, temos um número que representa a matrícula do aluno. As seis colunas seguintes mostram em quais algoritmos o aluno foi identificado como evadido. São eles: C5.0, Random Forest (RF), Arvore de Decisão (RPART), Regressão Logística (RegLog) e Rede Neural (RN). Na última coluna, apresentamos o total de ocorrências por modelo em que esse aluno foi identificado com previsão de evasão.

A lista apresentada foi confrontada com a lista de alunos que apresenta a situação deles em 19 de dezembro de 2022. Dessa forma, foram identificados 5 alunos em situação de evasão. Sendo 2 estudantes do curso de Engenharia de Software e 3 da Engenharia Aeroespacial.

Neste estágio, apenas as disciplinas foram consideradas variáveis independentes. Na seção seguinte, novas variáveis foram adicionadas para analisar o comportamento das previsões.

| Matricula         | C50        | RF         | RPART      | RegLog     | RN         | Total       |
|-------------------|------------|------------|------------|------------|------------|-------------|
| 28949282572242770 | 1          | 1          | 1          | 1          | 1          | 5           |
| 28952305785520440 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32404071007870860 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32404176134560000 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32405480151709280 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32405700490715560 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32405784378148490 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32405962954332970 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32406433519336900 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32406524968479380 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32406686984993480 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32407107149707080 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32408005814484170 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32409217015411200 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32410504571297330 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32410895955909700 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32419026992594020 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32419377016269490 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32419642416138920 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32419889811568080 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32422944340620730 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32424567095196440 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32436107614273440 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32449478195187770 | 1          | 1          | 1          | 1          | 1          | 5           |
| 22542385943088400 | 1          | 1          |            | 1          | 1          | 4           |
| 22543173290520960 | 1          | 1          |            | 1          | 1          | 4           |
| 25636248662618170 | 1          | 1          | 1          |            | 1          | 4           |
| 25638645293559840 | 1          | 1          | 1          |            | 1          | 4           |
| 25640936031759880 | 1          | 1          | 1          |            | 1          | 4           |
| 25642807640503520 | 1          | 1          | 1          |            | 1          | 4           |
| <b>Total</b>      | <b>232</b> | <b>277</b> | <b>193</b> | <b>214</b> | <b>229</b> | <b>1145</b> |

Figura 6.4: Listagem de matrículas previstas por algoritmos.

### 6.1.5 Inclusão de novos fatores

A partir dados extraídos dos SIGAA e SIGRA foi possível recuperar, de forma íntegra, o sexo, a idade, situação do aluno, forma de ingresso, ano de ingresso, curso, período do

curso, disciplinas cursadas, notas nas disciplinas e tipo de integralização.

Nesta etapa, foram incluídos na Matriz de Fatores: sexo; idade; forma de ingresso e o período do curso. Estes dados foram categorizados da seguinte forma: o sexo recebeu o valor 1 e 2 para feminino e masculino, respectivamente. O período do curso recebeu também os valores 1 e 2 para diurno e noturno. Para a forma ingresso Vestibular, Programa de Avaliação Seriada (PAS), Exame Nacional do Ensino Médio (ENEM), Portador de Diploma de Curso Superior, Mudança de Curso, Transferência Obrigatória, Transferência Facultativa e Outros <sup>1</sup>, foram categorizados com os valores de 1 a 8. Assim, a Matriz de Fatores, que continha o código das disciplinas, foi acrescida das novas variáveis. A estrutura da Matriz de Fatores, antes de passar pelo processo de categorização, é apresentada na Figura 6.5.

| FGA0250 | FGA0278 | IFD0171 | IFD0173 | MAT0025 | MAT0026 | MAT0031 | sexo | idade | forma_ingresso                | periodo_curso |
|---------|---------|---------|---------|---------|---------|---------|------|-------|-------------------------------|---------------|
| 0       | 0       | 0       | 0       | 0       | 0       | 0       | M    | 28    | MUDANÇA DE CURSO              | D             |
| 0       | 0       | 0       | 0       | 0       | 0       | 0       | M    | 25    | MUDANÇA DE CURSO              | D             |
| 1       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 26    | MUDANÇA DE CURSO              | D             |
| 1       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 25    | MUDANÇA DE CURSO              | D             |
| 1       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 23    | MUDANÇA DE CURSO              | D             |
| 0       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 25    | OUTROS                        | D             |
| 1       | 0       | 1       | 0       | 0       | 0       | 0       | M    | 26    | OUTROS                        | D             |
| 0       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 23    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 1       | 0       | 0       | 0       | 0       | 0       | F    | 22    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 22    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 0       | 0       | 0       | 0       | 0       | 0       | M    | 23    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 21    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 0       | 0       | 0       | 0       | 0       | 0       | M    | 22    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 1       | 0       | 0       | 0       | 0       | 0       | F    | 24    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 1       | 0       | 1       | 0       | 0       | 0       | M    | 23    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 0       | 1       | 0       | 0       | 0       | 0       | M    | 23    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 1       | 1       | 0       | 0       | 0       | 0       | 0       | F    | 23    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 23    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 1       | 0       | 0       | 0       | 0       | 0       | M    | 22    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |
| 0       | 0       | 1       | 0       | 0       | 0       | 0       | M    | 22    | PROGRAMA DE AVALIAÇÃO SERIADA | D             |

Figura 6.5: Novos fatores incluídos.

Uma vez que a matriz de fatores foi criada com as novas informações dos alunos e entregue aos algoritmos para a geração dos modelos, todo o processo para as previsões foi repetido com os novos dados.

Algumas diferenças foram observadas na fase de treinamento e geração dos modelos: a quantidade de alunos que participaram da geração dos modelos passou de 786 para 805; e as coortes passaram de 47 para 48. Isso indica que as novas variáveis forneceram informações aos modelos que permitiram gerar novas previsões antes não encontradas.

O número total de previsões de evasão foi de 1182. Filtrando pelas matrículas que aparecem repetidas em mais de um modelo, esse número foi de 498 alunos.

Desta vez houve uma melhora nas previsões realizadas pela Rede Neural (RN) que foi de 29,27%, sendo seguida pela Regressão Logística (RegLog) com 20,14% e o C5, que foi o terceiro melhor resultado empatado com o Random Forest (RF), com 18,02% conforme pode ser visto na Figura 6.6.

<sup>1</sup>A descrição Outros contempla todos dos casos que não foram citados como forma de entrada.

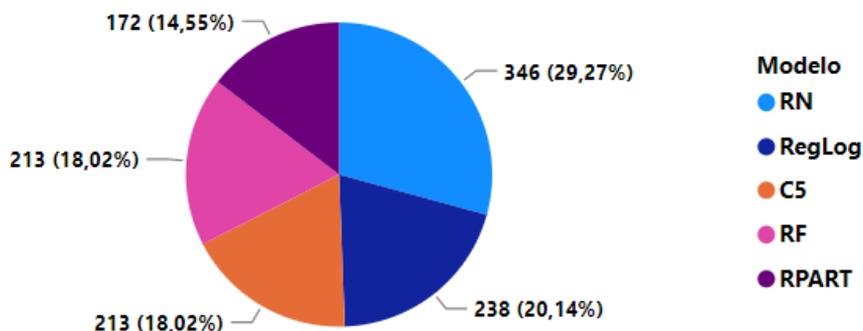


Figura 6.6: Total de previsões geral por modelo na FGA com novos fatores.

A Figura 6.7 mostra que uma grande parte das previsões continuaram sendo realizadas para o curso de Engenharia de Software, onde esse número representa 56,59% das previsões, seguido novamente pelo curso de Engenharia Aeroespacial com 18,10%.

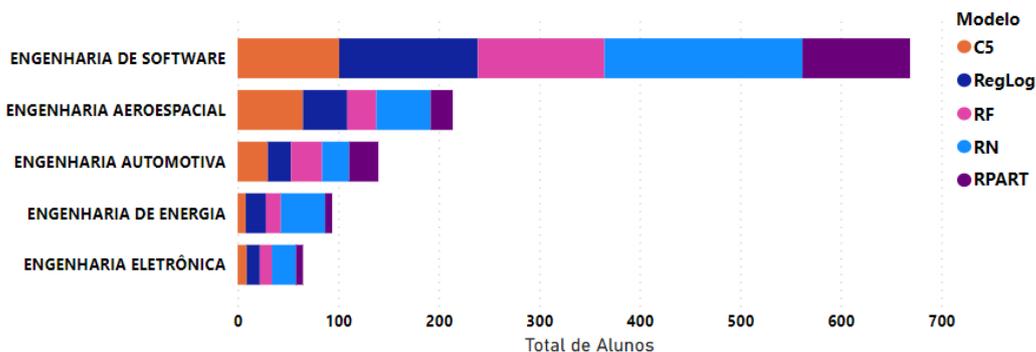


Figura 6.7: Total de previsões por curso e modelo na FGA com novos fatores.

A média de idade dos alunos nas previsões ficou em 24 anos, o percentual de alunos masculinos foi de 81,33% e feminino 18,67%. No caso do gênero, esse número já era de certa forma esperado, pois o conjunto de dados de alunos ativos fornecido para as previsões já estava apresentando uma diferença bem próxima ao que foi gerado nas previsões, sendo 80,21% de homens e 19,79% de mulheres.

Os estudantes que ingressam na universidade pelo Programa de Avaliação Seriada (PAS) são os que apresentam o maior número de previsões, enquanto o Vestibular aparece em segundo lugar. Na Figura 6.8 é apresentada distribuição completa da forma de entrada das previsões.

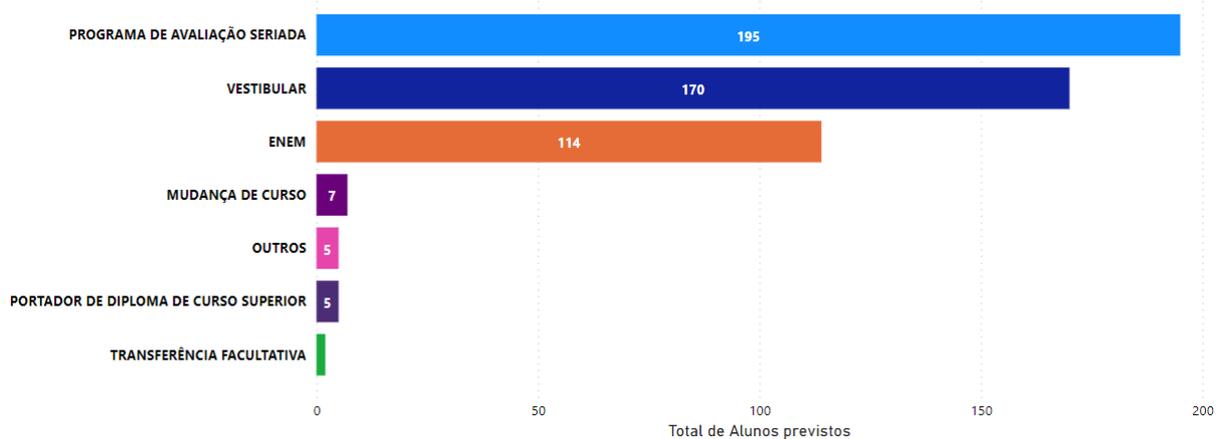


Figura 6.8: Total de previsões por forma de entrada na FGA.

Uma nova lista de alunos com previsão de evasão foi criada e apresentada na Figura 6.9. Esta lista, formada a partir das previsões realizadas com as novas variáveis fornecidas aos algoritmos, segue a mesma estrutura usada nas previsões somente com as disciplinas.

Nela, é possível verificar em quais algoritmos o aluno foi identificado e quais foram esses algoritmos. Nesta lista de alunos, é possível identificar os alunos previstos por mais de um modelo, reforçando o que chamamos de Força de Previsão, podendo ser usada para priorizar os alunos que constam da lista de previsões.

Ao comparar a lista de previsões com as disciplinas e a lista de previsões com as novas variáveis, observou-se uma diminuição no número de alunos que aparecem em 5 modelos. A primeira listagem apresentou 34 alunos, enquanto a segunda somente 8.

Essa queda pode significar que o MAGRA melhorou a precisão nas previsões. Ao analisar a lista de alunos no conjunto de dados de 19 de dezembro de 2022 que mostra a situação nesta data, houve um aumento no número de previsões confirmadas.

O número total de alunos confirmados em situação de evasão foi de 8, sendo 4 do curso de Engenharia de Software, 3 da Engenharia Aeroespacial e 1 da Engenharia Eletrônica.

A seção seguinte descreve o comportamento do MAGRA, que está sendo submetido a mais dados, e as análises das previsões na Universidade de Brasília.

## 6.2 2º Cenário — Universidade de Brasília

O conjunto de dados extraído da base de dados do SIGAA e do SIGRA constam 110 cursos, sendo 97 diurnos e 33 noturnos. O total de alunos foi de 58.195, sendo 26.071 masculino e 32.124 feminino, com média de idade de 25 anos para as mulheres e 26 para os homens. Neste conjunto de dados a forma de ingresso por Vestibular apresenta o maior número de alunos, somando mais de 19 mil e o Programa de Avaliação Seriada (PAS)

| Matricula         | C50        | RF         | RPART      | RegLog     | RN         | Total       |
|-------------------|------------|------------|------------|------------|------------|-------------|
| 28901506559633760 | 1          | 1          | 1          | 1          | 1          | 5           |
| 28914790191815020 | 1          | 1          | 1          | 1          | 1          | 5           |
| 28922284154049840 | 1          | 1          | 1          | 1          | 1          | 5           |
| 28948937179152380 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32404514917276680 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32406498685838600 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32410444801638120 | 1          | 1          | 1          | 1          | 1          | 5           |
| 32411911694617220 | 1          | 1          | 1          | 1          | 1          | 5           |
| 22504815557634600 | 1          | 1          |            | 1          | 1          | 4           |
| 22542385943088400 | 1          | 1          |            | 1          | 1          | 4           |
| 25636248662618170 | 1          |            | 1          | 1          | 1          | 4           |
| 25636718757237020 | 1          |            | 1          | 1          | 1          | 4           |
| 28902979156772640 | 1          | 1          |            | 1          | 1          | 4           |
| 28903830906945290 | 1          | 1          |            | 1          | 1          | 4           |
| 28904620104633740 | 1          |            | 1          | 1          | 1          | 4           |
| 28904790118471740 | 1          |            | 1          | 1          | 1          | 4           |
| 28905307983703320 |            | 1          | 1          | 1          | 1          | 4           |
| 28905730644057320 | 1          |            | 1          | 1          | 1          | 4           |
| 28908013335403490 | 1          |            | 1          | 1          | 1          | 4           |
| 28913534604278810 | 1          | 1          |            | 1          | 1          | 4           |
| 28919319647716970 | 1          |            | 1          | 1          | 1          | 4           |
| 28921620222034930 | 1          | 1          | 1          |            | 1          | 4           |
| 28922189597693000 | 1          |            | 1          | 1          | 1          | 4           |
| 28938886292034890 | 1          | 1          |            | 1          | 1          | 4           |
| 28939283340250000 | 1          | 1          |            | 1          | 1          | 4           |
| 28940927700030890 | 1          |            | 1          | 1          | 1          | 4           |
| 28940952537604840 | 1          | 1          |            | 1          | 1          | 4           |
| 28941600359736100 | 1          | 1          |            | 1          | 1          | 4           |
| 28947126536379600 |            | 1          | 1          | 1          | 1          | 4           |
| 28948014666427920 |            | 1          | 1          | 1          | 1          | 4           |
| 28952821013431220 | 1          | 1          |            | 1          | 1          | 4           |
| 28955071190638610 | 1          | 1          |            | 1          | 1          | 4           |
| 28955222210576000 | 1          | 1          |            | 1          | 1          | 4           |
| <b>Total</b>      | <b>213</b> | <b>213</b> | <b>172</b> | <b>238</b> | <b>346</b> | <b>1182</b> |

Figura 6.9: Listagem de previsão de alunos da FGA com novas variáveis por algoritmos.

aparece em segundo com mais de 18 mil. A Tabela 6.2 detalha as formas de ingresso presente no conjunto de dados avaliado.

As rotinas construídas para o protótipo executado na FGA foram aproveitadas para as análises com as informações da UnB. As próximas etapas descrevem os procedimentos.

### 6.2.1 Construção da Matriz de Fatores da UnB

O procedimento de construção da Matriz de Fatores seguiu o mesmo procedimento usado no desenvolvimento do protótipo na FGA.

Tabela 6.2: Forma de ingresso presente no conjunto de dados da UnB.

| <b>Forma de ingresso</b>              | <b>Total</b> |
|---------------------------------------|--------------|
| VESTIBULAR                            | 19.253       |
| PROGRAMA DE AVALIAÇÃO SERIADA         | 18.887       |
| ENEM                                  | 14.016       |
| MUDANÇA DE CURSO                      | 1.908        |
| OUTROS                                | 1.600        |
| PORTADOR DE DIPLOMA DE CURSO SUPERIOR | 1.263        |
| TRANSFERÊNCIA OBRIGATÓRIA             | 714          |
| TRANSFERÊNCIA FACULTATIVA             | 554          |
| Total de alunos                       | 58.195       |

### 6.2.2 Construção dos Modelos da UnB

A construção dos modelos seguiu o mesmo procedimento usado para na construção do protótipo na FGA.

Após a execução da rotina de geração dos modelos, percebeu-se uma diminuição no número de cursos, que, anteriormente, era de 110, passando para 84. A redução era, de certa forma, esperada devido à impossibilidade dos algoritmos encontrarem um padrão que pudesse ser construído no grupo de alunos e curso que estava sendo usado para o treinamento dos modelos.

Nessa etapa, somente as disciplinas que os alunos estudaram foram usadas como variáveis independentes. As análises dos resultados gerados durante a elaboração dos modelos serão feitas nas seções seguintes.

### 6.2.3 Avaliação dos Modelos UnB

A avaliação dos modelos seguiu o mesmo procedimento adotado na FGA. Aqueles que retornam F1 Score com o valor 0, nulo ou abaixo de 70% não foram selecionados. Dessa maneira, notamos que 6.817 alunos participaram da criação dos modelos e estão distribuídos em 84 cursos e em 450 coortes. A Figura 6.10 apresenta uma parte da lista de coortes que geraram os modelos válidos.

### 6.2.4 Análises das Previsões UnB

Após avaliar o cenário das previsões geradas, os cursos que conseguiram gerar as previsões diminuíram para 78. Os alunos previstos somam 16.531 matrículas. Ao excluir as matrículas que aparecem em mais de um modelo, o total passou para 6.536 alunos.

| Ano Currículo | Cod. Currículo | Curso                   | Ano/Período Ingresso | Alunos      |
|---------------|----------------|-------------------------|----------------------|-------------|
| 2020.1        | 8117/1         | ADMINISTRAÇÃO           | 2020/2               | 4           |
| 2018.2        | 8117/-2        | ADMINISTRAÇÃO           | 2019/1               | 6           |
| 2017.2        | 8117/-3        | ADMINISTRAÇÃO           | 2018/1               | 13          |
| 2016.1        | 8117/-4        | ADMINISTRAÇÃO           | 2016/2               | 27          |
| 2018.2        | 8150/1         | ADMINISTRAÇÃO           | 2018/2               | 13          |
| 2018.2        | 8150/1         | ADMINISTRAÇÃO           | 2019/2               | 11          |
| 2011.1        | 8150/-5        | ADMINISTRAÇÃO           | 2012/1               | 2           |
| 2011.2        | 6165/-2        | AGRONOMIA               | 2017/1               | 26          |
| 2016.1        | 5126/1         | ARQUITETURA E URBANISMO | 2017/1               | 13          |
| 2016.1        | 5126/1         | ARQUITETURA E URBANISMO | 2018/1               | 4           |
| 2013.1        | 5126/-2        | ARQUITETURA E URBANISMO | 2014/2               | 28          |
| 2016.1        | 5673/1         | ARQUITETURA E URBANISMO | 2016/1               | 13          |
| 2010.1        | 5673/-2        | ARQUITETURA E URBANISMO | 2010/2               | 14          |
| 2010.1        | 5673/-2        | ARQUITETURA E URBANISMO | 2011/2               | 17          |
| 2010.1        | 5673/-2        | ARQUITETURA E URBANISMO | 2012/2               | 21          |
| 2010.1        | 5673/-2        | ARQUITETURA E URBANISMO | 2013/2               | 19          |
| 2010.1        | 5673/-2        | ARQUITETURA E URBANISMO | 2014/2               | 25          |
| 2010.1        | 5673/-2        | ARQUITETURA E URBANISMO | 2015/2               | 25          |
| 2013.1        | 8192/1         | ARQUIVOLOGIA            | 2017/2               | 26          |
| 2016.2        | 5711/-3        | ARTES CÊNICAS           | 2018/1               | 2           |
| 2009.1        | 5720/-3        | ARTES CÊNICAS           | 2012/1               | 9           |
| 2009.1        | 5720/-3        | ARTES CÊNICAS           | 2013/1               | 7           |
| 2007.2        | 4936/-2        | ARTES VISUAIS           | 2007/2               | 3           |
| 2007.2        | 4936/-2        | ARTES VISUAIS           | 2011/1               | 52          |
| 2018.1        | 5649/1         | ARTES VISUAIS           | 2020/1               | 3           |
| 2016.1        | 5649/-2        | ARTES VISUAIS           | 2016/2               | 7           |
| 2014.1        | 5649/-3        | ARTES VISUAIS           | 2015/2               | 13          |
| 2015.2        | 2771/1         | BIOTECNOLOGIA           | 2016/1               | 30          |
| 2015.2        | 2771/1         | BIOTECNOLOGIA           | 2017/1               | 28          |
| 2015.2        | 2771/1         | BIOTECNOLOGIA           | 2018/1               | 16          |
| 2018.2        | 1856/1         | CIÊNCIA DA COMPUTAÇÃO   | 2018/2               | 9           |
| <b>Total</b>  |                |                         |                      | <b>6817</b> |

Figura 6.10: Listagem de modelos válidos por coorte.

Analisando as previsões com base nos modelos, a Rede Neural foi responsável por 25,81% do total geral de previsões, seguida pela Random Forest com 19,92% e C5.0 com 18,63% conforme pode ser visto na Figura 6.11.

Os cursos que apresentaram maior número de previsões foram Administração com, 1109 dos alunos previstos, Direito com 1085 e História com 833. A Figura 6.12 lista os 10 primeiros cursos e a segmentação de previsões por modelo nos cursos.

Seguindo o mesmo procedimento realizado para o protótipo na FGA, as análises nas

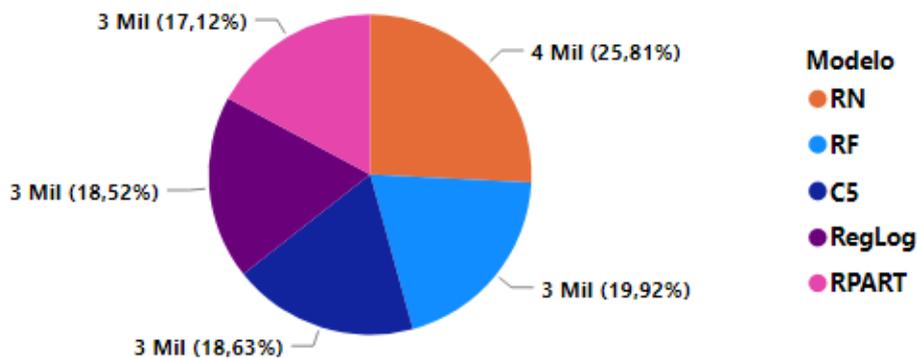


Figura 6.11: Visão Geral das Previsões por modelo na UnB.

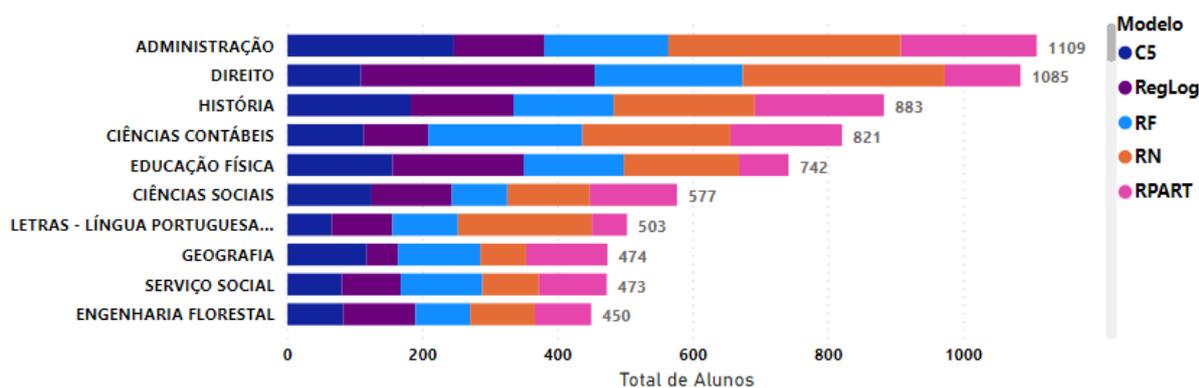


Figura 6.12: Visão Geral das Previsões por curso e modelo na UnB.

previsões começam priorizando os alunos que aparecem com previsão de evasão em 5 modelos. Mas conforme descrito durante as análises no protótipo na FGA, todas as previsões devem ser consideradas, inclusive os alunos que aparecem em apenas um modelo.

Os cursos de Administração e Direito foram escolhidos para análise por estarem nas duas primeiras posições no total geral de previsões conforme apresentado na Figura 6.12.

O conjunto de dados fornecido para a etapa de treinamento dos modelos para o curso de Administração contou com 29 disciplinas obrigatórias cursadas por 1529 alunos. Para as previsões, 1.103 alunos estavam presentes no conjunto de alunos ativos fornecido para os modelos. Desses, 412 alunos foram identificados com a previsão de evasão, representando 37% dos alunos.

Seguindo o procedimento descrito nas análises do protótipo na FGA, os 412 alunos estão em uma lista de alunos, onde alguns deles foram identificados por mais de um

modelo. Isso deve ser considerado um fator determinante para a prioridade de análises por parte dos coordenadores.

Para o curso de Direito as disciplinas obrigatórias somam 53 e 1.819 alunos foram selecionados para compor o conjunto de dados para construir os modelos. Nas previsões foram selecionados 1.267 alunos ativos. Desse total, 476 ou 37% foram indicados pelos modelos com previsão de evasão.

No conjunto de dados retirado do SIGAA que contém a situação dos alunos em 19 de dezembro de 2022 foram confirmados em situação de evasão em 4 alunos de Administração e 12 do Direito.

Uma nova previsão foi feita incluindo as novas variáveis e os procedimentos foram detalhados nas próximas seções.

### **6.2.5 Inclusão de novos fatores UnB**

Novamente, o sexo, a idade, a forma de ingresso e o período do curso foram inseridos no conjunto de informações para realizar as etapas de treinamento e previsão dos modelos.

Na fase de treinamento, o total de cursos que foram passíveis de formar as previsões passou de 84 para 87. Os alunos que participaram do conjunto de dados de treinamento subiram de 6.817 para 7.298 e o total de cortes que era de 450 passou para 486. Isso demonstra que a inclusão das novas variáveis melhorou como os modelos são construídos para as previsões. O comportamento é o mesmo que foi observado no protótipo apresentado na FGA.

### **6.2.6 Análises das Previsões com novas variáveis UnB**

Em relação aos dados gerais das previsões, notamos que a média de idade permaneceu em 23 anos. O gênero apresentou uma distribuição equilibrada, sendo 51,05% para o masculino e 48,95% para o feminino. O número de cursos previstos aumentou de 78 para 82, o número líquido de matrículas previstas (sem as matrículas que aparecem em mais de um modelo) aumentou de 6.536 para 7.379.

A Rede Neural apresentou o melhor percentual de previsões, chegando a 29,97% do total geral, seguida pelo C5.O com 19,29%. A distribuição das previsões por algoritmo foi descrita de forma completa na Figura 6.13.

Em relação à forma de ingresso na universidade, o Programa de Avaliação Seriada apresentou o maior número de previsões, seguido novamente pelo Vestibular, como pode ser observado na Figura 6.14.

Avaliando as previsões por curso, houve uma inversão entre os dois primeiros. Desta vez o curso de Direito apresentou o maior número de previsões e curso de Administração

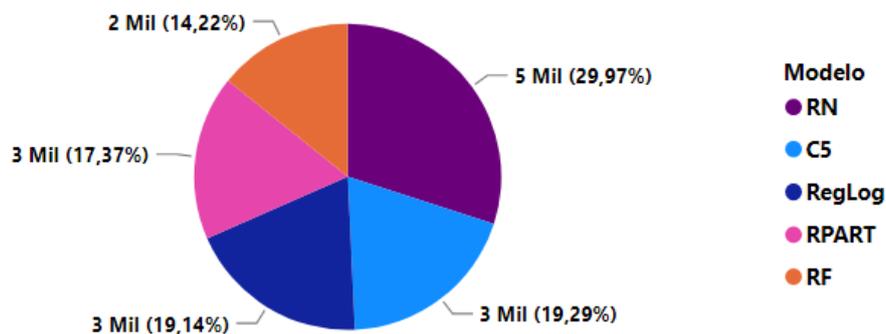


Figura 6.13: Visão Geral das Previsões por modelo na UnB com as novas variáveis.

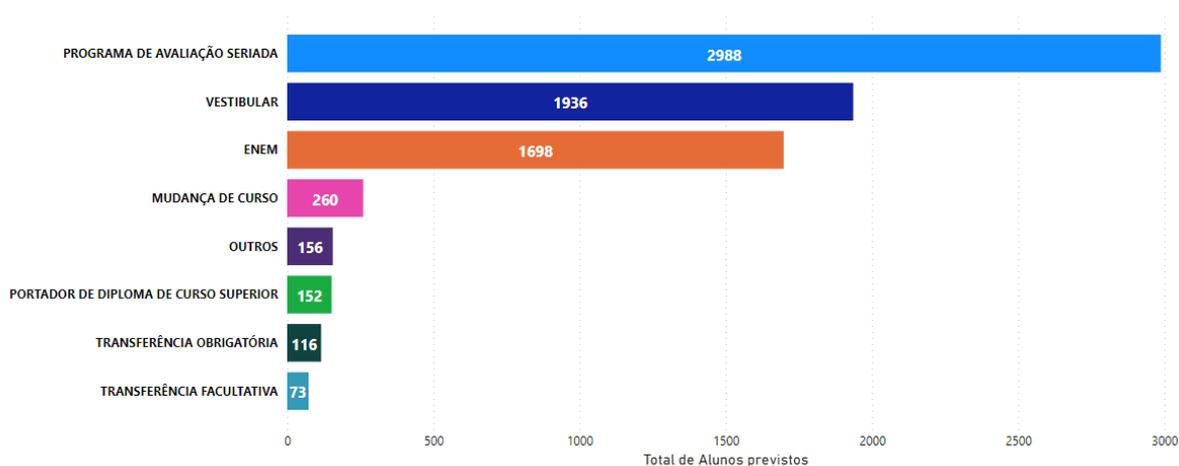


Figura 6.14: Total de previsões por forma de ingresso na UnB.

ficou em segundo. A Figura 6.15 apresenta a posição dos 10 primeiros cursos com mais previsões por modelo e curso.

As análises com as novas variáveis seguiram o mesmo critério da análise somente com as disciplinas, ou seja, foram realizadas as análises descritivas para os dois primeiros cursos.

O MAGRA previu, sem as novas variáveis, um total de 476 alunos para o curso de Direito. Com as novas variáveis, esse número passou para 517, distribuídos em 285 alunos do noturno e 232 do diurno. O sexo masculino representou 59% e 41% feminino. A média de idade dos alunos previstos é de 25 anos. A forma de ingresso pelo Programa de Avaliação Seriada (PAS) somou 204 e em segundo o Exame Nacional do Ensino Médio (ENEM) com 145 alunos.

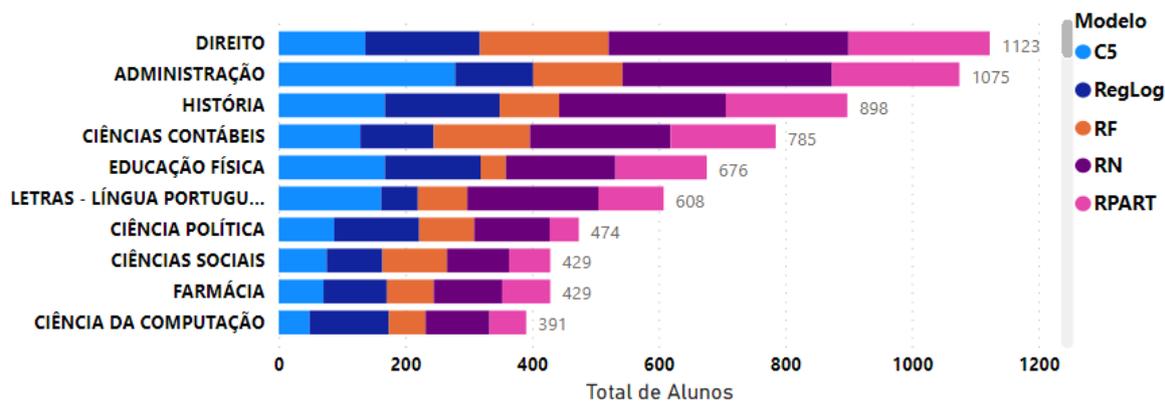


Figura 6.15: Visão Geral das Previsões por curso e modelo na UnB com novas variáveis.

Para o curso de Administração usando apenas disciplinas, o MAGRA previu 412 alunos, com as novas informações esse número passou para 454. Desse número, 58% são dos sexos masculino e 42% feminino, com média de idade em 23 anos. O maior número de alunos é do noturno com 237 e o diurno 217. A forma de ingresso pelo PAS somou 192 e o ENEM ficou novamente em segundo com 146.

A inclusão das novas informações permitiu perceber que houve uma contribuição para melhorar as características das turmas e, conseqüentemente, os modelos conseguiram identificar novos padrões, o que refletiu no aumento no número de alunos previstos.

As disciplinas mostraram serem instrumentos muito eficazes para identificar os alunos que estão em um processo de retenção. Aliar as disciplinas cursadas, as características pessoais e educacionais podem melhorar as previsões, seja para evasão ou retenção.

No conjunto de dados que contém a situação dos alunos de Direito até 19 de dezembro de 2022, foram confirmados 12 casos de evasão, sendo 6 masculinos e 6 femininos. Desse total, a forma de ingresso pelo ENEM somou 5 alunos, Vestibular com 3 e Mudança de Curso, PAS, Transferência Obrigatória e Outros com 1 cada.

Para curso de Administração foram confirmados 7 alunos, sendo 5 do sexo masculino e 2 do sexo feminino. Cinco desses alunos ingressaram pelo ENEM, 1 por Mudança de Curso e 1 Portador de Diploma Superior.

### 6.2.7 Implicações Práticas

O Modelo de Análise na Graduação (MAGRA) gera relatórios com a lista de alunos identificados com o perfil de evasão. Esses relatórios podem ser gerados em formato PDF e fornecidos aos coordenadores de cursos no início de cada semestre, após o período de matrícula nas disciplinas.

Procedimentos de feedback sobre o comportamento do MAGRA precisam ser definidos com base nas informações provenientes de coordenadores e professores, a fim de realizar ajustes nos processos de construção dos modelos. Isso permitirá a geração de versões do MAGRA mais precisas no futuro.

Para as futuras versões do MAGRA, é necessário integrá-lo ao SIGAA, para que os professores e coordenadores possam visualizar os alunos identificados na listagem da turma. Existem pacotes, como o RServe <sup>2</sup>, que podem ser instalados no servidor da aplicação Java que roda o SIGAA. Isso permite a execução de um servidor R e a comunicação com ele a partir do Java usando o protocolo TCP/IP. Uma vez realizada a integração, a rotina de previsão pode ser automatizada com base nas datas de matrícula nas disciplinas no início de cada semestre. As análises descritas estão no campo teórico, e a complexidade da integração e outros obstáculos devem ser discutidos com a equipe técnica que trabalha com o SIGAA.

Isso pode incluir restrições técnicas, complexidade da integração, necessidade de colaboração com outros departamentos ou equipes, entre outros obstáculos potenciais.

Considere aspectos como o treinamento de pessoal para utilizar e manter o sistema, a coleta e gerenciamento dos dados necessários para alimentar o modelo de previsão, e os possíveis impactos nos procedimentos existentes da instituição em que o sistema será implantado.

---

<sup>2</sup>O Rserve atua como um servidor de soquete (TCP/IP ou soquetes locais) que permite o envio de solicitações binárias ao R - <https://cran.r-project.org/web/packages/Rserve/index.html>

# Capítulo 7

## Conclusão

O trabalho avaliou o fenômeno da evasão nos cursos de graduação da Universidade de Brasília, buscando fatores que contribuem para a identificação precoce dos alunos com características de evasão.

A execução de uma Revisão Sistemática da Literatura (RSL) gerou resultados que direcionaram as decisões de quais os fatores poderiam ser avaliados na instituição para compor o conjunto de dados dos alunos. Serviu de base também na tomada de decisão sobre quais ferramentas usar durante o processo visando atingir os objetos do trabalho.

Após identificar os indicadores registrados nas bases educacionais e as ferramentas que seriam usadas durante as etapas de predição, elaborou-se o Modelo de Análise na Graduação (MAGRA)

O propósito do MAGRA é analisar as informações geradas pelos alunos usando algoritmos de Machine Learning visando realizar previsões de evasão. Uma vez gerados as previsões, elas devem servir de ferramenta de apoio para os coordenadores de cursos na identificação previa dos casos de evasão.

O grande desafio desse trabalho foi avaliar o ingente volume de dados gerados pelos estudantes, buscando informações que ajudassem a identificar alunos propensos a levar mais tempo para concluir o curso ou abandonar a universidade.

O presente trabalho contribui no sentido de lançar luz sobre dois pontos importantes para o estudo e, conseqüentemente, o desenvolvimento de modelos machine learning para previsão de evasão ou retenção. O primeiro ponto está na falta de integração entre os sistemas acadêmicos. A adoção do SIGAA trouxe uma simplificação nas atividades acadêmicas, permitiu acesso a várias informações sem a necessidade de usar múltiplas plataformas. No entanto, é preciso direcionar esforços para desenvolver rotinas para integração de outros sistemas, por exemplo, os ambientes de ensino remoto. Isso permitiria avaliar a saúde acadêmica do estudante em tempo real. O segundo ponto diz respeito à disponibilidade de acesso a dados de sistemas acadêmicos de forma proativa. Embora

meu cargo na UnB facilite o acesso a alguns sistemas, ainda assim tivemos dificuldades para acessar informações. Isso ocorreu tanto devido à falta de um formato apropriado para consumo, quanto à restrição de acesso apenas a perfis específicos.

Evidenciamos que a falta de acesso a dados abrangentes e integrados pode dificultar a identificação de padrões e fatores que contribuem para a evasão e retenção por parte dos pesquisadores. Isso compromete a capacidade da universidade de agir preventivamente e intervir de forma eficaz para reduzir os índices de abandono dos cursos.

Particionar o trabalho em dois cenários permitiu realizar testes com um universo menor de informação e depois ampliar em outro momento quando as estruturas do MAGRA já estivessem mais consolidadas.

O primeiro cenário realizado na Faculdade do Gama (FGA) funcionou como um protótipo, permitindo realizar a construção das estruturas para extração dos dados, treinamento, teste e avaliação dos modelos.

O segundo cenário tinha como objetivo contemplar todos os cursos da Universidade de Brasília que fossem passíveis de formar modelos válidos para as previsões.

Nos dois cenários, foi possível notar que os testes realizados na primeira etapa, somente com as disciplinas que os alunos cursaram, mostraram que o número de vezes que os alunos cursam uma disciplina pode ser um indicador da dificuldade desse aluno em transpor a disciplina no tempo determinado para a sua execução.

Na etapa dos testes com as novas variáveis, houve uma melhora no número de modelos válidos e, conseqüentemente, aumento nos cursos e turmas analisadas. A situação foi percebida na condução dos estudos na Faculdade do Gama e na Universidade de Brasília.

Os resultados apresentados pelos modelos foram de certa forma influenciados pela época em que foram retirados das bases educacionais. A pandemia trouxe uma série de mudanças as quais introduzem ruídos na massa de dados que devem ser considerados. Essas mudanças são vistas nas resoluções geradas pela UnB durante esse período. Os dados gerados até 2019 podem não ter as mesmas características dos dados entre 2020 e 2021. A expectativa dos resultados nesses dois anos é baixa em função dos efeitos da pandemia.

Com as previsões de cada curso, agora é possível elaborar um relatório no início de cada semestre para os coordenadores de curso de graduação presencial da UnB, permitindo um trabalho proativo por parte dos coordenadores.

A última análise com o curso de Direito, onde foram avaliados 1288 alunos, apontou que 40% deles foram previstos pelo MAGRA, com previsão de evasão. É possível considerar que esse número poderia ser mais baixo, fazendo melhorias na estrutura do MAGRA para aumentar a precisão das previsões. Estas melhorias estão descritas de próxima seção.

## 7.1 Trabalhos futuros:

Os Ambientes Virtuais de Aprendizagem, Aprender e o Microsoft Teams, utilizados na Universidade de Brasília, foram avaliados, mas por questões técnicas de acesso aos dados não foram contemplados por este trabalho. Compreendemos que os dados que persistem nesses sistemas podem gerar métricas relevantes para os estudos de evasão para a Universidade e devem ser considerados.

Ajustes devem ser feitos no trato com as disciplinas cursadas. As disciplinas equivalentes podem ter gerado uma leitura errada por partes dos algoritmos na previsão. O mesmo caso pode acontecer disciplinas com crédito concedido que não foram consideradas e isso pode levar a casos de falsos positivos.

Para melhorar o processo de identificação precoce dos alunos com características de evasão, é preciso de desenvolver mecanismos *feedback* por parte dos coordenadores de cursos, pois cada curso com especificidades no currículo gera peculiaridades que devem ser avaliadas visando melhorar a resposta dos algoritmos.

Neste estudo foram usados como variáveis (fatores) preditoras o curso, período do curso, gênero, forma de ingresso do aluno, disciplinas cursadas e idade. Outros 23 fatores identificados na revisão bibliográfica não foram examinados, podendo ser incluídos com objetivos de melhorar as previsões. Além da inclusão de mais variáveis, a execução de procedimentos para identificação de quais variáveis tem o maior peso nas previsões podem também melhorar os resultados.

Todos dos algoritmos de Aprendizagem de Máquina foram usados na sua forma padrão, ou seja, não foram realizados ajustes nos parâmetros dos algoritmos para melhorar as precisões nas previsões.

Ajustar o conjunto de dados selecionados no SIGAA para conter a informação do semestre. Isso pode evitar casos de falso positivos em alunos nos semestres iniciais do curso.

Outro ponto a ser considerado no MAGRA é aumentar o valor de F1 Score usado como limiar de corte dos modelos. Atualmente, está definido em 70%, podendo ser ajustado para valores próximos a 85%.

Para os próximos trabalhos é importante considerar mais uma métrica para garantir a confiabilidade nas previsões realizadas. O teste Kappa fornece testes estatísticos que podem ser usados com esse objetivo.

# Referências

- [1] Silva, Edmilson Cosme, Sergio Freitas, Cristiane Soares Ramos, Amanda Emilly Muniz De Menezes e Leticia Karla Soares Rodrigues De Araujo: *A systematic review of the factors that impact the prediction of retention and dropout in higher education*. International Conference on System Sciences (HICSS), 2023. xiii, xv, 12, 13, 14, 15
- [2] Hosmer Jr, David W, Stanley Lemeshow e Rodney X Sturdivant: *Applied logistic regression*, volume 398. John Wiley & Sons, 2013. xiii, 17, 18
- [3] Maimon, Oded Z e Lior Rokach: *Data mining with decision trees: theory and applications*, volume 81. World scientific, 2014. xiii, 19
- [4] Aggarwal, Charu C *et al.*: *Neural networks and deep learning*. Springer, 10:978–3, 2018. xiii, 20
- [5] De Freitas, Sara, David Gibson, Coert Du Plessis, Pat Halloran, Ed Williams, Matt Ambrose, Ian Dunwell e Sylvester Arnab: *Foundations of dynamic learning analytics: Using university student data to increase retention*. British journal of educational technology, 46(6):1175–1188, 2015. xiii, 22, 23
- [6] Adejo, Olugbenga Wilson e Thomas Connolly: *Predicting student academic performance using multi-model heterogeneous ensemble approach*. Journal of Applied Research in Higher Education, 2018. xiii, 24, 25, 26, 27
- [7] Silva Filho, Roberto Leal Lobo, Paulo Roberto Motejunas, Oscar Hipólito e Maria Beatriz de Carvalho Melo Lobo: *A evasão no ensino superior brasileiro*. Cadernos de pesquisa, 37:641–659, 2007. 1, 2
- [8] Tinto, Vincent: *Dropout from higher education: A theoretical synthesis of recent research*. Review of educational research, 45(1):89–125, 1975. 1, 6
- [9] Bean, John P: *Dropouts and turnover: The synthesis and test of a causal model of student attrition*. Research in higher education, 12(2):155–187, 1980. 1, 6
- [10] Cleveland-Innes, Martha: *Adult student drop-out at post-secondary institutions*. The Review of Higher Education, 17(4):423–445, 1994. 1, 6
- [11] Laws Jr, James E: *The influence of upward bound on freshman grade point average, drop-out rates, mathematics performance, and english performance*. Western Journal of Black Studies, 23(3):139, 1999. 1, 6

- [12] Rumberger, Russell W e Sun Ah Lim: *Why students drop out of school: A review of 25 years of research*. 2008. 1, 6, 7
- [13] Pereira, Alexandre Severino, Teresa Cristina Janes Carneiro, Gutemberg Hespanha Brasil e Maria Auxiliadora de Carvalho Corassa: *Perfil dos alunos retidos dos cursos de graduação presencial da universidade federal do espírito santo*. 2014. 1, 9
- [14] Lenning, Oscar *et al.*: *Retention and attrition: Evidence for action and research*. 1980. 1, 8
- [15] DesJardins, Stephen L, Dennis A Ahlburg e Brian P McCall: *An event history model of student departure*. *Economics of education review*, 18(3):375–390, 1999. 1, 2
- [16] ANDIFES, ABRUEM *et al.*: *Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas: resumo do relatório apresentado a andifes, abruem e sesu/mec pela comissão especial*. *Avaliação: Revista da Avaliação da Educação Superior*, 1(2), 1996. 2, 7, 8, 9, 10
- [17] Gil, Antonio Carlos *et al.*: *Como elaborar projetos de pesquisa*, volume 4. Atlas São Paulo, 2002. 4
- [18] Runeson, Per, Martin Host, Austen Rainer e Bjorn Regnell: *Case study research in software engineering: Guidelines and examples*. John Wiley & Sons, 2012. 4
- [19] Atif, Amara, Deborah Richards e Ayse Bilgin: *A student retention model: empirical, theoretical and pragmatic considerations*. 2013. 7, 8, 11, 12
- [20] Veloso, Tereza Christina MA e Edson Pacheco de Almeida: *Evasão nos cursos de graduação da universidade federal de mato grosso, campus universitário de cuiabá—um processo de exclusão*. *Série-Estudos-Periódico do Programa de Pós-Graduação em Educação da UCDB*, 2002. 7
- [21] Cislighi, Renato *et al.*: *Um modelo de sistema de gestão do conhecimento em um framework para a promoção da permanência discente no ensino de graduação*. 2008. 7, 10
- [22] Aguiar, Everaldo, Nitesh V Chawla, Jay Brockman, G Alex Ambrose e Victoria Goodrich: *Engagement vs performance: using electronic portfolios to predict first semester engineering student retention*. Em *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, páginas 103–112, 2014. 8, 15
- [23] Raju, Dheeraj e Randall Schumacker: *Exploring student characteristics of retention that lead to graduation in higher education using data mining models*. *Journal of College Student Retention: Research, Theory & Practice*, 16(4):563–591, 2015. 8, 12, 15, 16, 19
- [24] Rintala, Ulla, Staffan Andersson e Anna Kaarina Kairamo: *How to retain students in higher engineering education? findings of the attract project*. Em *SEFI Annual Conference, Lisbon, Portugal*. <http://goo.gl/tdVMj5>, 2011. 8

- [25] Pérez, Alexis Matheu, Claudio Ruff Escobar, Marcelo Ruiz Toledo, Luis Benites Gutierrez e Germán Morong Reyes: *Prediction model of first-year student desertion at universidad bernardo o´ higgins (ubo)*. Educação e Pesquisa, 44, 2018. 9
- [26] Pérez, Alfredo, Elizabeth E Grandón, Mónica Caniupán e Gilda Vargas: *Comparative analysis of prediction techniques to determine student dropout: Logistic regression vs decision trees*. Em *2018 37th International Conference of the Chilean Computer Science Society (SCCC)*, páginas 1–8. IEEE, 2018. 9
- [27] Pereira, Alexandre Severino: *Retenção discente nos cursos de graduação presencial da ufes*. 2013. <https://repositorio.ufes.br/handle/10/2471>, acesso em 2022-01-23. 9, 10
- [28] Silva Garcia, Léo Manoel Lopes da, Daiany Francisca Lara e Franciano Antunes: *Análise da retenção no ensino superior: Um estudo de caso em um curso de sistemas de informação*. Revista da Faculdade de Educação, 34(2):15–38, 2020. 9
- [29] Pereira, Alexandre Severino, Teresa Cristina Janes Carneiro, Gutemberg Hespanha Brasil e Maria Auxiliadora de Carvalho Corassa: *Fatores relevantes no processo de permanência prolongada de discentes nos cursos de graduação presencial: um estudo na universidade federal do espírito santo*. Ensaio: Avaliação e Políticas Públicas em Educação, 23:1015–1039, 2015. 9
- [30] Dos Santos, Adilson Pereira, Caroline Nascimento e Jaime Roberto Teixeira Rios: *Estudo da evasão e da retenção nos cursos de engenharia da escola de minas da universidade federal de ouro preto*, 2000. <http://www.abenge.org.br/cobenge/legado/arquivos/19/artigos/175.PDF>, acesso em 2022-01-23. 9, 10
- [31] CORREA, Ana CC, Adriana B NORONHA e Irene K MIURA: *Avaliação da evasão e permanência prolongada em um curso de graduação em administração de uma universidade pública*. Anais do Semead - Seminários de Administração, São Paulo, SP, Brasil, 7, 2004. 10
- [32] Ney, Otávio Abrantes de Sá *et al.*: *Sistemas de informação acadêmica para o controle da evasão*. 2010. 10
- [33] Esmat, Tiffany A e Joshua D Pitts: *Predicting success in an undergraduate exercise science program using science-based admission courses*. Advances in physiology education, 44(2):138–144, 2020. 11, 19
- [34] Tucker, Leslie e Oscar McKnight: *Assessing the validity of college success indicators for the at-risk student: Toward developing a best-practice model*. Journal of College Student Retention: Research, Theory & Practice, 21(2):166–183, 2019. 11
- [35] Atieh, Emily L, Darrin M York e Marc N Muñiz: *Beneath the surface: An investigation of general chemistry students’ study skills to predict course outcomes*. Journal of chemical education, 98(2):281–292, 2020. 11
- [36] Cohen, Richard e Angela M Kelly: *Mathematics as a factor in community college stem performance, persistence, and degree attainment*. Journal of Research in Science Teaching, 57(2):279–307, 2020. 11, 12

- [37] Cano, Alberto e John D Leonard: *Interpretable multiview early warning system adapted to underrepresented student populations*. IEEE Transactions on Learning Technologies, 12(2):198–211, 2019. 12, 20
- [38] Darabi, Houshang, Elnaz Douzali, Samuel Thomas Harford, Peter C Nelson e Fazle Shahnawaz Muhibul Karim: *Beyond grade point average and standardized testing: Incorporating a socio-economic factor in admissions to support minority success*. Em *2017 ASEE Annual Conference & Exposition*, 2017. 12
- [39] Sadati, Seyedhamed e Nicolas Ali Libre: *Development of an early alert system to predict students at risk of failing based on their early course activities*. Em *2017 ASEE Annual Conference & Exposition*, 2017. 12
- [40] ElAtia, Samira, Donald Ipperciel e Osmar R Zaiane: *Data mining and learning analytics: Applications in educational research*. John Wiley & Sons, 2016. 12
- [41] Al-Sheeb, Bothaina A, AM Hamouda e Galal M Abdella: *Modeling of student academic achievement in engineering education using cognitive and non-cognitive factors*. Journal of Applied Research in Higher Education, 2019. 12, 19
- [42] Frischenschlager, Oskar, Gerald Haidinger e Lukas Mitterauer: *Factors associated with academic success at vienna medical school: prospective survey*. Croatian medical journal, 46(1), 2005. 14, 15
- [43] Ford, George D, Robert Elliot Steffen e Robert Wayne Ford: *University freshman retention in north carolina*. Em *2012 ASEE Annual Conference & Exposition*, páginas 25–1400, 2012. 15
- [44] Tekin, Ahmet: *Early prediction of students' grade point averages at graduation: A data mining approach*. Eurasian Journal of Educational Research, 54:207–226, 2014. 15, 16, 20
- [45] Reed, Teri Kristine, PK Imbrie, Qu Jin e Joe JJ Lin: *Modeling student success of international undergraduate engineers*. Em *2012 ASEE Annual Conference & Exposition*, páginas 25–945, 2012. 15, 18, 20
- [46] Tharp, Jack: *Predicting persistence of urban commuter campus students utilizing student background characteristics from enrollment data*. Community College Journal of Research and Practice, 22(3):279–294, 1998. 15, 17
- [47] Beck, Hall P e Meg Milligan: *Factors influencing the institutional commitment of online students*. The Internet and Higher Education, 20:51–56, 2014. 15
- [48] Campbell, Corbin M e Jessica L Mislevy: *Student perceptions matter: Early signs of undergraduate student retention/attrition*. Journal of College Student Retention: Research, Theory & Practice, 14(4):467–493, 2013. 16, 19
- [49] Eubanks, David, William Evers Jr e Nancy Smith: *Finding predictors in higher education*. Data Mining and Learning Analytics, página 41, 2016. 16

- [50] Baker, Ryan, Seiji Isotani e Adriana Carvalho: *Mineração de dados educacionais: Oportunidades para o Brasil*. Revista Brasileira de Informática na Educação, 19(02):03, 2011. 16
- [51] Dey, Eric L e Alexander W Astin: *Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression*. Research in higher education, 34(5):569–581, 1993. 17
- [52] Litzler, Elizabeth e Jacob Young: *Understanding the risk of attrition in undergraduate engineering: Results from the project to assess climate in engineering*. Journal of Engineering Education, 101(2):319–345, 2012. 19
- [53] Figueroa-Cañas, Josep e Teresa Sancho-Vinuesa: *Early prediction of dropout and final exam performance in an online statistics course*. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje, 15(2):86–94, 2020. 20
- [54] Gustian, Dudih e Ria Dewi Hundayani: *Combination of ahp method with c4. 5 in the level classification level out students*. Em *2017 International Conference on Computing, Engineering, and Design (ICCED)*, páginas 1–6. IEEE, 2017. 20
- [55] SILVA, LA da, S PERES e C Boscaroli: *Introdução à mineração de dados-com aplicação em r*, 2016. 21, 51, 54
- [56] Adejo, Olugbenga e Thomas Connolly: *An integrated system framework for predicting students' academic performance in higher educational institutions*. AIRCC's International Journal of Computer Science and Information Technology, 9(3):149–157, 2017. 26
- [57] UnB: *Universidade de Brasília*, 2022. <https://unb.br/a-unb/missao>, acesso em 2022-02-16. 29
- [58] UnB: *Anuário estatístico 2020*, 2022. <https://anuario-estatistico-unb-2020.netlify.app/>, acesso em 2022-02-16. 29
- [59] UnB: *Plano de desenvolvimento insitucional*, 2022. [http://dpo.unb.br/index.php?option=com\\_content&view=article&id=48&Itemid=874](http://dpo.unb.br/index.php?option=com_content&view=article&id=48&Itemid=874), acesso em 2022-02-16. 29
- [60] UnB: *Projeto político-pedagógico institucional*, 2022. [http://dpo.unb.br/index.php?option=com\\_content&view=article&id=72&Itemid=915](http://dpo.unb.br/index.php?option=com_content&view=article&id=72&Itemid=915), acesso em 2022-02-16. 30
- [61] STI: *Secretaria de tecnologia da informação*, 2022. <http://www.sti.unb.br/index.php/sti-institucional/historia-da-sti>, acesso em 2022-02-05. 30
- [62] UnB: *Sistema integrado de gestão das atividades acadêmicas*, 2022. <https://boasvindas.unb.br/matricula/sigaa>, acesso em 2022-02-05. 32
- [63] UnB: *Unb inicia operações com novo sistema de gestão*, 2022. <https://noticias.unb.br/76-institucional/2293-unb-inicia-operacoes-com-novo-sistema-de-gestao>, acesso em 2022-02-05. 32

- [64] UnB: *Decanato de ensino de graduação*, 2022. [http://deg.unb.br/index.php?option=com\\_content&view=article&id=372&Itemid=110](http://deg.unb.br/index.php?option=com_content&view=article&id=372&Itemid=110), acesso em 2022-02-06. 32
- [65] CEAD: *Centro de educação a distância (cead/unb)*, 2021. <https://www.cead.unb.br/apresentacao>, acesso em 2021-07-26. 33
- [66] CEAD-REMOTO: *Ensino remoto*, 2021. <https://cead.unb.br/ensino-remoto>, acesso em 2021-08-10. 33
- [67] GigaCandanga: *Gigacandanga*, 2022. <https://gigacandanga.net.br/institucional/>, acesso em 2022-02-09. 33
- [68] REUNI: *Programa de apoio a planos de reestruturação e expansão das universidades federais*, 2008. [http://portal.mec.gov.br/index.php?option=com\\_docman&view=download&alias=2069-reuni-relatorio-pdf&category\\_slug=dezembro-2009-pdf&Itemid=30192](http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=2069-reuni-relatorio-pdf&category_slug=dezembro-2009-pdf&Itemid=30192), acesso em 2022-02-16. 34
- [69] Norte, Universidade Federal do Rio Grande do: *ResoluÇÃO no 051-2020-consad, de 17 de dezembro de 2020*, 2022. [https://arquivos.info.ufrn.br/arquivos/202102203500c584721146aee3f4f7d8a/res0512020\\_disciplina\\_o\\_relacionamento\\_entre\\_a\\_UFRN\\_e\\_os\\_interessados\\_nos\\_sistem.pdf](https://arquivos.info.ufrn.br/arquivos/202102203500c584721146aee3f4f7d8a/res0512020_disciplina_o_relacionamento_entre_a_UFRN_e_os_interessados_nos_sistem.pdf), acesso em 2022-05-18. 40
- [70] RALPH KIMBALL, Joe Caserta: *The data warehouse etl toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data*. Canada: Wiley Publishing, Inc, 2004. 48
- [71] Microsoft: *Microsoft - power bi*, 2022. <https://powerbi.microsoft.com/pt-br/what-is-power-bi/>, acesso em 2022-05-25. 49
- [72] Chicco, Davide e Giuseppe Jurman: *The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation*. BMC genomics, 21(1):1–13, 2020. 51
- [73] Kitchenham, Barbara e Stuart Charters: *Guidelines for performing systematic literature reviews in software engineering*. 2007. 81, 85, 86, 87
- [74] Kitchenham, Barbara Ann, David Budgen e Pearl Brereton: *Evidence-based software engineering and systematic reviews*, volume 4. CRC press, 2015. 81, 82, 83, 84, 85, 87
- [75] Parsifal: *Online tool designed to support researchers to perform systematic literature reviews within the context of Software Engineering*, 2021. <https://parsif.al/>, acesso em 2021-10-20. 83, 84, 86, 87
- [76] Kitchenham, Barbara: *Procedures for performing systematic reviews*. Keele, UK, Keele University, 33(2004):1–26, 2004. 83, 85
- [77] Moodle: *Sobre o moodle*, 2022. [https://docs.moodle.org/all/pt\\_br/Sobre\\_o\\_Moodle](https://docs.moodle.org/all/pt_br/Sobre_o_Moodle), acesso em 2022-05-19. 94

[78] Microsoft: *Teams para o ensino*, 2022. <https://docs.microsoft.com/pt-br/microsoftteams/expand-teams-across-your-org/teams-for-education-landing-page>, acesso em 2022-05-19. 95

# Apêndice A

## Revisão Sistemática da Literatura

De acordo com Kitchenham e Charters, [73], a Revisão Sistemática da Literatura (RSL) é uma forma de identificar, analisar e interpretar evidências disponíveis relacionadas com uma determinada questão de pesquisa, área ou fenômeno de interesse.

Uma RSL envolve um conjunto de nove atividades, agrupadas em três fases: Planejamento da Revisão, Condução da Revisão e Documentação da Revisão.

**Fase 1 e o Planejamento da Revisão.** Aborda como o estudo deve ser feito e três atividades importantes são executadas: Especificação das questões de pesquisa; Desenvolvimento do Protocolo da revisão; e Validação do Protocolo Kitchenham et al. [74].

**Fase 2 e a Condução da Revisão.** As atividades desta fase são: Identificação de artigos relevantes; Seleção dos estudos primários; Avaliação da Qualidade; Extração dos Dados e a Síntese dos Dados. As definições descritas no protocolo são executadas e quaisquer divergências onde sejam necessárias alterações no protocolo devem ser documentadas Kitchenham et al. [74].

**Fase 3 é a Documentação da Revisão.** Nesta fase os processos e resultados são registrados Kitchenham et al. [74].

Na Figura A.1 apresentam-se as fases e os estágios da RSL.

### A.1 Planejamento da Revisão

Nesta Seção é apresentado a fase de planejamento da revisão sistemática. São delineadas atividades importantes como a definição das questões e pesquisa abordada pela revisão sistemática, a construção de um protocolo que guia a revisão e por fim, a atividade avaliação do protocolo Kitchenham e Charters [73].

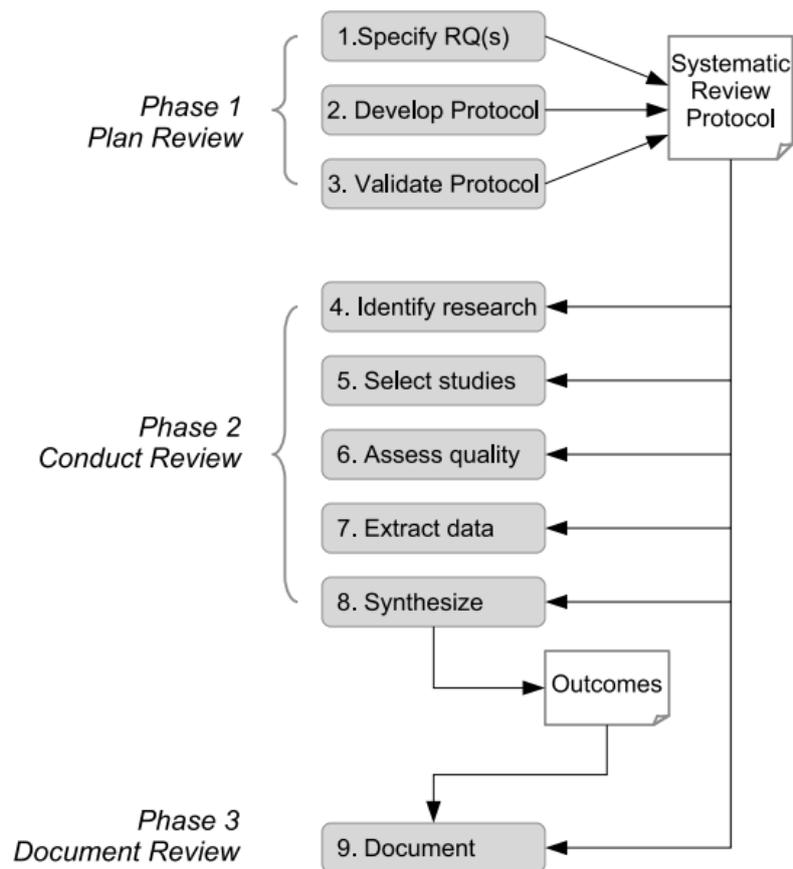


Figura A.1: Visão geral do processo de revisão sistemática  
 Fonte: Kitchenham et al. [74].

### A.1.1 Especificação das Questões de pesquisa para a RSL

A definição das questões de pesquisa é uma parte importante porque elas definem os motivos e impulsionam todo o processo de revisão Kitchenham et al. [74]. Para a RSL foram definidas três questões de pesquisa:

- QP1 — Quais fatores estão associados à previsão de retenção e desistência no ensino superior?
- QP2 — Quais são os fatores usados para prever a retenção do ensino superior?
- QP3 — Como os fatores são utilizados no processo de previsão?

### A.1.2 Desenvolvimento do Protocolo

O protocolo é um documento que detalha os procedimentos que serão executados na revisão, possibilitando a revisão por parte de outros pesquisadores com objetivo de dar *feedback* e mecanismos para resolver desacordos. Outra ajuda importante é a diminuição

de enviesamento por parte do pesquisador Kitchenham et al. [74]. As subseções a seguir descrevem os componentes usados no protocolo.

### **Estratégia de busca**

Na busca pelos artigos optou-se pela busca automática. Na realização da busca automática, Kitchenham et al. [74], recomenda do uso de ferramentas para facilitar o procedimento que envolve armazenamento sistemático, gerenciamento, validação e análise de uma considerável quantidade de dados propícios a erros. Desta forma, a ferramenta Parsifal [75] foi utilizada para dar suporte as fases planejamento da revisão e condução da revisão.

A *String* de busca usada nas bibliotecas digitais foi construída com base nas palavras-chave do objetivo da RSL, descrito na Seção 2.3, e considerou as questões de pesquisa descritas na Seção A.1.1. Depois desse procedimento, a estrutura da *String* de busca ficou definida desta forma:

( student OR undergraduate ) AND ( predict\* AND ( dropout OR retention OR attrition ) ) AND ( metric\* OR measurement OR indicat\* ) AND ( "higher education"OR "bachelor degree" ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) OR LIMIT-TO ( DOCTYPE , "cp" )

A *String* de busca está definida com a restrições por artigo e conferências e as bibliotecas digitas escolhidas para realizar a busca foram a SCOPUS e a IEEE.

### **Seleção dos Estudos**

Os critérios de seleção (inclusão ou exclusão) identificam os estudos primários que fornecem provas diretas sobre questão de pesquisa Kitchenham [76]. Estes critérios determinam se o estudo passará para as próximas fases da revisão sistemática. Os critérios de inclusão definidos foram: o artigo apresenta fatores relacionados a evasão; o artigo apresenta fatores relacionados com evasão e retenção; e os artigos apresentam fatores relacionados com a retenção. Para os critérios de exclusão foram definidos os seguintes tópicos: artigo de outra área de estudo; estudo duplicado; sem acesso ao texto completo; artigo não escrito em inglês, português ou espanhol; e estudos secundários;

### **Avaliando a qualidade dos estudos primários**

Nesta etapa uma lista de itens com critério de avaliação da qualidade é finidade com objetivo de garantir a confiabilidade da avaliação Kitchenham et al. [74]. Os itens da avaliação da qualidade ficaram assim definidos: os objetivos, questões de pesquisa e hipóteses (se aplicável) são claros e relevantes?; existe uma descrição adequada do contexto no qual a pesquisa foi conduzida?; os dados foram coletados de uma forma que abordasse

a questão da pesquisa?; existe uma declaração clara de resultados?; os autores descrevem limitações ao estudo?; as conclusões, implicações para a prática e pesquisas futuras, são adequadamente relatadas ao seu público?; as descobertas foram claramente relatadas?; e as questões éticas são devidamente tratadas (intenções pessoais, integridade, confidencialidade, consentimento, aprovação do conselho de revisão)?

Para fazer a análise da qualidade dos estudos, Kitchenham et al. [74] recomenda a contagem por votos, onde os estudos recebem uma pontuação para cada critério de qualidade que for atendido pelo artigo avaliado. Caso o artigo responda completamente à questão de qualidade em que está sendo avaliado recebe o valor de 1,0 pontos, caso responda só parcialmente recebe 0,5 pontos e caso não responda à questão é creditado nenhum ponto. A ferramenta Parsifal [75] dá suporte a esse método de pontuação como pode ser visto na Figura A.2.

| Answers                      |        |   |
|------------------------------|--------|---|
| Description                  | Weight |   |
| Yes                          | 1.0    |  edit  remove     |
| Partially                    | 0.5    |  edit  remove   |
| No                           | 0.0    |  edit  remove |
| <a href="#">+ Add Answer</a> |        |   |

| Quality Assessment Scores |                                  |  |
|---------------------------|----------------------------------|--|
| Max Score                 | <input type="text" value="8.0"/> | Calculated based on the number of questions and on the answer of greater weight          |
| Cutoff Score              | <input type="text" value="0.0"/> |  save |

Figura A.2: Definição da pontuação para as questões de qualidade dos estudos na ferramenta Parsifal

## Extração de Dados

Essa atividade define os procedimentos de extração dos dados que incluem detalhes da publicação e mais as informações necessárias para responder às questões de pesquisa da RSL, Kitchenham et al. [74].

Para executar a extração um formulário foi construído no Parsifal a extração das seguintes informações: Nome do pesquisador que realizou a extração; Total de citações; País em que ocorreu o estudo; Resumo do pesquisador; O tipo de faceta de pesquisa do artigo; A classificação do método de pesquisa usando; Especificação do curso de graduação; A descrição da disciplina; Informações relevantes para o pesquisador sobre a disciplina; O

indicador citado na pesquisa; Comentários adicionais sobre o indicador; Ferramentas de previsão citadas na pesquisa; Algoritmos de aprendizado de máquina utilizados; Descrever o algoritmo (subconjuntos, distribuição de dados, indicadores utilizados, etc.); Eficácia do algoritmo (cobertura, exatidão, precisão, *recall*, pontuação F1 e outros); e Outras informações relevantes para o pesquisador;

### **Estratégia de síntese e agregação de dados**

Nessa atividade do protocolo as estratégias para resumir, integrar, comparar e combinar os resultados encontrados nos trabalhos selecionados são definidas. Pela característica heterogenias das revisões sistemáticas em engenharia de software, uma abordagem qualitativa, como contagem de votos, pode ser usada Kitchenham et al. [74]. Para essa etapa, o Parsifal permite a extração das informações no formato de planilha que serão usadas para representar as informações no formato tabular e gráfico.

#### **A.1.3 Validar o Protocolo**

O protocolo tem um papel primordial na condução de uma revisão sistemática, os pesquisadores devem realizar acordos para avaliar o protocolo, Kitchenham, 2015 [73]. Esses acordos possibilitam gerar procedimentos para avaliação da *String* de busca e dos formulários de extração de dados, Kitchenham et al. [74]. A avaliação do protocolo foi realizada pelos professores orientadores do mestrado.

## **A.2 Condução da Revisão**

Esta fase é iniciada logo após os acordos de avaliação do protocolo, as próximas etapas são normalmente executadas sequencialmente, porém nada impede que algumas sejam realizadas simultaneamente, Kitchenham [76].

### **A.2.1 Identificação das pesquisas**

O objetivo desta etapa é a identificação das pesquisas relevantes. A definição de uma estratégia é fator importante para alcançar uma quantidade de estudos suficientes que respondam às questões de pesquisa, Kitchenham et al. [74].

A busca automática foi feita com a execução da *String* de busca nas bases de dados, como resposta foram obtidos 128 artigos, sendo 118 na SCOPUS e 10 na IEEE. A Figura A.3 mostra o percentual da distribuição dos artigos identificados por bases de pesquisa.

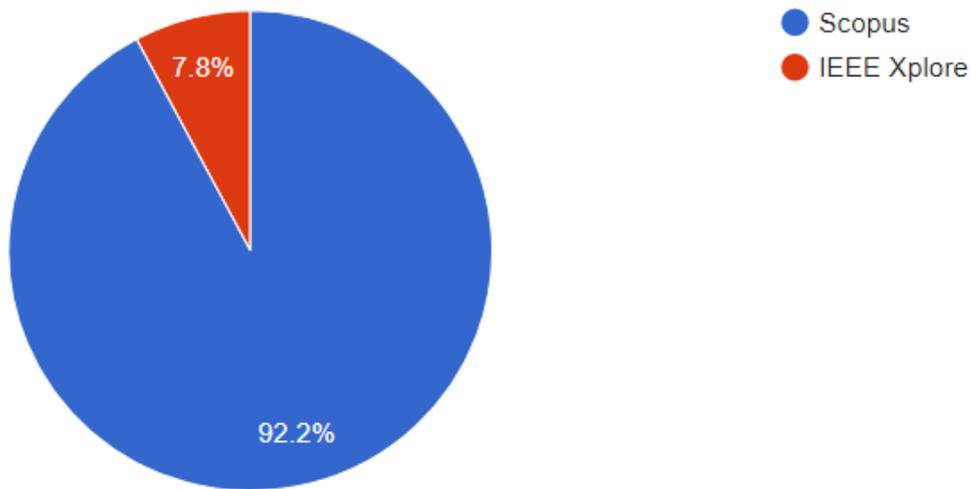


Figura A.3: Percentual da distribuição de artigos identificados por base de pesquisa.

## A.2.2 Seleção dos estudos

Nesta etapa, os estudos primários identificados passaram por processo de avaliação visando identificar a relevância desses estudos para a pesquisa [73].

A primeira atividade executada foi a identificação de estudos duplicados usando a funcionalidade nativa do Parsifal [75]. Isso reduziu de 129 para 120 artigos. A partir deste ponto os critérios de seleção (exclusão/inclusão) definidos na Seção A.1.2, que descreve o desenvolvimento do protocolo, foram utilizados.

Após este processo de análise, envolvendo o acesso ao texto completo, leitura dos títulos, resumos e palavras-chave, o total de estudos eleitos para a etapa de avaliação da qualidade foram de 58. A Figura A.4 mostra o número de artigos selecionados e aceitos por base.

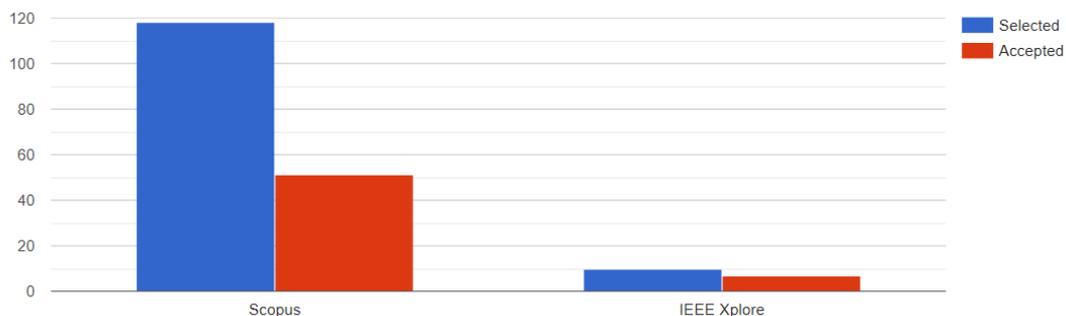


Figura A.4: Número de artigos selecionados e aceitos por base

### **A.2.3 Avaliar a qualidade**

A lista de avaliação da qualidade definida na Seção A.1.2, foi usada para atribuir a pontuação aos estudos. A avaliação da qualidade, segundo Kitchenham et al. [74], consiste em determinar se o estudo empírico é válido e isento de preconceitos. A autora também reforça que as notas de qualidade podem explicar diferenças nos resultados, interpretação dos resultados e a importância do estudo de uma forma geral para a RSL. Para a execução desta etapa a leitura integral do artigo foi feita.

### **A.2.4 Extração de dados**

O objetivo desta etapa é extrair as informações visando responder às questões de pesquisa da RSL. A estratégia de extração foi definida na etapa de desenvolvimento do protocolo executando a montagem o formulário de extração de dados no Parsifal [75]. Durante essa atividade o formulário de extração foi preenchido com as informações obtida pelo pesquisador. O procedimento de registro dessas informações foi realizado durante a etapa de avaliação da qualidade. Kitchenham et al. [74], descrevem que as duas etapas são distintas, podendo ser executadas em sequência ou em conjunto.

### **A.2.5 Sintetize dos dados**

Segundo Kitchenham e Charters [73], a compilação e o resumo dos resultados acontecem nesta etapa. Podem ser na forma de síntese descritiva ou quantitativa:

- Na síntese descritiva os dados devem ser tabulados de forma consistente com a pergunta da revisão e as tabelas dever dar prioridade as semelhanças e diferenças entre os resultados. Segundo os autores, é possível complementar a síntese descritiva com um resumo quantitativo.
- Nas análises quantitativas, que também devem ser a apresentadas em forma de tabelas, os resultados podem ser obtidos de maneiras diferentes se houverem diferentes tipos de medidas. Dente as formas diferentes de medida está probabilidade, risco, razão de probabilidade etc.

Nesse processo de análise dos artigos, a síntese descritiva foi utilizada para a apresentação dos achados nas pesquisas.

Foram avaliados 58 artigos, dispostos na Tabela A.4. O maior número de trabalhos é dos Estados Unidos, sendo seguido por países da Europa como Alemanha, Austrália e Holanda. A Figura A.5 mostra os países dos trabalhos selecionados na RSL.

Os fatores relacionados ao processo de evasão ou retenção foram identificados nos estudos e agrupados com base nas citações. Na Tabela A.1, são descritos os fatores e o

| Países       | Contagem de Países |
|--------------|--------------------|
| USA          | 19                 |
| Alemanha     | 4                  |
| Austrália    | 4                  |
| Holanda      | 4                  |
| Chile        | 3                  |
| Hungria      | 3                  |
| Espanha      | 2                  |
| Indonésia    | 2                  |
| Rússia       | 2                  |
| UK           | 2                  |
| Áustria      | 1                  |
| Brasil       | 1                  |
| Costa Rica   | 1                  |
| Cuba         | 1                  |
| El Savador   | 1                  |
| Finlândia    | 1                  |
| Grécia       | 1                  |
| India        | 1                  |
| Índia        | 1                  |
| Japão        | 1                  |
| Not reported | 1                  |
| Portugal     | 1                  |
| Qatar        | 1                  |
| Turquia      | 1                  |
| <b>Total</b> | <b>59</b>          |

Figura A.5: Total de artigos por país.

somatório dos fatores por artigo. Dentre os utilizados nas pesquisas, a pontuação média, gênero, notas do curso, curso e idade aparecem em maior número dos estudos.

Para realizar as análises nos fatores, os artigos mencionam ferramentas para dar suporte a esse procedimento. Essas ferramentas são algoritmos para construção dos modelos e softwares para dar suporte para a execução desses algoritmos.

A Tabela A.2, mostra um total de 13 algoritmos. A regressão foi o algoritmo mais citado pelos artigos, seguido pela árvore de decisão e redes neurais.

Na Tabela A.3, são disponibilizados os softwares identificados nos artigos. Um total de 8 softwares são usados para dar suporte a execução dos algoritmos e construção dos modelos. O IBM SPSS e o R foram os softwares mais citados nos artigos.

Tabela A.1: Fatores citados pelos artigos.

| <b>Fator</b>               | <b>Código do Artigo referente a Tabela A.4</b>   | <b>Total</b> |
|----------------------------|--|--------------|
| pontuação média            | A-2, A-9, A-14, A-16, A-18, A-19, A-20, A-22, A-23, A-24, A-25, A-27, A-29, A-30, A-31, A-33, A-34, A-36, A-37, A-38, A-39, A-41, A-43, A-44, A-45, A-51, A-55 | 27           |
| gênero                     | A-2, A-5, A-6, A-8, A-9, A-11, A-14, A-15, A-16, A-18, A-20, A-22, A-23, A-25, A-26, A-37, A-38, A-40, A-45, A-47, A-48, A-50, A-55                            | 23           |
| notas do curso             | A-2, A-5, A-11, A-12, A-16, A-17, A-18, A-19, A-22, A-23, A-33, A-34, A-37, A-39, A-45, A-51, A-57, A-58   | 18           |
| curso                      | A-1, A-9, A-18, A-22, A-25, A-27, A-30, A-31, A-33, A-37, A-40, A-44, A-45, A-46, A-48, A-50, A-51   | 17           |
| idade                      | A-2, A-8, A-9, A-11, A-14, A-15, A-17, A-19, A-20, A-22, A-23, A-25, A-48, A-50  | 14           |
| etnia                      | A-2, A-5, A-13, A-15, A-18, A-26, A-29, A-37, A-38, A-45, A-46, A-48, A-55   | 13           |
| bolsa de estudos           | A-1, A-2, A-6, A-9, A-13, A-17, A-22, A-26, A-29, A-39, A-47, A-48   | 12           |
| código postal              | A-2, A-6, A-9, A-14, A-16, A-19, A-26, A-37, A-47, A-54  | 10           |
| estudo regular             | A-5, A-9, A-10, A-13, A-23, A-25, A-44, A-58   | 8            |
| horário do curso           | A-4, A-13, A-17, A-36, A-38, A-44, A-48, A-51  | 8            |
| tipo de entrada            | A-16, A-17, A-19, A-22, A-23, A-50, A-52, A-54   | 8            |
| emprego                    | A-2, A-6, A-8, A-11, A-14, A-26, A-47  | 7            |
| matérias aprovadas         | A-4, A-13, A-18, A-22, A-27, A-35, A-49  | 7            |
| nível educacional dos pais | A-1, A-2, A-6, A-14, A-15, A-41  | 6            |
| acesso a fórum             | A-10, A-11, A-12, A-37, A-58   | 5            |
| acesso ao sistema          | A-12, A-36, A-37, A-56   | 5            |
| resolução de exercícios    | A-10, A-12, A-37, A-58   | 4            |
| acesso a vídeo             | A-10, A-21, A-58   | 3            |
| conhecimento prévio        | A-11, A-19, A-23   | 3            |
| estado civil               | A-9, A-11, A-48  | 3            |
| estrangeiro                | A-17, A-22, A-45   | 3            |
| matérias reprovadas        | A-13, A-17, A-19   | 3            |
| necessidades especiais     | A-36, A-41, A-52   | 3            |
| trancamento                | A-13, A-24, A-35   | 3            |
| curso matriculados         | A-12, A-13   | 2            |
| dados de pesquisa          | A-14, A-39   | 2            |
| filhos                     | A-8, A-11  | 2            |
| apresentação tardia        | A-10   | 1            |
| estágio                    | A-13   | 1            |

Tabela A.2: Algoritmos citados pelos artigos.

| <b>Algoritmos</b>               | <b>Código do Artigo referente a Tabela A.4</b>   | <b>Total</b> |
|---------------------------------|--|--------------|
| <i>Regression</i>               | A-1, A-3, A-4, A-5, A-6, A-8, A-10, A-11, A-13, A-14, A-15, A-16, A-18, A-20, A-23, A-24, A-25, A-26, A-30, A-32, A-33, A-37, A-38, A-43, A-44, A-45, A-46, A-48, A-51, A-52, A-54, A-55, A-57 | 33           |
| <i>Decision trees</i>           | A-1, A-7, A-8, A-9, A-11, A-12, A-17, A-22, A-23, A-38, A-41, A-49, A-53   | 13           |
| <i>Neural Networks</i>          | A-9, A-22, A-34, A-37, A-38, A-44, A-45, A-49, A-53  | 9            |
| <i>Random Forest</i>            | A-9, A-10, A-11, A-22, A-23, A-37, A-50  | 7            |
| <i>Naive Bayes</i>              | A-11, A-27, A-37, A-43, A-53   | 5            |
| <i>Support Vector Machines</i>  | A-10, A-23, A-34, A-49   | 4            |
| <i>Bayesian networks</i>        | A-11, A-19   | 2            |
| <i>C4.5</i>                     | A-7, A-12  | 2            |
| <i>K-NN</i>                     | A-11, A-53   | 2            |
| <i>Boosting and AdaBoost</i>    | A-49   | 1            |
| <i>Cluster analysis</i>         | A-6  | 1            |
| <i>Extreme Learning Machine</i> | A-34   | 1            |
| <i>Gradient Boosted Tree</i>    | A-22   | 1            |
| não reportado                   | A-2, A-21, A-28, A-29, A-31, A-35, A-36, A-39, A-40, A-42, A-47, A-56, A-58  | 13           |

Tabela A.3: Softwares citados pelos artigos.

| <b>Software</b>          | <b>Código do Artigo referente a Tabela A.4</b>  | <b>Total</b> |
|--------------------------|---|--------------|
| IBM SPSS                 | A-1, A-3, A-5, A-9, A-14, A-15, A-20, A-25, A-28, A-32, A-35, A-36, A-42, A-44, A-46, A-58  | 16           |
| R                        | A-8, A-12, A-20, A-23, A-53, A-54   | 6            |
| Weka                     | A-11, A-41  | 2            |
| JCLAL                    | A-11  | 1            |
| Python                   | A-4   | 1            |
| Python/Scikit            | A-39  | 1            |
| SAP Predictive Analytics | A-8   | 1            |
| SAS                      | A-38  | 1            |
| Não reportado            | A-2, A-6, A-7, A-10, A-13, A-16, A-17, A-18, A-19, A-21, A-22, A-24, A-26, A-27, A-29, A-30, A-31, A-33, A-34, A-37, A-40, A-43, A-45, A-47, A-48, A-49, A-50, A-51, A-52, A-55, A-56, A-57 | 32           |

Tabela A.4: Artigos selecionados na busca pelos fatores.

| <b>Código</b> | <b>Artigo</b>   |
|---------------|---|
| A-1           | A multinomial and predictive analysis of factors associated with university Dropout   |
| A-2           | A student retention model: Empirical, theoretical and pragmatic considerations  |
| A-3           | Analysis of students performance in relation to the results of state unified exam: The case of russian university   |
| A-4           | Analyzing the influence of online behaviors and learning approaches on academic performance in first year engineering                                     |
| A-5           | Beneath the Surface: An Investigation of General Chemistry Students' Study Skills to Predict Course Outcomes  |
| A-6           | Career consciousness and commitment to graduation among higher education students in Central and Eastern Europe   |
| A-7           | Combination of AHP Method with C4.5 in the level classification level out students  |
| A-8           | Comparative Analysis of Prediction Techniques to Determine Student Dropout: Logistic Regression vs Decision Trees   |
| A-9           | Drop-Out Prediction in Higher Education Among B40 Students  |
| A-10          | Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data  |
| A-11          | Early dropout prediction in distance higher education using active learning   |
| A-12          | Early Prediction of Dropout and Final Exam Performance in an Online Statistics Cou  |
| A-13          | Ensemble Regression Models Applied to Dropout in Higher Education   |
| A-14          | Factors associated with academic success at Vienna Medical School: Prospective survey   |
| A-15          | Factors influencing the institutional commitment of online students   |
| A-16          | Factors that affect student desertion in careers in Computer Engineering profile  |
| A-17          | Implementing a Machine Learning Approach to Predicting Students Academic Outcomes   |
| A-18          | Mathematics as a factor in community college STEM performance, persistence, and degree attainment   |
| A-19          | Modeling Students' Academic Performance Using Bayesian Networks   |
| A-20          | Motivation matters: predicting students' career decidedness and intention to drop out after the first year in higher education                            |
| A-21          | Perceived academic control and academic emotions predict undergraduate university student success: Examining effects on dropout intention and achievement |
| A-22          | Predicting Dropout Using High School and First-semester Academic Achievement Measures   |
| A-23          | Predicting Math Student Success in the Initial Phase of College With Sparse Information Using Approaches From Statistical Learning                        |

Tabela A.5: Artigos selecionados na busca pelos fatores - Continuação.

| <b>Código</b> | <b>Artigo</b>  |
|---------------|--|
| A-24          | Predicting performance in higher education using proximal predictors   |
| A-25          | Self-regulation of learning and MOOC retention   |
| A-26          | Student perceptions matter: Early signs of undergraduate student retention/attrition   |
| A-27          | The association of identity and motivation with students' academic achievement in higher education                                       |
| A-28          | The crucial first year: a longitudinal study of students' motivational development at a Swiss Business School                            |
| A-29          | Assessing the Validity of College Success Indicators for the At-Risk Student: Toward Developing a Best-Practice Model                    |
| A-30          | Bachelor completion and dropout rates of selected, rejected and lottery-admitted medical students in the Netherlands                     |
| A-31          | Beyond grade point average and standardized testing: Incorporating a socioeconomic factor in admissions to support minority success      |
| A-32          | Considering the role of the distance student experience in student satisfaction and retention  |
| A-33          | Development of an early alert system to predict students at risk of failing based on their early course activities                       |
| A-34          | Early prediction of students' grade point averages at graduation: A data mining approach   |
| A-35          | Early predictors of study success in a Dutch advanced nurse practitioner education program: A retrospective cohort study                 |
| A-36          | Effects of emotional intelligence and supportive text messages on academic outcomes in first-year undergraduates                         |
| A-37          | Engagement vs performance: Using electronic portfolios to predict first semester engineering student retention                           |
| A-38          | Exploring student characteristics of retention that lead to graduation in higher education using data mining models                      |
| A-39          | Finding predictors in higher education   |
| A-40          | How to retain students in higher engineering education? Findings of the ATTRACT project  |
| A-41          | Interpretable Multiview Early Warning System Adapted to Underrepresented Student Populations   |
| A-42          | Learning analytics in higher education: Assessing learning outcomes  |
| A-43          | Measuring, Manipulating, and Predicting Student Success: A 10-Year Assessment of Carnegie R1 Doctoral Universities Between 2004 and 2013 |
| A-44          | Modeling of student academic achievement in engineering education using cognitive and non-cognitive factors                              |
| A-45          | Modeling student success of international undergraduate engineers  |
| A-46          | Persistence and engagement among first-year Hispanic students  |
| A-47          | Predicting college success with high school grades and test scores: Limitations for minority students                                    |

Tabela A.6: Artigos selecionados na busca pelos fatores - Continuação.

| <b>Código</b> | <b>Artigo</b>   |
|---------------|---|
| A-48          | Predicting persistence of urban commuter campus students utilizing student background characteristics from enrollment data      |
| A-49          | Predicting student academic performance using multi-model heterogeneous ensemble approach                                       |
| A-50          | Predicting student degree completion using random forest  |
| A-51          | Predicting success in an undergraduate exercise science program using science-based admission courses                           |
| A-52          | Prediction model of first-year student desertion at Universidad Bernardo O'Higgins (UBO)  |
| A-53          | Predictive modelling of student dropout using ensemble classifier method in higher education                                    |
| A-54          | Predictors of performance in business administration degrees: The effect of the high-school specialty                           |
| A-55          | Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression |
| A-56          | Technological barriers and incentives to learning analytics adoption in higher education: insights from users                   |
| A-57          | University freshman retention in North Carolina   |
| A-58          | Validating the effectiveness of the moodle engagement analytics plugin to predict student academic performance                  |

# Apêndice B

## Ambientes Virtuais de Aprendizagem

### B.1 Aprender UnB — Moodle

O Aprender é uma plataforma baseada no Moodle, um sistema de código aberto desenvolvido para fornecer a educadores, administradores e alunos um sistema robusto, seguro e integrado para criar ambientes de aprendizagem personalizados [77].

Para possibilitar a análise no modelo de dados do Moodle, uma instância do ambiente foi montada em ambiente local.

A versão 3.11 foi instalada e está disponível para download na página oficial do Moodle [77]. Em seu formato padrão, o Moodle é composto por 444 tabelas, das quais apenas 9 foram selecionadas para atender aos fatores de aprendizagem identificados na revisão sistemática. A Figura B.1 mostra as tabelas selecionadas que atendem aos seguintes fatores: curso matriculado, estudos regulares, acesso a fórum, acesso ao sistema, resolução de exercícios, acesso a vídeo e submissão tardia de tarefas.

Após a instalação da aplicação, foi possível criar disciplinas e atividades fictícias, gerando, dessa forma, uma massa de dados que permitiu identificar as tabelas e, consequentemente, gerar *scripts* em linguagem SQL para analisar e construir a estrutura de dados composta pelos fatores de aprendizagem selecionados na revisão sistemática. A Figura B.2 exemplifica um *script* SQL e a consulta gerada para identificação das datas de acesso ao sistema e páginas acessadas por um determinado usuário.

### B.2 Microsoft Teams

O Microsoft Teams para Educação é um software permite estudantes, professores e funcionários se encontrem, trabalhando juntos, criando conteúdo e compartilhando recursos

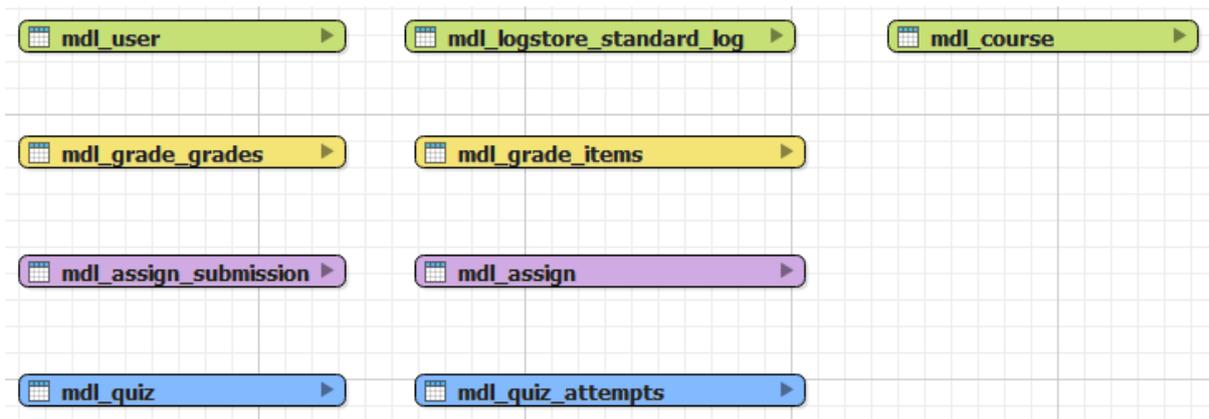


Figura B.1: Tabelas selecionadas do modelo do Moodle.

```

select
  mu.username as login,
  FROM_UNIXTIME(mu.firstaccess) as data_entrada_sistema,
  FROM_UNIXTIME(mu.lastlogin) as ultimo_login,
  FROM_UNIXTIME(mu.currentlogin) as login_corrente,
  FROM_UNIXTIME(msl.timecreated) as data_ultimo_evento,
  mc.shortname, mc.fullname,
  msl.action as acao,
  msl.target as alvo,
  msl.component,
  msl.crud
from mdl_logstore_standard_log msl
  inner join mdl_user mu on mu.id = msl.userid
  left join mdl_course mc on mc.id = msl.courseid
where msl.userid = 4

```

| login | data_entrada_sistema    | ultimo_login            | login_corrente          | data_ultimo_evento      | mc_shortname   | mc_fullname                   | acao   | alvo         | comp           | crud |
|-------|-------------------------|-------------------------|-------------------------|-------------------------|----------------|-------------------------------|--------|--------------|----------------|------|
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:10:48.000 | [NULL]         | [NULL]                        | viewed | user_profile | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:11:14.000 | [NULL]         | [NULL]                        | viewed | dashboard    | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:11:22.000 | [NULL]         | [NULL]                        | viewed | dashboard    | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:31:20.000 | [NULL]         | [NULL]                        | viewed | dashboard    | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-22 09:42:21.000 | [NULL]         | [NULL]                        | viewed | dashboard    | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:11:22.000 | Ambiente Virtu | AVA                           | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-22 20:01:06.000 | FGA0273        | MELHORIA DE PROCESSOS DE SOFT | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-22 20:02:18.000 | FGA0273        | MELHORIA DE PROCESSOS DE SOFT | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-22 20:03:55.000 | FGA0273        | MELHORIA DE PROCESSOS DE SOFT | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-23 07:52:46.000 | FGA0273        | MELHORIA DE PROCESSOS DE SOFT | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-23 10:12:32.000 | FGA0273        | MELHORIA DE PROCESSOS DE SOFT | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-22 20:07:26.000 | FGA0273        | MELHORIA DE PROCESSOS DE SOFT | graded | user         | core           | u    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:11:30.000 | REst           | Estatística com R             | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:34:36.000 | REst           | Estatística com R             | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:35:15.000 | REst           | Estatística com R             | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-22 09:42:25.000 | REst           | Estatística com R             | viewed | course       | core           | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-22 09:43:01.000 | REst           | Estatística com R             | viewed | grade_report | gradereport_u  | r    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 08:43:09.000 | FGA0278        | QUALIDADE DE SOFTWARE 1       | ended  | tour         | tool_usertours | c    |
| cosme | 2022-04-21 08:30:14.000 | 2022-04-22 09:42:21.000 | 2022-04-23 07:52:37.000 | 2022-04-21 09:10:39.000 | FGA0278        | QUALIDADE DE SOFTWARE 1       | viewed | user_list    | core           | r    |

Figura B.2: Exemplo da consulta no Moodle.

no Office 365 para Educação [78].

O foco principal durante a análise no Microsoft Teams foi obter os dados que alimentam a funcionalidade chamada de *Insights*. Essa funcionalidade promove relatórios sobre as atividades dos alunos em uma determinada turma registrada pelo professor. A Figura B.3 mostra a página *Insights* com a visualização das funcionalidades.

A análise dessa funcionalidade permitiu identificar que os dados fornecidos pelo Microsoft Teams podem prover informações que, individual ou combinadamente, podem atender aos fatores identificados na revisão da literatura. Um exemplo disso é a funci-

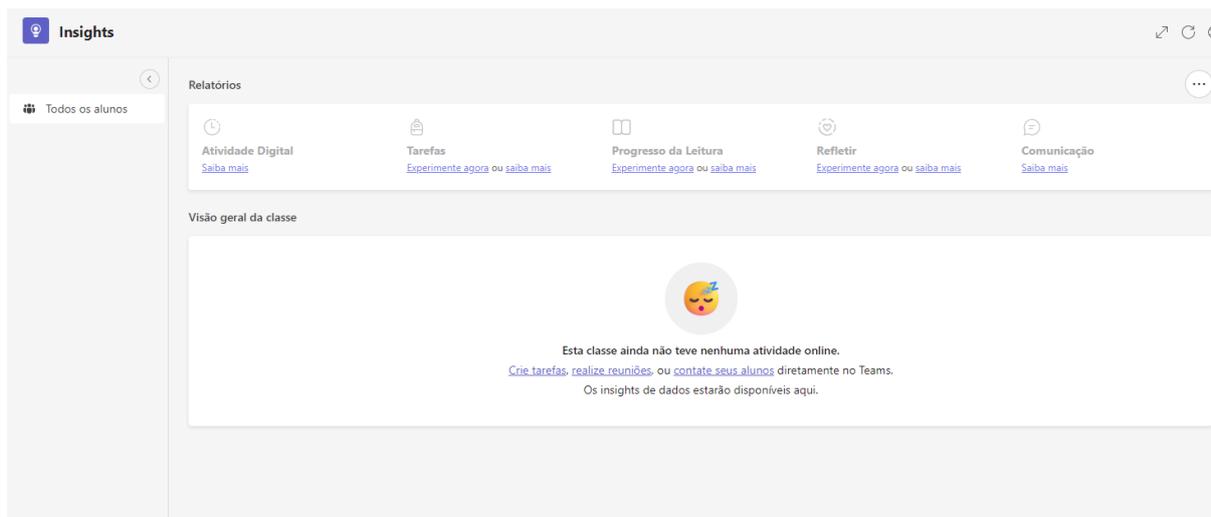


Figura B.3: Tela da funcionalidade Insights do Teams.

onalidade Tarefas, onde é possível definir uma pontuação para a tarefa e saber se foi realizada ou não pelo aluno. Os fatores atendidos pelo Teams também estão apresentados na Tabela 4.1.

O acesso às informações da página de *Insights* é feito por meio de relatórios em formato de planilhas do Excel, não sendo possível acessar o bando de dados. Dessa forma, não foi possível fazer o mapeamento dos fatores com as tabelas e atributos do banco de dados desse sistema.

## B.3 Disponibilidade dos Dados

Em contato com o Centro de Educação a Distância (CEAD), unidade responsável pela condução das políticas para gestão dos cursos de graduação a distância, foi informado que os dados poderiam ser disponibilizados, mas não em um formato de relatório baseado na consulta construída com comandos em linguagem SQL. Os dados seriam disponibilizados no formato de uma cópia integral do banco de dados.

Esse procedimento tornou a inclusão dessa base de dados complexa, tendo em vista que seria de nossa responsabilidade a guarda e a anonimização de toda e qualquer informação dos alunos contida no banco de dados. Desta forma, foi tomada a decisão de não incluir neste trabalho os fatores contemplados por essa base de dados.

Os dados do sistema Microsoft Teams não puderam ser usados devido ao administrador do Microsoft Teams na STI não ter acesso, de forma geral, aos dados dos *Insights*. Essas informações só podem ser obtidas pelo professor, o qual é o responsável por criar a

turma. Assim, não sendo possível acessar essas informações pelo perfil de administrador do Microsoft Teams da UnB, o sistema em questão não foi incluído neste estudo.

# Apêndice C

## Fatores e Atributos dos Bancos de Dados

O propósito desta seção foi relacionar os fatores identificados na RSL e os atributos das tabelas dos sistemas acadêmicos mapeados neste trabalho.

Essa análise obteve como produto a Tabela C.1, onde mostram os fatores, sistemas, tabelas e atributos mapeados. Os fatores que não estão relacionados com os atributos da tabela não foram encontrados no sistema e foram marcados com a letra “x” e não foram considerados neste trabalho.

Tabela C.1: Mapeamento de Fatores, Tabelas e Atributos

| <b>FACTOR</b>   | <b>SIGAA</b>                                    | <b>SIGRA</b>                              | <b>MOODLE</b>                       |
|-----------------|---|---|-------------------------------------|
| pontuação média | ensino-matricula-componente- <b>media-final</b> | bdsiac-historicoescolar- <b>HesMencao</b> | mdl-grade-grades- <b>finalgrade</b> |
| curso           | public-discente- <b>id-curso</b>                | bdsiac-curso- <b>curcurso</b>             | mdl-course                          |
| gênero          | comum.pessoa- <b>sexo</b>                       | BDSia.Aluno- <b>alusexo</b>               |                                     |
| notas do curso  | ensino-matricula-componente- <b>media-final</b> | bdsiac-historicoescolar- <b>HesMencao</b> |                                     |
| idade           | comum-pessoa- <b>data-nascimento</b>            | BDSiac-dbo-Aluno- <b>aludtnasc</b>        |                                     |
| etnia           | <b>X</b>  | <b>X</b>                                  |                                     |

Continua...

| <b>FACTOR</b>                   | <b>SIGAA</b>   | <b>SIGRA</b>  | <b>MOODLE</b>                 |
|---------------------------------|--|---|-------------------------------|
| bolsa de estudos                | administrativo-<br>bolsista- <b>id-pessoa</b>                      | <b>X</b>  |                               |
| código postal                   | comum-pessoa- <b>cep</b>   | BDSiac-dbo-Aluno-<br><b>aluendcep</b>                   |                               |
| estudo regular                  |  |   | mdl-logstore-<br>standard-log |
| horário do curso                | ensino-turno- <b>id-<br/>turno</b>                                 | BDSiac-curso-<br><b>CurTurno</b>                        |                               |
| tipo de entrada                 | ensino-forma-<br>ingresso- <b>id-forma-<br/>ingresso</b>           | BDTabelaApoio-<br>TB-FormaIngresso-<br><b>FInCodigo</b> |                               |
| emprego                         | <b>X</b>   | <b>X</b>  |                               |
| matérias aprova-<br>das         | ensino.matricula-<br>componente- <b>id-<br/>situacao-matricula</b> | bdsiac-<br>historicoescolar-<br><b>HesMencao</b>        |                               |
| nível educacio-<br>nal dos pais | <b>X</b>   | <b>X</b>  |                               |
| acesso a fórum                  |  |   | mdl-logstore-<br>standard-log |
| acesso ao sis-<br>tema          |  |   | mdl-logstore-<br>standard-log |
| resolução de<br>exercícios      |  |   | mdl-logstore-<br>standard-log |
| acesso a vídeo                  |  |   | mdl-logstore-<br>standard-log |
| conhecimento<br>prévio          | <b>X</b>   | <b>X</b>  | <b>X</b>                      |
| estado civil                    | comum.pessoa. <b>id-<br/>estado-civil</b>                          | TB-EstadoCivil-<br><b>ECiCodigo</b>                     |                               |
| estrangeiro                     | comum.pessoa. <b>id-<br/>pais-nacionalidade</b>                    | Tabela-<br>DadosPessoais-<br><b>Nacionalidade</b>       |                               |
| matérias repro-<br>vadas        | ensino-matricula-<br>componente- <b>id-<br/>situacao-matricula</b> | bdsiac-<br>historicoescolar-<br><b>HesMencao</b>        |                               |

Continua...

| <b>FACTOR</b>          | <b>SIGAA</b>  | <b>SIGRA</b>                                  | <b>MOODLE</b>         |
|------------------------|---|---|-----------------------|
| necessidades especiais | comum-pessoa- <b>id-tipo-necessidade-especial</b>         | BDSiac-dbo-Aluno- <b>alupne</b>               |                       |
| trancamento disciplina | ensino-matricula-componente- <b>id-situacao-matricula</b> | bdsiac-historicoescolar- <b>HesMencao</b>     |                       |
| cursos matriculados    | public- <b>discente-id-curso</b>                          | bdsiac-dbo-dadosopcao- <b>dadalumatricula</b> |                       |
| dados de pesquisa      | <b>X</b>  | <b>X</b>                                      |                       |
| filhos                 | <b>X</b>  | <b>X</b>                                      |                       |
| apresentação tardia    |   |   | mdl-assign-submission |
| estágio                | estagio-estagio- <b>id-discente</b>                       | <b>X</b>                                      |                       |

# Apêndice D

## Consulta SQL e Tipos de Dados

## Consulta executada na base de dados do SIGAA

```
select
--ALUNO--
public.discente.matricula,
comum.pessoa.id_pessoa,
comum.pessoa.codmergpe as origem_informacao , comum.pessoa.sexo,
comum.pessoa.data_nascimento,
comum.pessoa.id_tipo_necessidade_especial,
comum.pessoa.cep, comum.pais.nome as nacionalidade,
comum.pessoa.id_raca, comum.tipo_raca .descricao as raca_descricao,
comum.estado_civil.descricao as estado_civil,
public.status_discente.descricao,
public.discente.ano_ingresso, public.discente.periodo_ingresso,
ensino.forma_ingresso.descricao as forma_ingresso,
public.discente.data_colacao_grau, public.discente.prazo_conclusao,
-- CURSO --
public.curso.id_curso, public.curso.nome as nome_curso, ensino.turno.periodo
as periodo_curso, public.curso.codmergpapos as curso_codmergpapos,
public.curso.codmergpa as curso_codmergpa, comum.campus_ies.sigla as
sigla_campus , comum.campus_ies.nome as nome_campus,
-- MATRIZ CURRICULAR --
graduacao.curriculo.id_curriculo as opcao, graduacao.curriculo.id_matriz,
graduacao.curriculo.codigo as codigo_curriculo,
graduacao.curriculo.anoentradavigor||' '||graduacao.curriculo.periodoentradavi
gor as ano_curriculo,
cast(extract (year from
graduacao.matriz_curricular.data_inicio_funcionamento)as varchar)||
case
when date_part('month',graduacao.matriz_curricular.data_inicio_funcionamento)
< 6 then 1
when date_part('month',graduacao.matriz_curricular.data_inicio_funcionamento)
> 6 then 2
end as ano_curso,
-- DISCIPLINA --
ensino.componente_curricular.codigo as codigo_comp_curricular,
ensino.componente_curricular.codigo_antigo as codigo_comp_curricular_antigo,
ensino.componente_curricular.id_disciplina,
-- DESCRICAO COMPONENTE CURRICULAR --
ensino.componente_curricular_detalhes.nome as nome_comp_curricular,
ensino.componente_curricular.codmergpa , ensino.componente_curricular.codmerg
as cod_dis_sigra,
ensino.componente_curricular_detalhes.equivalencia as
equivalencia_disciplina,
comum.unidade.sigla as sigla_dep, comum.unidade.nome as nome_dep,
ensino.tipo_componente_curricular.descricao as descricao_tipo_disciplina,
ensino.situacao_matricula.descricao as sit_mat_descricao,
ensino.matricula_componente.tipo_integralizacao,
ensino.matricula_componente.ano as ano_mat_mc,
ensino.matricula_componente.periodo as periodo_mat_mc,
ensino.matricula_componente.mes as mes_mat_mc,
ensino.matricula_componente.mes_fim as mes_fim_mat_mc,
ensino.componente_curricular.ativo as comp_curricular_ativo,
ensino.matricula_componente.numero_faltas as numero_faltas_mc,
ensino.matricula_componente.porcentagem_frequencia as
porcentagem_frequencia_mc,
ensino.matricula_componente.media_final as media_final_mc,
ensino.conceito_valor.conceito,
```

```

-- TURMA --
ensino.turma.id_turma, ensino.turma.codigo as codigo_turma, ensino.turma.ano
as ano_turma, ensino.turma.periodo as periodo_turma,
ensino.turma.descricao_horario as descricao_horario_turma
from public.discente
inner join comum.pessoa on comum.pessoa.id_pessoa = public.discente.id_pessoa
left join comum.estado_civil on comum.estado_civil.id_estado_civil =
comum.pessoa.id_estado_civil
left join comum.pais on comum.pais .id_pais =
comum.pessoa.id_pais_nacionalidade
left join comum.tipo_raca on comum.tipo_raca.id_tipo_raca =
comum.pessoa.id_raca
left join comum.tipo_etnia te on te.id_tipo_etnia =
comum.pessoa.id_tipo_etnia
inner join ensino.forma_ingresso on ensino.forma_ingresso.id_forma_ingresso =
public.discente.id_forma_ingresso
inner join public.status_discente on public.status_discente.status =
public.discente.status
inner join ensino.matricula_componente on
ensino.matricula_componente.id_discente = public.discente.id_discente
inner join ensino.situacao_matricula on
ensino.situacao_matricula.id_situacao_matricula =
ensino.matricula_componente.id_situacao_matricula
inner join ensino.componente_curricular_detalhes on
ensino.componente_curricular_detalhes.id_componente_detalhes =
ensino.matricula_componente.id_componente_detalhes
inner join ensino.componente_curricular on
(ensino.componente_curricular.id_detalhe =
ensino.componente_curricular_detalhes.id_componente_detalhes and
ensino.componente_curricular.id_disciplina =
ensino.matricula_componente.id_componente_curricular)
inner join comum.unidade on comum.unidade.id_unidade =
ensino.componente_curricular.id_unidade
inner join graduacao.curriculo on graduacao.curriculo.id_curriculo =
public.discente.id_curriculo
inner join public.curso on public.curso.id_curso =
graduacao.curriculo.id_curso
inner join comum.campus_ies on comum.campus_ies.id_campus =
public.curso.id_campus
left join ensino.turno on ensino.turno.id_turno = public.curso.id_turno
left join ensino.turma on ensino.turma.id_turma =
ensino.matricula_componente.id_turma
inner join graduacao.matriz_curricular on
graduacao.matriz_curricular.id_matriz_curricular =
graduacao.curriculo.id_matriz
left join ensino.conceito_valor on
(ensino.conceito_valor.id_configuracao_consolidacao =
ensino.turma.id_cofiguracao_consolidacao and
ensino.matricula_componente.media_final = ensino.conceito_valor.valor)
inner join ensino.tipo_componente_curricular on
ensino.tipo_componente_curricular.id_tipo_disciplina =
ensino.componente_curricular_detalhes.id_tipo_componente

```

Descrição de tipo e tamanho dos campos selecionados no SIGAA.

| Nome do Campo                 | Tipo    | Tamanho |
|-------------------------------|---------|---------|
| matricula                     | Integer | 15      |
| id_pessoa                     | Integer | 9       |
| sexo                          | String  | 1       |
| data_nascimento               | Date    |         |
| id_tipo_necessidade_especial  | Integer | 9       |
| cep                           | String  | 10      |
| nacionalidade                 | String  | 80      |
| id_raca                       | Integer | 9       |
| raca_descricao                | String  | 80      |
| estado_civil                  | String  | 80      |
| status_discente               | String  | 60      |
| ano_ingresso                  | Integer | 9       |
| periodo_ingresso              | Integer | 1       |
| forma_ingresso                | String  | 80      |
| data_colacao_grau             | Date    |         |
| prazo_conclusao               | String  | 9       |
| id_curso                      | Integer | 9       |
| nome_curso                    | String  | 200     |
| opcao                         | String  | 60      |
| ano_curso                     | String  | 10      |
| ano_curriculo                 | String  | 10      |
| curso_codmergpa               | String  | 60      |
| id_matriz                     | Integer | 9       |
| codigo_curriculo              | String  | 60      |
| curso_codmergpapos            | String  | 60      |
| sigla_campus                  | String  | 20      |
| nome_campus                   | String  | 120     |
| periodo_curso                 | String  | 2       |
| codmergpa                     | String  | 60      |
| codigo_comp_curricular        | String  | 20      |
| codigo_comp_curricular_antigo | String  | 30      |
| id_disciplina                 | Integer | 9       |
| nome_comp_curricular          | String  | 150     |
| sigla_dep                     | String  | 20      |
| nome_dep                      | String  | 200     |
| descricao_tipo_disciplina     | String  | 255     |
| sit_mat_descricao             | String  | 80      |
| tipo_integralizacao           | String  | 3       |
| ano_mat_mc                    | Integer | 4       |
| periodo_mat_mc                | Integer | 4       |
| mes_mat_mc                    | Integer | 9       |

|                           |         |      |
|---------------------------|---------|------|
| mes_fim_mat_mc            | Integer | 9    |
| id_turma                  | Integer | 9    |
| codigo_turma              | String  | 10   |
| conceito                  | String  | 3    |
| numero_faltas_mc          | String  | 9    |
| porcentagem_frequencia_mc | String  | 9    |
| media_final_mc            | String  | 9    |
| ano_turma                 | Integer | 9    |
| periodo_turma             | Integer | 9    |
| descricao_horario_turma   | String  | 600  |
| comp_curricular_ativo     | Integer |      |
| origem_informacao         | String  | 60   |
| cod_dis_sigra             | Integer | 10   |
| equivalencia_disciplina   | String  | 5000 |

**Descrição de tipo dos atributos selecionados para as previsões.**

| <b>Nome do Campo</b>                          | <b>Tipo</b> | <b>Valor alterado para uso nos algoritmos</b>  |
|---|-------------|--|
| codigo_comp_curricular                        | Texto       | Usado conforme veio do SIGAA   |
| sexo  | Inteiro     | Feminino - 1 e Masculino -2  |
| idade (calculado a partir da data_nascimento) | Inteiro     |  |
| forma_ingresso                                | Inteiro     | Vestibular -1, ENEM - 2, Portador de Diploma de Curso Superior - 3, Mudança de Curso - 4, Transferência Obrigatória - 5, Transferência Facultativa - 6 e Outros - 8. |
| periodo_curso                                 | Character   | Diurno -1 e Noturno -2   |

# Anexo I

## Legislação



**Presidência da República**  
**Casa Civil**  
**Subchefia para Assuntos Jurídicos**

**DECRETO Nº 7.234, DE 19 DE JULHO DE 2010.**

Dispõe sobre o Programa Nacional de Assistência Estudantil - PNAES.

**O PRESIDENTE DA REPÚBLICA**, no uso da atribuição que lhe confere o art. 84, inciso VI, alínea "a", da Constituição:

**DECRETA:**

Art. 1º O Programa Nacional de Assistência Estudantil – PNAES, executado no âmbito do Ministério da Educação, tem como finalidade ampliar as condições de permanência dos jovens na educação superior pública federal.

Art. 2º São objetivos do PNAES:

- I – democratizar as condições de permanência dos jovens na educação superior pública federal;
- II - minimizar os efeitos das desigualdades sociais e regionais na permanência e conclusão da educação superior;
- III - reduzir as taxas de retenção e evasão; e
- IV - contribuir para a promoção da inclusão social pela educação.

Art. 3º O PNAES deverá ser implementado de forma articulada com as atividades de ensino, pesquisa e extensão, visando o atendimento de estudantes regularmente matriculados em cursos de graduação presencial das instituições federais de ensino superior.

§ 1º As ações de assistência estudantil do PNAES deverão ser desenvolvidas nas seguintes áreas:

- I - moradia estudantil;
- II - alimentação;
- III - transporte;
- IV - atenção à saúde;
- V - inclusão digital;
- VI - cultura;
- VII - esporte;
- VIII - creche;
- IX - apoio pedagógico; e
- X - acesso, participação e aprendizagem de estudantes com deficiência, transtornos globais do desenvolvimento e altas habilidades e superdotação.

§ 2º Caberá à instituição federal de ensino superior definir os critérios e a metodologia de seleção dos alunos de graduação a serem beneficiados.

Art. 4º As ações de assistência estudantil serão executadas por instituições federais de ensino superior, abrangendo os Institutos Federais de Educação, Ciência e Tecnologia, considerando suas especificidades, as áreas estratégicas de ensino, pesquisa e extensão e aquelas que atendam às necessidades identificadas por seu corpo discente.

Parágrafo único. As ações de assistência estudantil devem considerar a necessidade de viabilizar a igualdade de oportunidades, contribuir para a melhoria do desempenho acadêmico e agir, preventivamente, nas situações de retenção e evasão decorrentes da insuficiência de condições financeiras.

Art. 5º Serão atendidos no âmbito do PNAES prioritariamente estudantes oriundos da rede pública de educação básica ou com renda familiar **per capita** de até um salário mínimo e meio, sem prejuízo de demais requisitos fixados pelas instituições federais de ensino superior.

Parágrafo único. Além dos requisitos previstos no **caput**, as instituições federais de ensino superior deverão fixar:

I - requisitos para a percepção de assistência estudantil, observado o disposto no **caput** do art. 2º; e

II - mecanismos de acompanhamento e avaliação do PNAES.

Art. 6º As instituições federais de ensino superior prestarão todas as informações referentes à implementação do PNAES solicitadas pelo Ministério da Educação.

Art. 7º Os recursos para o PNAES serão repassados às instituições federais de ensino superior, que deverão implementar as ações de assistência estudantil, na forma dos arts. 3º e 4º.

Art. 8º As despesas do PNAES correrão à conta das dotações orçamentárias anualmente consignadas ao Ministério da Educação ou às instituições federais de ensino superior, devendo o Poder Executivo compatibilizar a quantidade de beneficiários com as dotações orçamentárias existentes, observados os limites estipulados na forma da legislação orçamentária e financeira vigente.

Art. 9º Este Decreto entra em vigor na data de sua publicação.

Brasília, 19 de julho de 2010; 189º da Independência e 122º da República.

LUIZ INÁCIO LULA DA SILVA  
*Fernando Haddad*

Este texto não substitui o publicado no DOU de 20.7.2010



**Presidência da República**  
**Casa Civil**  
**Subchefia para Assuntos Jurídicos**

**DECRETO Nº 6.096, DE 24 DE ABRIL DE 2007.**

**Institui o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais - REUNI.**

**O PRESIDENTE DA REPÚBLICA**, no uso da atribuição que lhe confere o art. 84, inciso VI, alínea "a", da Constituição, e considerando a meta de expansão da oferta de educação superior constante do item 4.3.1 do Plano Nacional de Educação, instituído pela Lei nº 10.172, de 9 de janeiro de 2001,

**DECRETA:**

Art. 1º Fica instituído o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais - REUNI, com o objetivo de criar condições para a ampliação do acesso e permanência na educação superior, no nível de graduação, pelo melhor aproveitamento da estrutura física e de recursos humanos existentes nas universidades federais.

§ 1º O Programa tem como meta global a elevação gradual da taxa de conclusão média dos cursos de graduação presenciais para noventa por cento e da relação de alunos de graduação em cursos presenciais por professor para dezoito, ao final de cinco anos, a contar do início de cada plano.

§ 2º O Ministério da Educação estabelecerá os parâmetros de cálculo dos indicadores que compõem a meta referida no § 1º.

Art. 2º O Programa terá as seguintes diretrizes:

I - redução das taxas de evasão, ocupação de vagas ociosas e aumento de vagas de ingresso, especialmente no período noturno;

II - ampliação da mobilidade estudantil, com a implantação de regimes curriculares e sistemas de títulos que possibilitem a construção de itinerários formativos, mediante o aproveitamento de créditos e a circulação de estudantes entre instituições, cursos e programas de educação superior;

III - revisão da estrutura acadêmica, com reorganização dos cursos de graduação e atualização de metodologias de ensino-aprendizagem, buscando a constante elevação da qualidade;

IV - diversificação das modalidades de graduação, preferencialmente não voltadas à profissionalização precoce e especializada;

V - ampliação de políticas de inclusão e assistência estudantil; e

VI - articulação da graduação com a pós-graduação e da educação superior com a educação básica.

Art. 3º O Ministério da Educação destinará ao Programa recursos financeiros, que serão reservados a cada universidade federal, na medida da elaboração e apresentação dos respectivos planos de reestruturação, a fim de suportar as despesas decorrentes das iniciativas propostas, especialmente no que respeita a:

I - construção e readequação de infra-estrutura e equipamentos necessárias à realização dos objetivos do Programa;

II - compra de bens e serviços necessários ao funcionamento dos novos regimes acadêmicos; e

III - despesas de custeio e pessoal associadas à expansão das atividades decorrentes do plano de reestruturação.

§ 1º O acréscimo de recursos referido no inciso III será limitado a vinte por cento das despesas de custeio e pessoal da universidade, no período de cinco anos de que trata o art. 1º, § 1º.

§ 2º O acréscimo referido no § 1º tomará por base o orçamento do ano inicial da execução do plano de cada universidade, incluindo a expansão já programada e excluindo os inativos.

§ 3º O atendimento dos planos é condicionado à capacidade orçamentária e operacional do Ministério da Educação.

Art. 4º O plano de reestruturação da universidade que postule seu ingresso no Programa, respeitados a vocação de cada instituição e o princípio da autonomia universitária, deverá indicar a estratégia e as etapas para a realização dos objetivos referidos no art. 1º.

Parágrafo único. O plano de reestruturação deverá ser aprovado pelo órgão superior da instituição.

Art. 5º O ingresso no Programa poderá ser solicitado pela universidade federal, a qualquer tempo, mediante proposta instruída com:

I - o plano de reestruturação, observado o art. 4º;

II - estimativa de recursos adicionais necessários ao cumprimento das metas fixadas pela instituição, em atendimento aos objetivos do Programa, na forma do art. 3º, vinculando o progressivo incremento orçamentário às etapas previstas no plano.

Art. 6º A proposta, se aprovada pelo Ministério da Educação, dará origem a instrumentos próprios, que fixarão os recursos financeiros adicionais destinados à universidade, vinculando os repasses ao cumprimento das etapas.

Art. 7º As despesas decorrentes deste decreto correrão à conta das dotações orçamentárias anualmente consignadas ao Ministério da Educação.

Art. 8º Este Decreto entra em vigor na data de sua publicação.

Brasília, 24 de abril de 2007; 186º da Independência e 119º da República.

LUIZ INÁCIO LULA DA SILVA

*Fernando Haddad*

*Paulo Bernardo Silva*

Este texto não substitui o publicado no DOU de 25.4.2007

## **Resolução CEPE nº 10/2007**

RESOLUÇÃO DO CONSELHO DE ENSINO, PESQUISA E EXTENSÃO N. 10/2007

Cria o Programa de Tutoria Especial (PTE), normaliza o apoio acadêmico a estudantes com necessidades especiais e dá outras providências.

O PRESIDENTE DA FUNDAÇÃO, REITOR DA UNIVERSIDADE DE BRASÍLIA E PRESIDENTE DO CONSELHO DE ENSINO, PESQUISA E EXTENSÃO, no uso de suas atribuições, conforme deliberação do referido Órgão Colegiado, em sua 411ª Reunião, realizada em 15/3/2007,

### **R E S O L V E:**

Art. 1º Criar o Programa de Tutoria Especial (PTE) no âmbito da Universidade de Brasília.

Parágrafo único. O PTE consiste no apoio acadêmico a estudantes da UnB com necessidade educacional especial, realizado por estudante da Universidade, sob a supervisão do professor da respectiva disciplina e da sua Unidade Acadêmica, com o acompanhamento do Programa de Apoio às Pessoas com Necessidades Especiais (PPNE) da Diretoria de Acompanhamento e Integração Acadêmica do Decanato de Ensino de Graduação (DAIA/DEG) e do Decanato de Pesquisa e Pós-Graduação (DPP).

### **Capítulo I – Do Programa de Tutoria Especial**

Art. 2º O PTE tem como objetivo:

- I. oferecer apoio acadêmico a estudantes regularmente matriculados nos cursos da UnB, que apresentem necessidades educacionais especiais e que estejam inscritos no Programa de Apoio às Pessoas com Necessidades Especiais/PPNE;
- II. subsidiar a Universidade na promoção da inclusão, em atividades acadêmicas, dos estudantes com necessidades educacionais especiais;
- III. possibilitar ao estudante universitário formação acadêmica sintonizada com a perspectiva da sociedade inclusiva.

### **Capítulo II – Dos Participantes do PTE**

Art. 3º Compõem o PTE:

- I.as Unidades Administrativas PPNE/VRT, DAIA/DEG e DPP;
- II.o estudante de graduação e de pós-graduação regularmente matriculado na UnB, doravante denominado Tutor Especial;
- III.professor em cuja disciplina o Tutorado estiver matriculado;
- IV.o Coordenador do curso e o Professor Orientador;
- V.o estudante de graduação ou de pós-graduação, com necessidades educacionais especiais, regularmente matriculado na UnB, doravante denominado Tutorado.

§ 1º O estudante de pós-graduação, com necessidades educacionais especiais, poderá ser tutorado por estudante de graduação inscrito como Tutor Especial junto ao PTE.

§ 2º Somente fará jus à participação no PTE o estudante com necessidade educacional especial que for cadastrado no PPNE e que demande os apoios estabelecidos nesta resolução.

### Capítulo III – Das Categorias da Tutoria Especial

Art. 4º A Tutoria Especial está classificada em duas categorias:

- I.Tutoria Especial Voluntária (não remunerada);
- II.Tutoria Especial Remunerada (por meio da concessão de bolsa de tutoria).

§ 1º A Tutoria Especial não gera qualquer tipo de vínculo empregatício entre o estudante e a FUB.

§ 2º O Tutor Especial não assumirá nenhuma responsabilidade funcional no âmbito das funções administrativas da UnB.

### Capítulo IV – Das Funções e Atividades de Apoio

PPNE, Professor, Coordenador, Orientador, Tutorado e Tutor Especial

Art. 5º Caberá ao PPNE:

- I. planejar as ações dos envolvidos no PTE;
- II. treinar e acompanhar o grupo de Tutores Especiais, Tutorados e professores participantes do PTE;
- III. prever estratégias de apoio e indicar recursos pedagógicos, institucionais e tecnológicos para atender às necessidades educacionais do Tutorado;
- IV. articular os diversos setores da Universidade na busca de alternativas para o atendimento dessas necessidades.

Art. 6º Caberá ao Professor:

- I. apresentar, no início de cada semestre, o Plano de Ensino da disciplina e cumprir o cronograma de atividades acadêmicas, informando, antecipadamente, quaisquer modificações no cronograma, com o objetivo de evitar prejuízo dos estudantes participantes do PTE;
- II. participar, juntamente com o PPNE, da busca de alternativas que visem ao atendimento das necessidades educacionais especiais dos estudantes participantes do PTE;
- III. repassar ao Tutor Especial ou ao Tutorado, antecipadamente, material didático a ser adaptado, sempre que necessário;
- IV. entregar diretamente ao PPNE, com antecedência, instrumentos de avaliação da disciplina por ele utilizados os quais necessitem de adaptação para os estudantes com necessidades educacionais especiais;
- V. acompanhar o desenvolvimento das atividades do PTE, pelos estudantes matriculados em sua disciplina, e solicitar apoio do PPNE sempre que julgar necessário;
- VI. responder formulário de avaliação do PTE, ao final do período letivo.

Art. 7º Caberá ao Coordenador do curso e/ou Orientador do Tutorado colaborar na indicação do Tutor Especial.

Art. 8º Caberá ao Tutorado:

- I. firmar Termo de Compromisso junto ao PPNE;
- II. definir, juntamente com o Professor e o Tutor Especial, o tipo de apoio e adaptações necessários para o bom andamento da disciplina;
- III. elaborar, juntamente com o Tutor Especial, o cronograma de atividades previstas no Plano de Ensino da disciplina;
- IV. informar sobre o Trancamento da Matrícula, geral ou parcial, e quaisquer modificações no cronograma, com o objetivo de evitar prejuízo de suas atividades acadêmicas e de seu(s) Tutor(es) Especial(is);
- V. providenciar, em tempo hábil, o material a ser adaptado, bem como o material de consumo e o equipamento a serem utilizados;
- VI. comunicar ao PPNE quando o Tutor Especial negligenciar o cumprimento das atividades previstas;
- VII. responder formulário de avaliação do PTE, ao final do período letivo.

Art. 9º Caberá ao Tutor Especial:

- I.firmar Termo de Compromisso junto ao PPNE;
- II.submeter-se a treinamento sempre que determinado pelo PPNE;
- III.cumprir o cronograma elaborado em conjunto com o Tutorado;
- IV.recorrer ao PPNE para dirimir dúvidas a respeito da adaptação de material;
- V.responsabilizar-se pela conservação dos materiais e equipamentos fornecidos como empréstimo pelo PPNE, pelo Professor ou pelo Tutorado;
- VI.informar sobre Trancamento de Matrícula, geral ou parcial, e quaisquer modificações no cronograma, com o objetivo de evitar prejuízo de suas atividades acadêmicas e de seu Tutorado;
- VII.devolver os materiais e equipamentos emprestados pelo PPNE, pelo Professor ou pelo Tutorado, até o último dia do período letivo;
- VIII.responder formulário de avaliação do PTE, ao final do período letivo;
- IX.oficializar sua desistência do PTE junto ao Tutorado, ao Professor da disciplina e ao PPNE.

Art. 10. Providenciar para que o horário de exercício das atividades, em hipótese alguma, sobreponha-se e/ou interfira nos horários das outras disciplinas nas quais estiver matriculado ou em outras atividades necessárias a sua formação acadêmica.

Art. 11. A Coordenação-Geral do PTE será exercida pelo PPNE, em conjunto com a DAIA/DEG e o DPP, que acompanharão e avaliarão o desenvolvimento do PTE e apresentarão os resultados ao Conselho de Ensino, Pesquisa e Extensão (CEPE), com vistas à revisão permanente da política do PTE na UnB.

Art. 12. Caberá à DAIA/DEG e ao DPP, em conjunto com o PPNE, a revisão e a distribuição das vagas e de recursos orçamentários e financeiros.

## Capítulo VI – Da Organização e Administração do Programa de Tutoria Especial

Art. 13. A organização e a administração do PTE serão conduzidas pelo PPNE, pela DAIA/DEG e pelo DPP.

Art. 14. Caberá ao PPNE:

- I.coordenar a execução do processo de seleção dos participantes, comunicando à DAIA/DEG e ao DPP, por meio de formulário específico;
- II.providenciar assinatura, pelos Tutores Especiais, de Termo de Compromisso, e manter os arquivos referentes ao processo seletivo à disposição da DAIA/DEG;
- III.encaminhar à DAIA/DEG, mensalmente, em formulários específicos, a frequência dos

- Tutores Especiais;
- IV. encaminhar aos Tutores Especiais, Tutorados e professores da disciplina, semestralmente, formulários de avaliação do PTE;
- V. encaminhar à DAIA/DEG e ao DPP, semestralmente, os resultados das avaliações respondidas pelo Tutores Especiais, Tutorados e professores da disciplina, para posterior apreciação conjunta, com vistas à melhoria do PTE.

Art. 15. Caberá à DAIA/DEG:

- I. elaborar calendário das atividades da Tutoria Especial;
- II. encaminhar, mensalmente, folha de frequência do PTE à Secretaria de Recursos Humanos (SRH);
- III. encaminhar à Secretaria de Administração Acadêmica, ao final de cada período letivo, a lista de Tutores Especiais que terão direito à concessão de créditos.

## Capítulo VII – Dos Recursos Financeiros

Art. 16. Caberá à FUB o provimento de recursos destinados a custear as bolsas do PTE para cada período.

Parágrafo único. O valor da bolsa do Programa de Tutoria Especial e a forma de pagamento serão iguais ao do Programa de Monitoria.

Art. 17. No caso da participação remunerada no PTE, o pagamento mensal da bolsa será efetuado pela SRH, mediante solicitação da DAIA/DEG, por meio da agência bancária designada pela Diretoria de Contabilidade e Finanças do Decanato de Administração (DCF/DAF) para esse fim.

## Capítulo VIII – Do Cronograma de Atividades do Programa de Tutoria Especial

Art. 18. O cronograma do PTE observará o calendário acadêmico da Universidade de Brasília.

## Capítulo IX – Da Inscrição de Atividades do Programa de Tutoria Especial

Art. 19. Poderão participar da seleção para Tutor do PTE:

- I. estudantes regularmente matriculados em disciplina de graduação ou de pós-graduação da Unidade Acadêmica em que houver estudantes com necessidades educacionais especiais e que necessitem de apoio especial;
- II. estudantes não-matriculados na disciplina em que se encontra um estudante com necessidades educacionais especiais, desde que já tenham cursado a referida disciplina com aprovação;
- III. estudantes de outras universidades nacionais ou estrangeiras, em caráter excepcional e na modalidade não-remunerada, somente durante o tempo de permanência na UnB e em virtude de convênios que assim o permitam;
- IV. estudante que não tenha nenhuma reprovação no semestre anterior em disciplinas obrigatórias do seu curso.

Parágrafo único. Nos apoios a estudante em disciplina de final de curso ou de pós-graduação, o candidato a Tutor Especial deve ter, no mínimo, 50% dos créditos do seu curso concluídos, e não ter reprovação no semestre anterior em disciplinas obrigatórias.

Art. 20. A seleção será feita, prioritariamente, mediante indicação do Tutorado e/ou do Professor da disciplina, do Coordenador do curso e do Orientador, em comum acordo com o PPNE..

Art. 21. Na falta de tutores com o perfil estabelecido, o PPNE poderá indicar Tutor Especial, a critério do PPNE, em consonância com a presente Resolução.

Art. 22. Para efeito de seleção e exercício da atividade, o estudante deverá estar regularmente matriculado no número mínimo de créditos do seu curso e não poderá ter sido identificado no semestre, pela SAA, como estudante em situação de risco de desligamento.

Art. 23. O candidato não poderá exceder o número máximo de créditos do curso por semestre, a menos que seja provável formando naquele semestre.

Art. 24. Será vedada a participação no PTE de estudantes que estejam exercendo atividades de Monitoria, mesmo que voluntária.

Art. 25. A concessão de Tutoria Especial remunerada será feita pelo PPNE e DAIA/DEG, segundo os critérios a seguir:

- I.no caso de já ter sido Tutor Especial, ter apresentado bom desempenho;
- II.ser preferencialmente estudante de baixa renda, com estudo socioeconômico elaborado pela Diretoria de Desenvolvimento Social do Decanato de Assuntos Comunitários (DDS/DAC);
- III.não ter qualquer outra atividade remunerada paga pela mesma fonte;
- IV.ter recebido bolsa em menor número de semestres.

Art. 26. Ao Tutor Especial que tiver concluído com aprovação a disciplina que tenha prestado apoio, após homologação do relatório pela DAIA/DEG, serão concedidos:

- I.dois créditos pela atividade no período;
- II.registro da atividade em Histórico Escolar.

§ 1º A concessão de créditos integralizará o limite permitido para o Módulo Livre do curso do Tutor Especial.

§ 2º É vedada a concessão de créditos em mais de uma disciplina por semestre, pela atividade de tutoria.

Art. 27. Ao Tutor Especial de disciplina “Trabalho de final de curso” ou equivalente ou de estudante de pós-graduação que tiver concluído com aprovação todas as disciplinas do semestre, após homologação do relatório pela DAIA/DEG, serão concedidos:

- I.dois créditos pela atividade no período;
- II.registro da atividade em Histórico Escolar.

§ 1º A concessão de créditos integralizará o limite permitido para o Módulo Livre do curso do Tutor Especial.

§ 2º É vedada a concessão de créditos em mais de uma disciplina por semestre, pela atividade de tutoria.

Art. 28. No caso de integralizado o limite permitido para o Módulo Livre, o estudante manterá o direito às demais concessões.

Art. 29. A concessão de créditos valerá para o cumprimento das normas vigentes de acompanhamento acadêmico.

Art. 30. Não serão concedidos créditos ao Tutor Especial que desistir, por qualquer motivo, da Tutoria Especial.

Art. 31. O Tutorado poderá ter mais de um Tutor Especial na mesma disciplina, desde que a necessidade seja justificada e aprovada pelo PPNE.

#### Capítulo XI – Da suspensão da participação no Programa de Tutoria Especial

Art. 32. A participação do Tutor Especial no PTE poderá ser suspensa no caso de:

- I. não-cumprimento das atividades da programação da disciplina;
- II. desistência de participação no PTE;
- III. ocorrência de abandono ou trancamento de disciplina pelo estudante Tutorado.

Parágrafo único. O PPNE, em comum acordo com o Tutor Especial, poderá incumbir a este último de desenvolver tutoria especial junto a outro tutorado, situação na qual o Tutor Especial fica autorizado a atuar com outro tutorado e em outra(s) disciplina(s).

Art. 33. A participação do Tutorado e do Tutor Especial no PTE poderá ser suspensa a qualquer momento pelo PPNE, quando não houver o cumprimento das presentes normas.

#### Capítulo XII – Das Disposições Gerais e Transitórias

Art. 34. O PTE será implantado no Sistema de Informações Acadêmicas (SIAC) da UnB.

Art. 35. A cada dois períodos letivos, após a implantação plena do PTE, proceder-se-á a avaliação integral do sistema pelo PPNE, DAIA, DEG e DPP.

Art. 36. Os casos omissos serão analisados pelo PPNE, DAIA/DEG e DPP e decididos nas instâncias superiores competentes.

Art. 37. Esta Resolução entra em vigor a partir desta data, revogadas disposições em contrário.

Brasília, 3 de abril de 2007.

Timothy Mulholland

Reitor