Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Análise do Desempenho Acadêmico dos Estudantes de Engenharia Computação da Universidade de Brasília

Italo Franklin C. Vaz

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Orientadora
Prof.a Dr.a Maristela Terdo de Holanda

Brasília
2024

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Análise do Desempenho Acadêmico dos Estudantes de Engenharia Computação da Universidade de Brasília

Italo Franklin C. Vaz

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Prof.a Dr.a Maristela Terdo de Holanda (Orientadora)
CIC/UnB

Prof. Dr. Jonathan Rosa Moreira
CIC/UnB

Prof. Dr. Marcelo Grandi Mandelli
CIC/UnB

Prof. Dr. João Luiz Azevedo de Carvalho
Coordenador do Curso de Engenharia da Computação

Brasília, 20 de Setembro de 2024

Dedicatória

Dedico este trabalho à minha esposa Bianka Fernanda, que sempre me apoiou e incentivou. Aos amigos Camila Coutinho, Jônatas Júnior e Lucas Borges, por terem sido parte importante e indispensável da minha graduação, sem eles eu não teria chegado até aqui. Dedico também à minha amiga Nina Viegas, que sempre tinha as respostas certas para as questões complicadas da vida.

Agradecimentos

Agradeço primeiramente a Deus, que me concedeu a sabedoria necessária para a realização deste trabalho. Agradeço à Prof.a Dr.a Maristela Holanda pela oportunidade e por me orientar durante este trabalho. Agradeço às colegas Giovana Pinho e Kailany Ketulhe, que me ajudaram e contribuíram para a realização deste trabalho. Também agradeço à Rita de Cássia, Antônio Geraldo e Vitor Gustavo, minha família, que me acompanharam em cada momento desta jornada.

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por disponibilizar o Acesso ao Portal de Periódicos.

Resumo

A Lei 12.711/2012, conhecida como a Lei de Cotas, foi implementada no Brasil com o objetivo de ampliar o acesso à educação superior para pessoas pretas, pardas, indígenas, de escola pública e portadoras de deficiência física. A implementação foi gradual, de maneira que em 2013 a reserva teria de ser 12,5%, atingindo os 50% das vagas reservadas em 2016. A lei busca corrigir desigualdades sociais e raciais e promover maior inclusão no ensino superior.

No contexto da Universidade de Brasília, este trabalho buscou entender o desempenho dos alunos cotistas e não cotistas que foram beneficiados pela Lei das Cotas, foram analisados os dados referentes ao curso de Engenharia de Computação. Utilizando ferramentas de visualização e análise de dados, a análise focou-se em três fases distintas: o ingresso dos alunos no curso, o desempenho nas disciplinas iniciais e análise dos fatores que podem causar a evasão dos alunos. Além disso, também foi realizada a análise por sexo.

Por meio de algoritmos de predição baseados em aprendizado de máquina, observou-se o padrão de evasão dos alunos do curso e identificaram-se os principais fatores que podem contribuir com à desistência. Além disso, foi possível apontar áreas que necessitam de maior atenção e melhorias, especialmente relacionadas às disciplinas iniciais cursadas pelos estudantes.

Palavras-chave: Engenahria de Computação, Lei das Cotas, Evasão, Aprendizado de Máquina, Power BI, Python

Abstract

Law 12.711/2012, known as the Quota Law, was implemented in Brazil to expand access to higher education for Black, Brown, Indigenous students, those from public schools, and individuals with physical disabilities. The implementation was gradual, beginning with a 12.5% seat reservation in 2013, reaching 50% by 2016. The law aims to address social and racial inequalities and promote greater inclusion in higher education.

In the context of the University of Brasília, this study aimed to understand the performance of both quota and non-quota students who benefited from the Quota Law, with a focus on the Computer Engineering program. Using data visualization and analysis tools, the research examined three distinct phases: student enrollment, performance in initial courses, and factors contributing to student dropout. Additionally, an analysis by gender was conducted.

Through machine learning prediction algorithms, patterns of student dropout were identified, along with key factors that may contribute to attrition. Furthermore, areas requiring greater attention and improvement were highlighted, particularly in relation to the initial courses taken by students.

Keywords: Computer Engineering, Quata Law, Dropout Analysis, Machine Learning, PowerBI , Python

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Analisando os registros históricos da Universidade de Brasília em relação às cotas raciais, nota-se que desde a primeira discussão sobre este tema na universidade, houve divergências nas opiniões acerca do real propósito e da eficiência desta política pública [3]. Segundo Carvalho, a primeira proposta de cotas para negros na universidade foi apresentada em 1999, em resposta a um caso de desigualdade racial no programa de mestrado do Departamento de Antropologia da UnB [4]. O caso Arí, como ficou conhecido na época, chamou a atenção de muitos após o primeiro estudante negro no programa de mestrado, que já existia há 20 anos na UnB, levar mais de dois anos para ter sua nota revisada e finalmente ser aprovado na primeira matéria do curso. Neste contexto, o pensamento que motivou a primeira proposta de cotas raciais na UnB foi [4]: "Se é tão difícil manter um único aluno negro na UnB, vamos propor um sistema de cotas, para que pelo menos alguns negros permaneçam.".

Após quase cinco anos de debates, a Universidade de Brasília entendeu que a reserva de vagas para negros seria uma ação afirmativa na busca por inclusão e desenvolvimento igualitário, além de reduzir a desigualdade racial. Em 2004, no segundo semestre, realizou-se o primeiro vestibular com cotas para negros [5]. Naquele semestre foram ofertadas 2 mil vagas para 61 cursos e 20% dessas vagas foram reservadas para candidatos negros [6] [7].

Apesar das afirmações contrárias às cotas, que diziam que a adoção dessa política levaria a uma ruptura ideológica acerca da mistura etnológica brasileira, hoje é possível observar, por meio de comprovações, a existência da desigualdade racial no meio acadêmico e por meio do aumento da representatividade de alunos cotistas na Universidade é possível perceber o impacto positivo resultante da política de cotas raciais adotada pela UnB [3].

No entanto, foi no ano de 2012 que o cenário de desigualdade racial no meio acadêmico começou a mudar, a Lei 12.711/2012 foi decretada, com o objetivo de observar e estudar

estratégias para mudar a realidade acadêmica de grupos sociais como afrodescendentes e indígenas brasileiros [8].

A lei determinava que as universidades federais deveriam reservar 50% das vagas oferecidas para os cursos de nível superior para estudantes que cursaram todo o ensino médio em instituições de ensino público. Segundo disposição do art. 3o da Lei no 12.711/2012 esses estudantes deveriam ter cursado o ensino médio em escolas públicas (não PPI), ser autodeclarados pretos, pardos, indígenas (PPI) ou pessoas com deficiência. Após esse marco histórico nas universidades, é possível perceber uma grande mudança na presença de PPIs (pretos, pardos e indígenas) nas universidades federais do Brasil. O número de ingressos de estudantes cotistas vem aumentando significativamente nos últimos anos, sem prejudicar os alunos não cotistas, visto que as vagas ofertadas nas universidades federais quase triplicaram, de 109.184 vagas em 2017 para 393.550 vagas em 2018 [8].

Segundo Santos, frequentemente se questiona o desempenho dos alunos cotistas em comparação aos não cotistas e se a qualidade dos cursos se mantém a mesma após a implantação da política pública de cotas, visto que os ingressantes cotistas são, em alguns casos, classificados com nível de aprendizado inferior ao dos alunos não cotistas, ou que não teriam a capacidade necessária para ingressar na universidade. No entanto, a autora afirma que "tal afirmação não passa de uma racionalidade de senso comum, não resistindo a uma análise empírica da situação."[9]. Em contrapartida, em 2020 foram feitos estudos nas três principais universidades federais do Brasil (Universidade de Brasília, Universidade Estadual do Rio de Janeiro e Universidade Federal do Espirito Santo) mostram que o desempenho dos alunos cotistas ao longo de cursos como Engenharia de Computação, Administração, Direito, Ciências da Computação e Medicina quase se assemelha ao dos alunos não cotistas, demonstrando que, mesmo que os alunos cotistas não possuam a mesma base de ensino, estão se equiparando no desenvolvimento das matérias ao longo do curso [9].

Carvalho e Segato em 2002, reafirmam a importância das cotas pela observação de alguns pontos relevantes: 45% da população brasileira é negra e representa apenas 2% dos universitários. Por outro lado, a população branca, que constitui 54%, representa 98% dos universitários do país. Em 2002 na UnB 99% do corpo discente era branco e 90% dos alunos também eram brancos [7].

Neste contexto, este trabalho tem como objetivo a análise dos dados educacionais da UnB, considerando o desempenho acadêmico dos alunos cotistas beneficiados pela Lei 12.711/2012, conhecida pela lei da cotas, o estudo foi especificamente direcionado para os alunos do curso de Engenharia de Computação.

Foram definidas questões de pesquisa para trabalhar com os dados coletados, e as informações foram separadas em três grupos: ingresso dos candidatos do curso de Enge-

nharia de Computação, desempenho durante o curso e, por último, conclusão do curso. Em cada grupo, a análise foi feita observando principalmente os seguintes aspectos: Tipo de cota e sexo.

Apesar da UnB já adotar cotas para negros desde 2004, neste trabalho as análises se voltaram para as cotas que se iniciaram em 2013 com a Lei das Cotas. Como a Lei de Cotas subdivide os alunos em diferentes grupos (incluindo estudantes pretos, pardos, indígenas (PPI), alunos de escolas públicas, pessoas com deficiência física e aqueles com renda familiar limitada), é fundamental realizar uma análise mais detalhada sobre o desempenho acadêmico desses alunos no curso. Essa análise permite identificar diferenças no desempenho acadêmico e na taxa de evasão entre os grupos.

A diversidade de gênero nos cursos de Engenharia de Computação é uma questão estudada na literatura, dado que o curso é majoritariamento masculino. Entre os anos de 2000 e 2013, apenas 17% dos alunos que concluíram o curso eram mulheres. No contexto da Universidade de Brasília o ingresso de alunas e bem menor que o ingresso de alunos, o menor registro para o curso de Engenharia de computação foi em 2015, onde apenas 5% dos alunos ingressantes eram meninas [10]. Além disso, o baixo índice de graduação de alunas é evidente: nesse período, o número de homens que concluíram o curso aumentou 98%, enquanto o número de mulheres diminuiu 8% [5], somado ao preconceito e estereótipos estruturados na sociedade desmotivam meninas e mulheres a ingressar no meio tecnológico [11], tornando este assunto relevante para a análise dos dados educacionais.

O curso de Engenharia de Computação da UnB começou em 2009. Esse curso é oferecido no campus Darcy Ribeiro, e sua criação foi parte de uma expansão dos programas das engenharias da universidade para atender à crescente demanda por profissionais qualificados na área de tecnologia. No curso de Engenharia de Computação os alunos aprendem principalmente a projetar sistemas de hardware digital, focado no desenvolvimento de software, com ênfase em software para dispositivos digitais e suas interfaces com usuários e outros dispositivos [12].

Foi realizada uma análise geral do desempenho dos alunos durante o curso. Paralelamente, foi destacado os dados das principais matérias do curso com o intuito de identificar padrões de aprovação e reprovação, a fim de propor possíveis melhorias para as disciplinas em questão. Desta forma, foi possível ter uma visão mais crítica dos efeitos a longo prazo das cotas sociais e raciais e como essa importante política pública vem interferindo positivamente na universidade e, principalmente, nos grupos sociais por ela alcançados.

1.1 Objetivo

Este trabalho teve como objetivo geral realizar uma análise de desempenho acadêmico dos alunos cotistas e não cotistas do curso de Engenharia da Computação da UnB. A análise foi feita em três momentos específicos da jornada: ingresso, desempenho durante o curso e conclusão do curso. Assim, foram comparados principalmente o gênero, o tipo de escola e o tipo de cota.

1.1.1 Objetivos Específicos

Há muitas especulações sobre a real importância das cotas raciais e sociais nas universidades brasileiras, muitas das quais estão relacionadas a como as cotas afetam os cursos nas instituições públicas. Portanto, para analisar este contexto, os objetivos específicos deste trabalho são:

1. Analisar os dados referentes ao ingresso de aluno no curso de Engenharia da Computação.
2. Identificar as disciplinas com maiores taxas de aprovação e reprovação no início e no final do curso.
3. Visualizar os dados referentes ao ingresso de alunos cotistas e não cotistas.
4. Realizar análises dos alunos cotista e não cotistas, separado por gênero.
5. Analisar os resultados comparando o desempenho dos alunos ao longo das disciplinas cursadas.

1.2 Estrutura do Trabalho

Este trabalho foi organizado em quatro capítulos específicos, que explicam separadamente os pontos mais importantes para a compreensão e análise do tema proposto, estruturados da seguinte forma:

- No Capítulo 2, foi apresentado o referencial teórico necessário para o entendimento deste trabalho, contextualizando a análise realizada. Foram examinadas estatísticas descritivas com o Microsoft Power BI, uma ferramenta de BI (*business intelligence*) que permite a visualização interativa de dados e a criação de relatórios. Além disso, foram explorados algoritmos de *machine learning* para treinar modelos preditivos e analisar as tendências de desempenho dos alunos cotistas e não cotistas da Universidade de Brasília. Também foram utilizadas linguagens de programação como

Python e DAX (*Data Analysis Expressions*), uma linguagem de fórmula que permite criar cálculos e manipular dados.

- No Capítulo 3, é apresentado a metodologia realizada para executar a análise descritiva dos dados, bem como a análise realizada por meio dos resultados obtidos da previsão dos modelos de aprendizado de máquina utilizadas neste trabalho.
- No Capítulo 4, foram explorados de maneira detalhada os dados coletados e as análises realizadas para cada grupo específico, como tipo de cota, gênero, forma de ingresso na UnB e origem escolar (pública ou particular), entre outros. Nesse contexto, foi analisado o desempenho desses grupos ao longo do curso de Engenharia de Computação, desde o início até a saída do curso. A partir da análise dos dados coletados para cada grupo, foi possível responder as questões inicialmente levantadas neste trabalho.
- No Capítulo 5, com base nas análises dos capítulos anteriores, foram apresentadas as conclusões deste trabalho. Este capítulo sintetizou os pontos explorados e as conclusões alcançadas, justificando cada objetivo proposto anteriormente. Além disso, foram apresentadas recomendações futuras com base nos resultados obtidos.

1.3 Metodologia de Pesquisa

O desenvolvimento deste trabalho foi dividido da seguinte forma:

1. Tratamento dos dados do *dataset* do Sistema Integrado de Gestão das Atividades Acadêmicas (SIGAA).
2. Tratamento dos dados do Sigra (Sistema de Informações Acadêmicas de Graduação).
3. Importação dos dados para o Power BI.
4. Análise e criação da visualização dos dados.
5. Treinamento dos modelos preditivos e análise das tendências de desempenho dos alunos cotistas e não cotistas do curso de Engenharia de Computação.
6. Análise e exposição dos resultados obtidos.
7. Conclusão dos dados observados.

Capítulo 2

Análise de Dados Educacionais

Neste capítulo, são apresentados os principais conceitos para a compreensão do trabalho, contextualizando a análise realizada. Expondo as tecnologias utilizadas para análise dos dados e os algoritmos de *machine learning*, que foram considerados fundamentais para a interpretações dos resultados.

2.1 Sistema de Cotas

A Universidade de Brasília foi pioneira na implementação do sistema de cotas no Brasil, sendo uma das primeiras instituições a adotar essa política inclusiva. Desde o segundo semestre de 2004, a UnB passou a reservar 20% das vagas do vestibular para candidatos autodeclarados negros. Com a entrada em vigor da Lei das Cotas (Lei 12.711/2012) em 2013, a distribuição das vagas foi reformulada, com 5% destinadas a alunos negros, 50% a estudantes oriundos de escolas públicas, e 45% para alunos não cotistas.

Para concorrer às vagas reservadas pela Lei das Cotas, os estudantes devem atender aos seguintes critérios:

- Baixa Renda: Com renda familiar bruta igual ou inferior a $1\frac{1}{2}$ salário mínimo per capita;
- Alta Renda: Com renda familiar bruta superior a $1\frac{1}{2}$ salário mínimo per capita;
- PPI: Que se declaram pretos, pardos ou indígenas;
- Não PPI: Que não se declaram pretos, pardos ou indígenas;
- PCD: Com deficiência.

Este estudo considera tanto as cotas previstas pela Lei das Cotas quanto aquelas especificamente voltadas para alunos negros.

2.2 Estudo dos Dados

Os dados são informações coletadas de diversas fontes, que podem ser estruturadas ou não, e servem como base para a análise e a tomada de decisões. De acordo com a Enciclopédia Britannica [13], a análise de dados envolve coletar, limpar, transformar, descrever, modelar e interpretar essas informações para obter informações úteis. O conceito de ciência de dados tem se expandido como uma disciplina que abrange análise avançada de dados, mineração de dados, aprendizado de máquina e outras técnicas para extrair conhecimento útil e transformá-lo em estratégias acionáveis. A ciência de dados, como descrito na revista Springer Nature Computer Science [14], é interdisciplinar e engloba estatísticas, informática e várias outras disciplinas, com o objetivo de transformar dados em conhecimento, aplicando uma metodologia que vai da análise de dados à geração decisões.

No contexto educacional, os dados educacionais referem-se a informações coletadas e analisadas para melhorar o desempenho dos estudantes, avaliar a eficácia dos professores e aperfeiçoar programas educacionais. A análise de dados nessa área inclui a avaliação de aprendizado, análises de programas e monitoramento do progresso dos estudantes, ajudando na identificação de áreas que necessitam de melhorias e no desenvolvimento de estratégias educativas mais eficazes. Segundo um estudo publicado pela Research Method, a análise de dados educacionais é fundamental para medir o progresso dos estudantes, avaliar o impacto de intervenções pedagógicas e orientar políticas educacionais baseadas em evidências [15].

2.3 Tratamento de Dados

O processo de ETL (*Extract, Transform, Load*) é uma prática essencial no gerenciamento e integração de dados. O processo inicia com a extração, na qual os dados são coletados de diversas fontes, como bancos de dados. Essa fase precisa ser realizada de forma eficiente para minimizar o impacto nos sistemas de origem e garantir a consistência dos dados. Em seguida, ocorre a transformação, onde os dados são adaptados para atender às necessidades específicas do sistema de destino ou da análise que será realizada. Isso pode incluir limpeza, formatação ou aplicação de filtros, assegurando que os dados estejam prontos para a análise ou armazenamento. De acordo com Simitsis e Vassiliadis em [16], essa etapa pode exigir uma modelagem complexa para lidar com as particularidades dos dados de entrada e as demandas da análise final.

Na fase de carregamento, os dados transformados são inseridos no sistema de destino, que frequentemente é um banco de dados ou software de análise de dados. Este estágio é crítico para assegurar que os dados sejam carregados sem interrupções, mesmo quando

volumes de dados são elevados. A importância de um ETL bem estruturado está em sua capacidade de garantir que os dados, uma vez carregados, estejam prontos para uso em análises e relatórios [17].

2.4 Visualização de Dados

A visualização de dados é um recurso fundamental na análise de dados. Seu objetivo principal é transmitir informações ou conceitos complexos de forma clara e precisa, facilitando a análise e a compreensão dos dados pelos usuários. Embora a prática de visualizar dados seja antiga, ela ganhou maior importância com o aumento da quantidade de informações processadas atualmente e a evolução do mercado tecnológico [18].

Para realizar a análise dos dados extraídos foi utilizado o recurso de visualização dos dados em forma de gráficos, principalmente os gráficos de pizza (Figura 2.2), gráficos de barras (Figura 2.1) e gráficos de linhas (Figura 4.1), pois foram os tipos gráficos que mais se encaixavam com o modelo de dados estudados.

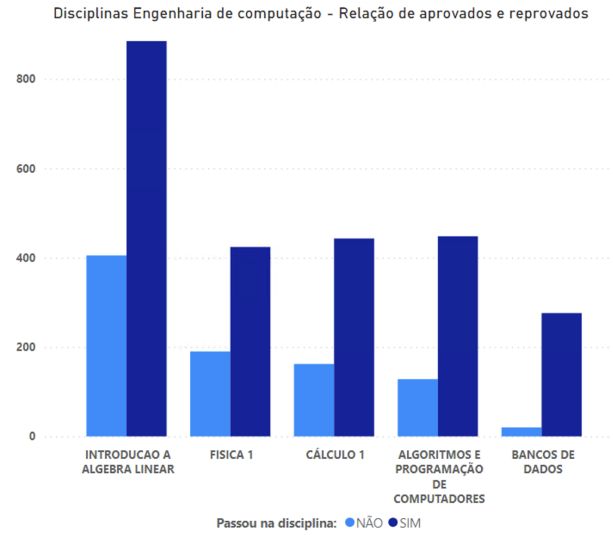

Figura 2.1: Gráfico de Barras.

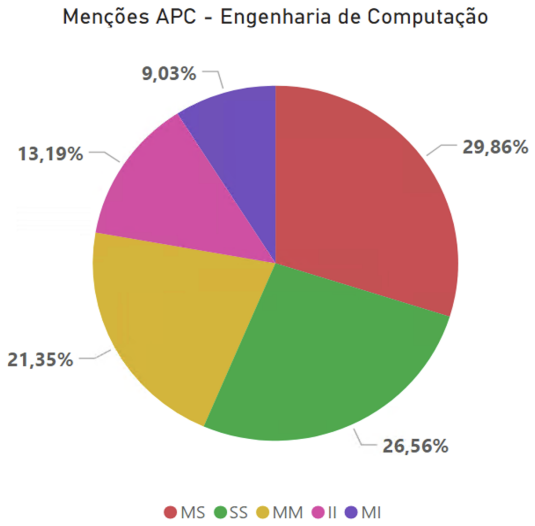

Figura 2.2: Gráfico de Pizza.

Por meio da visualização de dados, obter uma compreensão de sistemas complexos se torna mais fácil e intuitivo. Em sua dissertação [19], Hansen menciona que a visualização de dados representa graficamente informações, facilitando comparações, identificação de padrões e detecção de alterações. E quando os dados são analisados por meio de imagens e gráficos, pode-se destacar as principais características deste dados, tornando mais natural a interpretação da informação em questão.

2.5 Algoritmos de Aprendizado de Máquina

Com o avanço das tecnologias em IA (Inteligência Artificial), os estudos e análises nas ciências de dados ganharam um grande aliado: o Aprendizado de Máquina. Simplificadamente, os modelos de aprendizado de máquina podem ser classificados em dois grupos [20]: aprendizado supervisionado, onde os dados de treinamento incluem uma porcentagem de exemplos corretos da informação, servindo como referência para a previsão; e aprendizado não supervisionado, onde os modelos precisam identificar padrões presentes em conjuntos de dados não rotulados ou não treinados.

Para este trabalho, foram utilizados trés modelos de treinamento: GBM (*Gradient Boosting Machine*), SVM (*Support Vector Machine*) e *Random Forest*. Todos eles são considerados como modelos supervisionados, pois uma porcentagem da amostra foi treinada com a informação correta do dataset para que fossem comparada com o restante da amostra.

No aprendizado de máquina supervisionado utilizado neste trabalho, é utilizado dados de treinamento que incluem exemplos de como uma variável de entrada x está relacionada a uma variável de saída y. Então se aplica um método a esses dados de treinamento para adaptá-los, para então prever a variável y para um novo conjunto de dados de teste, não conhecido, no qual apenas o x é conhecido. É nesse processo que ocorre o treinamento de um modelo a partir dos dados de treinamento [21].

A partir da obtenção dos dados tratados, foi possível executar os algoritmos de predição para obter os valores relacionados a conclusão ou evasão baseada em cada variável. Para medir o desempenho dos classificadores, e entender se um modelo esta tendo um desempenho adequado, utilizou-se as medidas de desempenho da biblioteca Scikit Learn [1]. Existem 5 métricas principais que mostram se aquele algoritmo está realizando boas previsões:

- Acurácia: fração de previsões corretas feitas no conjunto de teste. Para dataset desbalanceados a Acurácia necessita de complementação de outras métricas.
- Precisão: é a fração das previsões positivas que o algoritmo retorna, podendo estar correto ou não.
- Recall: é a fração de todas as previsões positivas que o algoritmo retorna, sendo elas todas a previsões que de fato são positivas.
- F1-Score: é a média harmônica entre precisão e recall, é útil em casos que são necessário os resultados de ambas as métricas.

[1] https://scikit-learn.org

- Support: indica dentro do conjunto de teste a quantidade de cada feature classificada.

Essas métricas de classificação são baseadas na matriz de confusão (Figura 2.3). Segundo a documentação da biblioteca Scikit Learn [22], cada linha representa as classes reais, enquanto cada coluna representa as classes previstas. Dessa matriz, obtem-se quatro conceitos importantes: verdadeiros positivos (TP), falsos positivos (FP), falsos negativos (FN) e verdadeiros negativos (TN)[2]. Tanto a precisão quanto o recall usam esses conceitos nas classificações de modelos. A precisão é uma fração onde: $\frac{TP}{TP+FP}$, enquanto o recall: $\frac{TP}{TP+FN}$, essas duas métricas compõem a Acurácia: $\frac{previsõescorretas}{totaldeprevisões}$, que indica se a previsão do algoritmo foi boa ou não. por sua vez a métrica F1-score é representada da seguinte forma:

$$F1 = \frac{2}{\frac{1}{\text{precisão}} + \frac{1}{\text{recall}}}$$

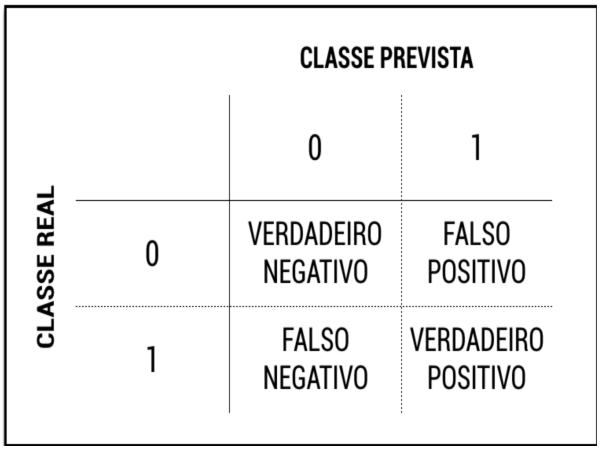

Figura 2.3: Matriz de confusão

2.5.1 GBM (Gradient Boosting Machine)

Por conta de sua eficiência de treinamento e boa precisão, um dos algoritmos utilizados foi o GBM, conhecido por ser rápido e de alto desempenho é um algoritmo baseado na técnica de árvore de decisão, ele prever a qualidade da ligação entre as amostras escolhidas, podendo ser usado para ordenação, classificação e regressão em problemas de aprendizado de máquina [23].

Neste trabalho usaremos o algoritmo para classificar categorias e suas importâncias em relação a saída dos alunos do curso de Engenharia de Computação. O GBM é um

[2]https://nbviewer.org/github/programacaodinamica/machine-learning/

algoritmo de "boosting"[3] que relaciona os modelo que tem o desempenho de predição mais fraco, obtendo modelos que tem um desempenho de predição significativamente melhor.

O algoritmo implementa uma série de árvores de decisão de forma sequencial, onde cada nova árvore tenta corrigir os erros cometidos pelas árvores anteriores. O processo começa com uma previsão inicial simples, como a média dos valores da variável escolhida (forma_saida_curso). Em seguida, a primeira árvore é criada para ajustar os resíduos (distancia entre a reta da função do algoritmo e a amostra de teste) da previsão inicial. Cada árvore subsequente ajusta os resíduos restantes das árvores anteriores, aprimorando progressivamente a precisão do modelo. A cada iteração, as previsões são atualizadas somando-se as contribuições das novas árvores, multiplicadas por uma taxa de aprendizado que controla o impacto de cada árvore. Esse ciclo de construção e ajuste de árvores continua até que o número predefinido de árvores seja atingido ou que os erros sejam suficientemente pequenos. O resultado final é a combinação das previsões de todas as árvores [24].

Após a execução do pipeline aplica-se a validação cruzada (Figura 2.4), uma técnica que permite estimar o desempenho do modelo em todo o conjunto de dados, realizando várias divisões e alternando as amostras para teste. O conjunto de dados original é dividido em cinco partes iguais. Em seguida, cada uma das cinco partes é usada, uma de cada vez, como conjunto de teste, enquanto as outras quatro partes são usadas para treinar o modelo. Isso resulta em cinco diferentes estimativas de precisão, que podem ser usadas para calcular a precisão média e a variância [25].

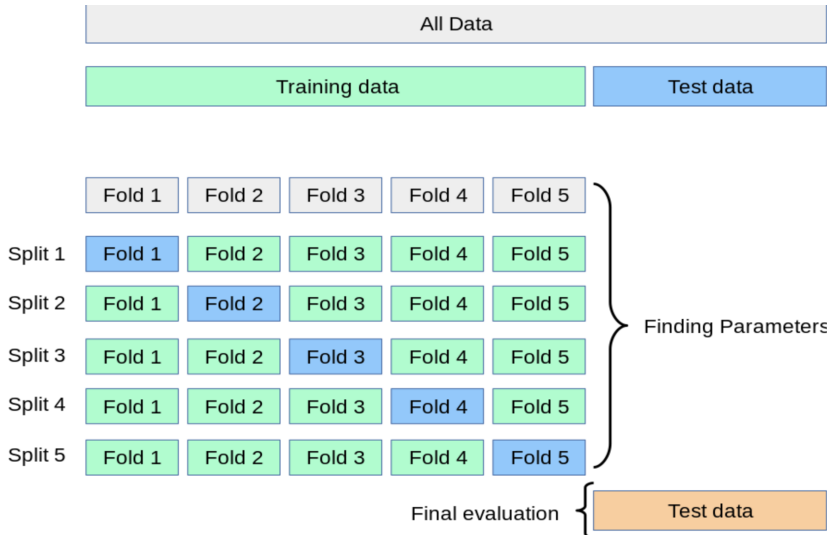

Figura 2.4: Validação Cruzada. Fonte: Validação k-fold Scikit-Learn.

[3]É uma técnica que cria um aprendiz forte a partir de muitos aprendizes fracos. O algoritmo de boosting treina sequencialmente uma série de aprendizes fracos, onde cada um corrige os erros do anterior.

2.5.2 SVM (Support Vector Machine)

O algoritmo Support Vector Machine neste trabalho foi utilizado para classificação porém é um modelo muito potente sendo possível ser utilizado também para regressão.

O algoritmo SVM é bem conhecido por sua capacidade de separar classes que podem ser divididas por uma linha reta, um conceito conhecido como classes linearmente separáveis [1]. Para entender melhor este conceito, utilizamos a biblioteca Seaborn do Python para visualizar a fronteira entre os dados. Neste exemplo, faremos uma análise das features 'ira' e 'APC', pois o pairplot considera apenas as classes numéricas do dataset. Olhando os gráficos, é possível ter uma boa intuição de que os dados são sim linearmente separáveis e é possível traçar uma fronteira entre os pontos como podemos observar na Figura 2.5.

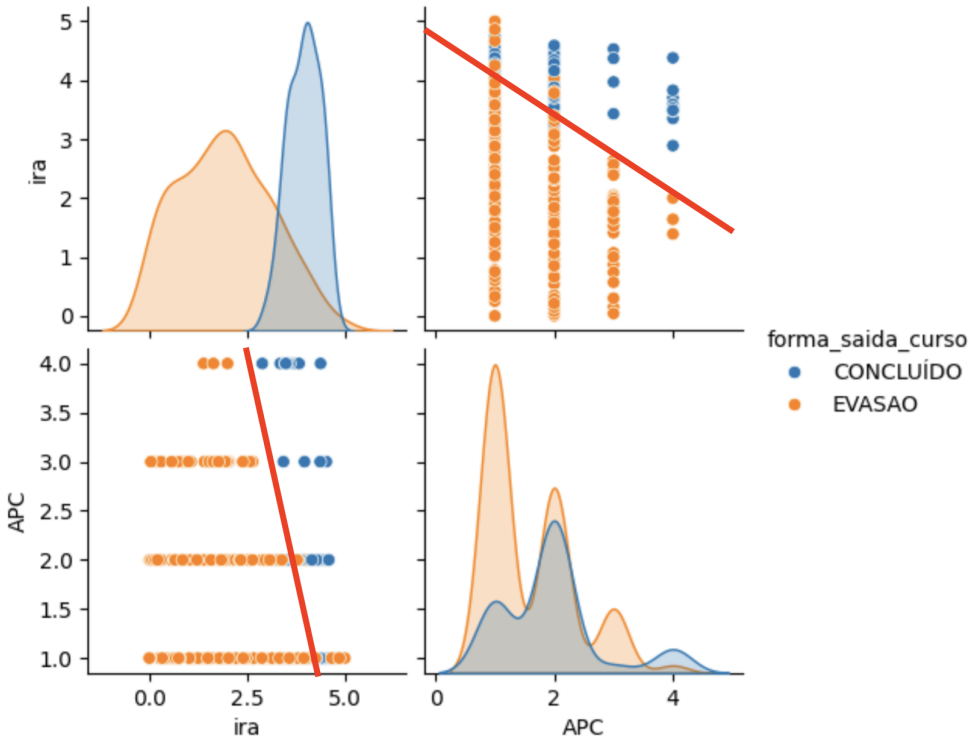

Figura 2.5: Pairplot: Grafico de conclusão e evasão das features ira e APC

A lógica por trás do modelo Support Vector Machine é baseada na identificação da melhor fronteira entre as classes. Como pode-se observar na Figura 2.6, quando um modelo é linearmente separável, existem várias retas que podem ser traçadas para separar as classes.

O SVM utiliza uma métrica para determinar qual é a melhor reta, otimizando o que podemos chamar de margem, que é a separação entre as classes. O SVM ajusta essa reta de forma a maximizar a margem, ou seja, a distância entre a reta e os pontos mais

próximos de ambas as classes. Esses pontos mais próximos são conhecidos como Vetores de Suporte. Na Figura 2.7, a margem M é definida como a distância do hiperplano (reta) aos Vetores de Suporte das duas classes. O objetivo do SVM é encontrar a reta que maximize essa margem, garantindo que a separação entre as classes seja a maior possível [1].

Se as duas classes não são linearmente separáveis, o SVM pode utilizar os SVM-Kernels, permitem ao SVM determinar a separação entre as amostras usando funções que não são retas, aumentando a flexibilidade do modelo. Alternativamente, o SVM pode procurar um hiperplano no espaço original que maximize a margem entre as classes, ao mesmo tempo em que minimiza uma quantidade proporcional ao número de falhas na margem, conhecidas como variáveis de folga. Este método, chamado SVM com margem suave, permite que algumas amostras estejam no lado errado do hiperplano ou dentro da margem, equilibrando a maximização da margem e a minimização dos erros de classificação [26].

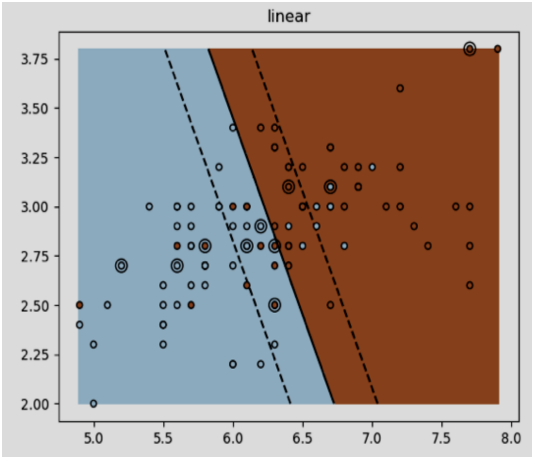

Figura 2.6: Vetores de suporte. Fonte: Scikit Learn

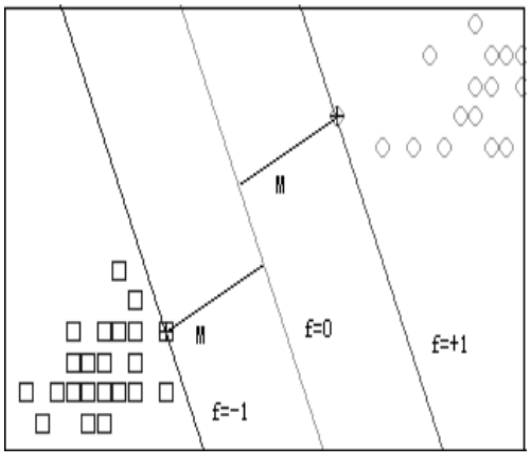

Figura 2.7: vetores de suporte [1]

2.5.3 Random Forest

O principal objetivo do algoritmo Random Forest quando se trata de problemas de classificação, é melhorar a precisão das previsões realizadas pelo algoritmo, ele faz isso construindo uma floresta de muitas árvores de decisão independentes durante o treinamento e combinando suas previsões para obter uma única previsão [24].

O Random Forest utiliza a simplicidade e flexibilidade das árvores de decisão, resultando em uma grande melhoria na precisão dos dados de saída. Pode-se resumir a execução do algoritmo em dois passos principais [27]:

- Passo 1: é criar um conjunto de dados amostrado (bootstrapped dataset), do mesmo tamanho que o original, selecionamos aleatoriamente amostras do conjunto de dados original, permitindo que uma mesma amostra seja escolhida mais de uma vez.
- Passo 2: Criar uma árvore de decisão usando o conjunto de dados amostrado (X). Inicialmente para criar a raiz da arvore o algoritmo utiliza um subconjunto aleatório de variáveis, selecionando a variável que melhor separa as amostras. Então se repete os mesmos passos para os outros n nós, selecionando aleatoriamente as variáveis e construindo a árvore de decisões (Figura 2.8), até termos a primeira árvore completa.

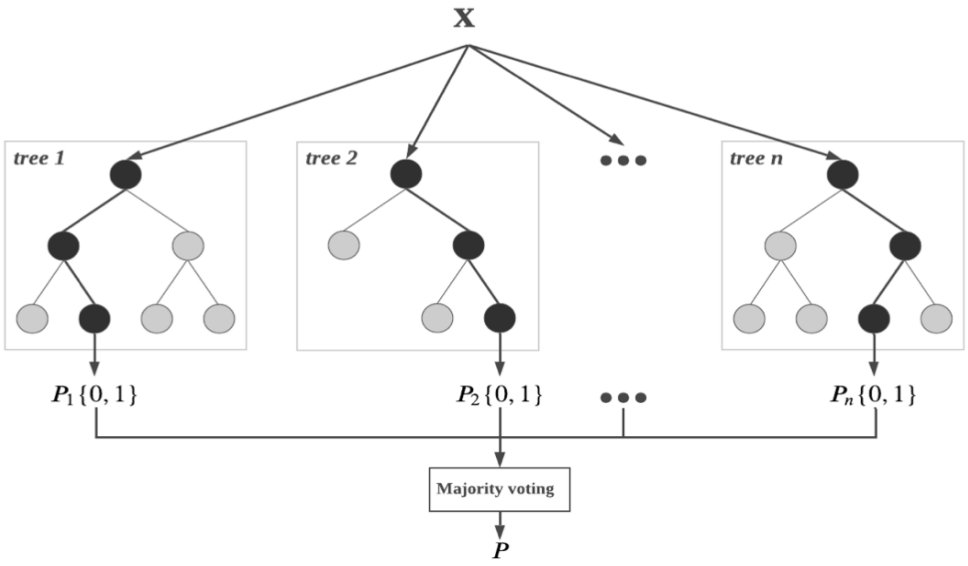

Figura 2.8: Arvore de decisões Randon Forest. [2]

O algoritmo repete esses passos centenas de vezes, resultando em uma grande variedade de árvores. Essa variedade é o que torna o Random Forest mais eficaz do que árvores de decisão individuais. Para usar a floresta de decisões que foi criada, o algoritmo pega um novo conjunto de dados de teste e usa todas as árvores de decisão para classificar esse novo conjunto de dados de teste [2].

Por fim, utilizamos um algoritmo que acessa o dado que foi treinado e armazenado na etapa de classificação do pipeline, e obtém as importâncias das classes do dataframe, extraindo e combinando os nomes das classes categóricas e numéricas. Depois, cria um novo *Data Frame* que associa cada classe à sua importância, ordenando-as da mais importante para a menos importante, destacando as features que mais influenciam as previsões do modelo.

Capítulo 3

Metodologia

Neste capítulo é apresentada a metodologia utilizada para análise dos dados educacionais do curso de engenharia de computação da UnB. Esta análise tem duas partes: Análise descritiva e Aplicação de algoritmos de aprendizado de máquina.

3.1 Descrição da Metodologia

A Figura 3.1 apresenta os passos realizados para realização da análise descritiva que iniciou-se com a seleção dos dados brutos, identificando e extraindo os dados mais relevantes para o estudo (Curso de Engenharia de Computação) tendo como as fontes dos dados a base do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) e do Sistema de Informações Gerenciais de Registros Acadêmicos (SIGRA), sistemas acadêmicos da Universidade de Brasília durante o período da análise dos dados. Em seguida, foi realizado o processo de ETL, onde os dados selecionados foram processados e transformados no formato mais adequado para análise, garantindo o entendimento das informações, gerando uma base de dados tratados. A etapa de transformação envolveu a conversão dos dados tratados em visualizações gráficas, facilitando a interpretação dos resultados.

Também foi realizado a aplicação de algoritmos de aprendizado de máquina para entender os padrões relacionados aos motivos que podem contribuir com a desistência dos alunos. Utilizando a linguagem Python foi realizada uma nova filtragem nos dados tratados, essa filtragem teve como objetivo isolar informações mais específicas e de interesse para análise da evasão, em seguida as variáveis foram fornecidas aos algoritmos de predição (*Gradient Boosting Machine, Random Forest, e Support Vector Machine*). Dessa forma, foi possível observar como cada variável escolhida influenciou a evasão dos alunos.

Por fim, foi conduzida a análise dos dados, onde as informações processadas foram examinadas para possibilitar e fundamentar as conclusões deste trabalho.

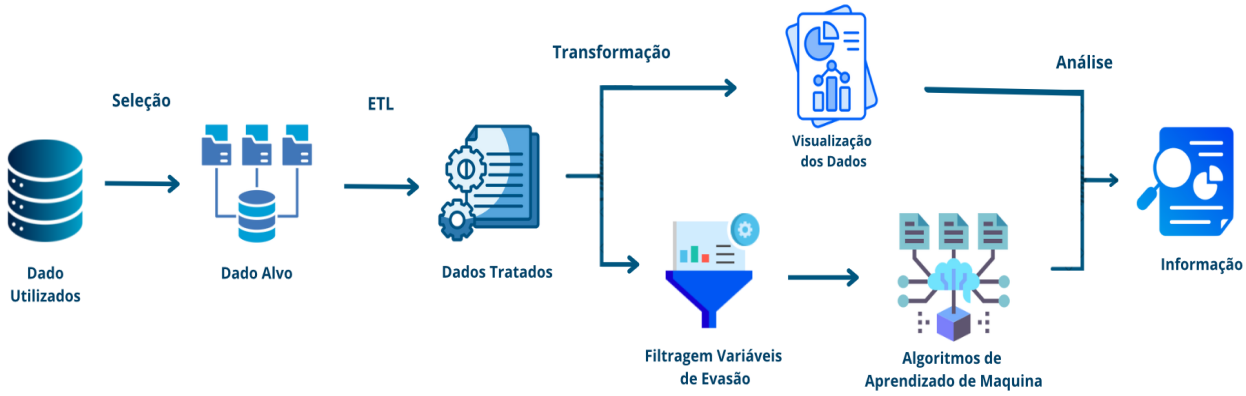

Figura 3.1: Metodologia do projeto.

3.2 Dados Utilizados

Os dados utilizados neste trabalho foram extraídos de duas bases diferentes: SIGAA e SIGRA. Os dados das bases utilizadas foram extraídas em arquivos .xlsx que é o formato de arquivo padrão usado pelo Microsoft Excel. Em ambas as bases encontram-se informações gerais sobre o histórico do aluno, bem como informações sobre o período em que o aluno cursou as matérias, ano que ingressou no curso e ano que saiu do curso.

Originalmente o dado bruto possuía 32 colunas com informações gerais do aluno. Para fins de uma análise mais direcionada, aplicou-se o processo de ETL nos dados para deixar apenas os dados referentes ao curso de Engenharia da Computação. Após o ETL, o *dataset* ficou com 19 colunas contendo as variáveis que seriam analisadas, sendo elas:

- aluno: código do aluno no sistema;
- id_pessoa: id único do aluno utilizados para relacionar as bases utilizadas;
- IRA: Índice de Rendimento Acadêmico;
- gênero: informa qual o gênero do aluno (Feminino ou Masculino);
- nascimento: data do nascimento do aluno;
- estado_nascimento: Sigla da Unidade Federativa (UF) em que o aluno nasceu;
- sistema_cotas: informa por qual modalidade de cotas o aluno ingressou:

  Candidato Negro; Escola Pública Baixa Renda - Não PPI; Escola Pública Alta Rena - Não PPI; Escola Pública Baixa Renda - PPI; Escola Pública Baixa Renda - Não PPI; Escola Pública Baixa Renda Não PPI - PCD
- cota: informa se o aluno entrou por cota ou não (sim ou não);
- raça: raça do aluno;

Branco; Pardo; Preta; Amarelo (de origem oriental); Não Cadastrada; Não Informado

- segundo_grau_tipo_escola: tipo de escola de ensino médio do aluno (Pública ou Particular);
- curso: nome do curso que o aluno estava cursando, neste estudo apénas Engenharia de Computação;
- periodo_ingressou_unb: semestre e ano referente ao ingresso do aluno na UnB;
- forma_ingresso_unb: forma pela qual o aluno ingressou na UnB:

  Vestibula; Programa De Avaliação Seriada (PAS); SISU - Sistema De Selecão Unificada; Portador de Diploma de Cuso Superior; Mudança de Curso; Transferência Obrigatória; Transferência Facultativa; ENEM - UnB; Matrícula Cortesia
- periodo_saida_curso: semestre e ano referente a saída do aluno da UnB;
- forma_saida_curso: o motivo informado no momento da saída:

  Abandono (Nenhuma Matrícula); Efetivação de Novo Cadastro; Solicitação Espontânea; Desligamento - não cumpriu condição; Mudança de Curso; Reprovar 3 vezes na mesma disciplina obrigatoria; Novo Vestibular; Desligamento Falta Documentação; Falecimento; Trans. P/ outra IES; Integralização de Discente
- periodo_cursou_disciplina: semestre e ano em que o aluno cursou cada disciplina;
- modalidade_disciplina: informação se a disciplina é obrigatória ou optativa;
- nome_disciplina: informação do nome da disciplina cursada;
- creditos_disciplina: quantidade de créditos das respectivas matérias cursadas;
- mencao_disciplina: a menção final da disciplina cursada;

  SS (Superior); MS (Médio Superior); MM (Médio); MI (Médio Inferior); II (Inferior); SR (Sem Rendimento)

Com o *dataset* já tratada, os dados foram convertidos de um data frame no python para um arquivo CSV para então ser importada no Power BI.

3.3 Tratando os Dados

Grande parte dos dados foi processada e tratada em ferramentas de programação e manipulação de dados. O Google Colab, ou Colaboratory[1], é uma ferramenta desenvolvida e disponibilizada pelo Google que permite a escrita e execução de códigos de forma virtual

[1] https://colab.google

através do navegador. Essa ferramenta oferece um ambiente online onde é possível criar notebooks nos quais se pode executar células em Python, o que se mostrou ideal para realizar o processo de ETL dos dados recebidos.

Os dados foram tratados utilizando o Python que é uma linguagem interpretada, dinâmica e funcional, de código aberto, com sintaxe semelhante à língua inglesa, que vem ganhando popularidade no meio da tecnologia [28]. Uma linguagem diversificada, incluindo desenvolvimento web simples até a ciência de dados, essa linguagem possuem uma sintaxe um pouco mais clara, por se tratar de uma linguagem de alto nível, o que facilita a leitura do código, aumentando a produtividade do programador [29].

Os processos de filtragem dos dados utilizando o python foram: Renomeação de colunas tendo como base as colunas dos dados do SIGAA, retirada de dados duplicados, retirada de algumas colunas do *data frame*, refatoração de abreviações e nomeclaturas, identificar e tratar os campos nulos, utilizacao das bibliotecas Pandas e Numpy para manipulacao e organizacao dos dados e separar informações uteis em grupos.

Dentre as várias funções do Colab, a integração com o Google Drive[2] foi particularmente útil para a transformação e carregamento dos dados de tratados. A base de dados era grande, e a capacidade de salvar e carregar notebooks do Colab diretamente do Google Drive facilitou o processamento dos dados. Isso permitiu que os *datasets* fossem armazenados e acessados de maneira mais fácil e rápida.

3.4 Análise Visual Descritiva

Para visualização dos dados foi utilizado a ferramenta Microsoft Power BI[3], uma plataforma de *business intelligence* (BI) desenvolvida pela Microsoft[4] que permite a visualização de dados e a criação de relatórios em formato *dashboard*. Essa ferramenta foi utilizada para transformar dados brutos coletados da base em informações e gráficos analisados neste trabalho. A finalidade principal do Uso do Power BI neste trabalho foi permitir a análise de dados de maneira prática, possibilitando principalmente:

1. Conectar-se a fontes de dados resultantes do tratamento dos *datasets* (arquivo CSV).
2. Transformar e limpar a base recebida, utilizando apenas os dados do curso de Engenharia de Computação.
3. Criar visualizações gráficas.
4. Analisar e compartilhar os dados extraídos.

[2]https://drive.google.com/drive
[3]https://www.microsoft.com/power-bi
[4]https://www.microsoft.com/

Além disso, vale ressaltar que foram identificadas muitas vantagens em utilizar essa ferramenta, pois com o Power BI tem-se a possibilidade de conexão e integração com outras ferramentas online e local. A ferramente possibilita também a modelagem dos dados necessárias para organizar grandes massas de dados em tabelas de fácil compreensão e análise. Nesse mesmo sentido, o Power BI facilita a criação de *dashboards* e relatórios através de recursos de arrastar e soltar (*drag and drop)* tornando a experiencia do usuário simples. Outra grande vantagem identificada é a atualização em tempo real, onde os *dashboards* e relatórios são automaticamente atualizados conforme as fontes de dados são modificadas [30].

Após a importação e tratamento dos dados no Power BI, foi necessário realizar manipulações adicionais para obter informações mais detalhadas, que não haviam sido previamente aplicadas. Para esse processo final, utilizou-se a linguagem DAX (Data Analysis Expressions), uma linguagem de fórmulas do Power BI, que permite a criação de cálculos avançados e a exploração mais aprofundada dos dados.[31].

3.5 Algoritmos de Aprendizado de Máquina

Nesta seção, é descrita a metodologia utilizada para a execução da análise preditiva, baseada no modelo CRISP-DM (*Cross Industry Standard Process for Data Mining*) [32]. Este modelo, amplamente aplicado em projetos de análise de dados, segue seis etapas principais:

1. Entendimento do Negócio: A Universidade de Brasília, fundada em 1962, é um símbolo de transformação e modernização no cenário educacional brasileiro. Idealizada pelo antropólogo Darcy Ribeiro, a universidade foi projetada para ser um ambiente de autonomia acadêmica, incentivando o desenvolvimento livre do pensamento científico e cultural. Nos primeiros anos, a UnB enfrentou grandes obstáculos, principalmente durante o período da ditadura militar, mas superou essas adversidades e se estabeleceu como uma das instituições mais respeitadas do país. Atualmente, o campus Darcy Ribeiro, localizado na Asa Norte, é o principal e maior da universidade [33]. A UnB oferece uma ampla variedade de programas de graduação e pós-graduação e se destaca por seus projetos que abordam questões sociais e ambientais. Além disso, a universidade tem uma forte presença em pesquisa, inovação e extensão, consolidando-se como uma referência tanto no Brasil quanto no exterior [33]. Segundo Assunção e Nogueira, antes da implementação da Lei de Cotas (Lei 12.711/2012), as universidades públicas brasileiras refletiam uma realidade de exclusão de grupos sociais historicamente marginalizados, como afrodescendentes e indígenas. A baixa representatividade desses grupos era evidente: Em 2002, apesar

de 45% da população brasileira ser negra, apenas 2% desses indivíduos alcançavam o ensino superior [8]. A Lei de Cotas, implementada a partir de 2013, buscou corrigir essa desigualdade, reservando vagas para estudantes de escolas públicas, com uma atenção especial para negros, indígenas e pessoas de baixa renda [7]. No contexto da UnB, o estudo sobre a eficácia e os impactos dessa política no curso de Engenharia de Computação é de suma importância, pois permite avaliar como essas ações afirmativas contribuem para a transformação do cenário acadêmico e social na universidade, promovendo uma educação mais inclusiva e equitativa.

2. Entendimento dos Dados: A descrição dos dados utilizados foi feita da Seção 3.2.

3. Preparação dos Dados: O tratamento e preparação dos dados utilizados foi descrito na Seção 3.3 deste trabalho.

4. Modelagem: Para treinar os modelos de predição escolhidos, realizou-se um filtro nos dados tratados, selecionando as principais variáveis (*features*) que seriam investigadas como possíveis fatores que poderiam influenciar na conclusão ou evasão do curso, para aplicar os algoritmos foi necessário tratar os dados novamente. Este tratamento final realizado com a linguagem Python, resultou em uma base com 358 alunos, apenas alunos que ingressaram a partir do início do sistema de cotas pela Lei das Cotas, que não estavam ativos, ou seja, ja haviam saído da universidade.

   O intervalo de tempo selecionado para o recorte dos dados utilizados foram os semestres 01/2013 (ano em que a Lei das Cotas entrou em vigor) até o semestre 02/2022, dessa forma tornaríamos o dataset mais preciso em relação a analises especificas acerca do período em que as cotas entraram em vigor. As variáveis utilizadas pelo algoritmo de treinamento e suas respectivas porcentagens foram:

   - id_pessoa: Código único de identificação do aluno.

     Total de aluno : 358 alunos.

     Evasões: 229 alunos.

     Conclusões: 129 alunos.

     Porcentagem de Evasão: 63,96%

     Para esta medida foi considerado os seguintos motivos de evasão: Abandono (Nenhuma Matrícula), Efetivação De Novo Cadastro, Solicitação Espontânea, Desligamento - Não Cumpriu Condição, Mudança De Curso, Reprovação 3 Vezes Na Mesma Disciplina Obrigatória, Novo Vestibular, Desligamento Por Falta De Documentação, Falecimento e Transferência Para Outra IES
   - IRA: Índice de Rendimento Acadêmico do aluno (Tabela 3.1).

- sexo: Sexo do aluno. Para este trabalho foi utilizado os sexos masculino e feminino (Tabela 3.3).
- idade: Idade do aluno. Para facilitar os resultados das previsões separamos em grupos de idade: Até 17, 18 a 20, 21 a 25, 26 a 30, Mais de 30 (Tabela 3.8).
- estado_nascimento: Para fins de análise a UF foi separada em DF ou Fora do DF (Tabela 3.4).
- Cota: Tipo de cota do aluno. Para os algoritmos de predição os tipos de cotas foram agrupados da seguinte forma: Escola Pública Alta Renda-PPI, Escola Púb. Alta Renda-Não PPI, Escola Púb Alta Renda-PPI-PCD, foram consideradas Escola Pública Alta Renada. Escola Pública Baixa Renda-PPI e Escola Púb Baixa Renda-Não PPI, foram consideradas Escola Pública Baixa Renda. Canditado Negro foi considerado como Negros. Os demais registros foram considerados como Universal (Tabela 3.5).
- segundo_grau_tipo_escola: Tipo de escola de formatura do aluno (pública ou particular) (Tabela 3.2).
- curso: Curso do aluno. Apenas foram selecionados os alunos de Engenharia de Computação.
- forma_ingresso_unb: Tipo de prova feita para ingresso na UnB. Para facilitar os resultados das previsões separamos a forma de ingresso em grupos principais: Vestibular, PAS e ENEM (SISU e ENEM - UnB). Os motivos considerados como "OUTROS"são: Transferência Facultativa, Portador de Diploma de Curso Superior, Transferência Obrigatória, Mudança de Curso e Matrícula Cortesia (Tabela 3.6).
- APC: Quantidade de vezes que o aluno realizou a matéria APC. Apenas foram selecionados os alunos que pegaram a matéria entre 1 a 4 vezes (Tabela 3.7).
- forma_saida_curso: Aluno concluiu ou evadiu ao sair da UnB.

Na realização do treinamento dos modelos separou-se a variável 'forma_saida_curso', como era a variável que seria prevista e utilizou-se as demais variáveis do *data frame* para os testes. Nos testes, 20% dos dados foram usados, e 80% para o treino.

5. Avaliação: A confiabilidade dos dados extraídos, tratados e analisados das bases do SIGAA e do SIGRA foi validada pela acurácia dos modelos de treinamento, que é calculada pela divisão do número de predições corretas pelo total de predições.

6. Implantação:

Tabela 3.1: Porcentagem de Alunos por Faixa de IRA

Faixa de IRA	Porcentagem	Total Alunos
0-1	16,20%	58
1-2	17,04%	61
2-3	16,76%	60
3-4	29,05%	104
4-5	20,95%	75

Tabela 3.2: Porcentagem de Alunos por Tipo de Escola

Escola	Porcentagem	Total
Pública	38,38%	139
Particular	61,62%	219

A análise e o estudo dos dados apresentados neste trabalho geraram informações úteis, assim como orientações para ações futuras que foram detalhadas no Capítulo 5.

Tabela 3.3: Porcentagem por tipo de Sexo.

Tipo	Porcentagem	Total Alunos
Masculino	90%	324
Feminino	10%	34

Tabela 3.4: Porcentagem de Alunos por Estado de Nascimento

Estado	Porcentagem	Total
DF	32,32%	242
Fora do DF	67,68%	116

Tabela 3.5: Porcentagem por tipo de Cota.

Tipo	Porcentagem	Total Alunos
Pública Baixa Renada	10,31%	38
Negros	6,19%	24
Pública Alta Renda	18,56%	67
Universal	60,95%	229

Tabela 3.6: Porcentagem por tipo de Ingresso.

Tipo de Ingresso	Porcentagem	Total Alunos
Vestibulas	45,45%	162
PAS	30,30%	108
ENEM	6,06%	22
Outros	18,18%	66

Tabela 3.7: Porcentagem de vezes que cursou APC

Tipo de Ingresso	Porcentagem	Total Alunos
1 vez	44,90%	161
2 vezes	42,86%	152
3 vezes	8,16%	30
4 vezes	4,08%	15

Tabela 3.8: Porcentagem de Idade

Tipo de Ingresso	Porcentagem	Total Alunos
18 a 20	77,32%	271
até 17	10,31%	37
21 a 25	8,25%	32
26 a 30	2.06%	9
Mais de 30	2.06%	9

Capítulo 4

Desenvolvimento e Análise de Resultados

Neste capítulo, são apresentados os resultados obtidos a partir da análise do desempenho acadêmico dos alunos ao longo de diferentes períodos. A análise incluiu dados sobre o ingresso dos alunos, a identificação das disciplinas com maiores taxas de reprovação, o desempenho de alunos cotistas e não cotistas e a visualização dos dados analisados, bem como as possíveis influências do sexo e do tipo de escola no ambiente acadêmico.

4.1 Resultados e Análises da Entrada de Alunos no Curso de Engenharia da Computação

Um dos objetivos centrais deste trabalho foi identificar a quantidade de alunos ingressantes no curso de Engenharia de Computação para observar os efeitos da Lei 12.711/2012. A Lei das Cotas determina que as universidades federais deveriam reservar 50% das vagas oferecidas para os cursos de nível superior a estudantes que cursaram todo o ensino médio em instituições de ensino público [8]. A lei entrou em vigor no primeiro semestre de 2013 iniciando com 12,5% das vagas reservadas para cotas, atingindo 50% das vagas reservaas em 2016. A Figura 4.1 apresenta o crescimento gradual da entrada de alunos cotistas a partir do início da implementação da Lei em 2013 na UnB, Embora o curso tenha dado início na UnB no segundo semestre de 2009, para analisar os números referente aos ingressos dos alunos, foram investigados os dados de um ano completo (2 semestres), iniciando em 2010 até 2022. Observa-se uma convergência no gráfico a partir do ano de 2013, ano em que a Lei 12.711/2012 entrou em vigor.

O comportamento do gráfico nos semestres anteriores a 2013 é diferente do comportamento observado após esse mesmo ano. A linha azul escura, que representa os alunos

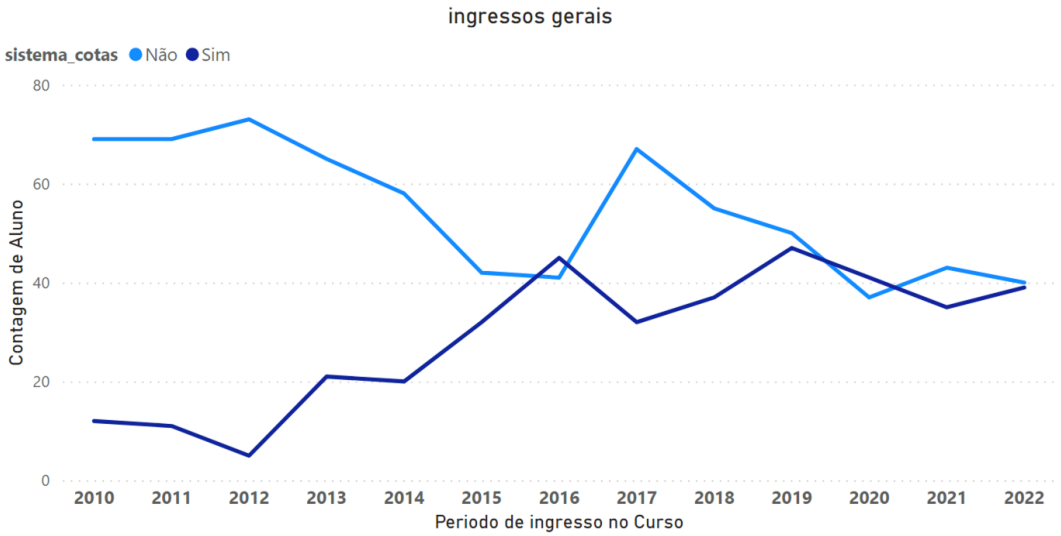

Figura 4.1: Ingressos de alunos cotistas no curso de Engenharia de Computação

cotistas, estava sempre muito abaixo da linha azul clara, que representa os alunos não cotistas. Este fato deve-se à reserva de 20% das vagas do curso para alunos negros desde o segundo semestre de 2004 [7].

Após o ano de 2013, a mudança no cenário de entrada de alunos no curso é principalmente atribuída à Lei da cotas, que permitiu que não só alunos negros, mas também alunos de baixa renda e indígenas tivessem a reserva de 50% das vagas oferecidas, garantindo o ingresso desses grupos no curso.

Em 2017, o gráfico apresenta um número elevado de alunos não cotistas ingressando no curso. Esse pico foi resultado de uma entrada atípica de alunos por meio de diferentes modalidades: transferências obrigatórias e facultativas (4 alunos), mudança de curso (2 alunos), matrícula cortesia (1 aluno), ingresso com diploma de curso superior (8 alunos) e pelo Sistema de Seleção Unificada (6 alunos). Por padrão estes motivos são considerados não cotistas nos sistemas da UnB. A quantidade de alunos ingressantes por Vestibular e Programa de Avaliação Seriada foi de 40 alunos ao total.

4.1.1 Entrada de Alunos por Tipo de Escola

Na Figura 4.2 observa-se uma mudança significativa nos ingressos de alunos provenientes de escolas públicas. Nos anos anteriores a 2013, há uma discrepância nos registros de entradas entre alunos de escolas particulares e públicas. A partir de 2013 observa-se uma tendência de crescimento da entrada de alunos de escolas públicas na UnB. Nota-se que as quantidades de ingresso por tipo de escola praticamente se igualam a partir do ano de 2015, diferente da Figura 4.1 que se refere aos ingressos por tipo de cota, onde esse fenômeno ocorreu apenas no ano seguinte, em 2016. Após o ano de 2019, a informação

do tipo de escola do aluno deixou de ser obrigatória, justificando a grande quantidade de alunos que optaram por não informar o tipo de escola a partir de 2020, como mostra o gráfico.

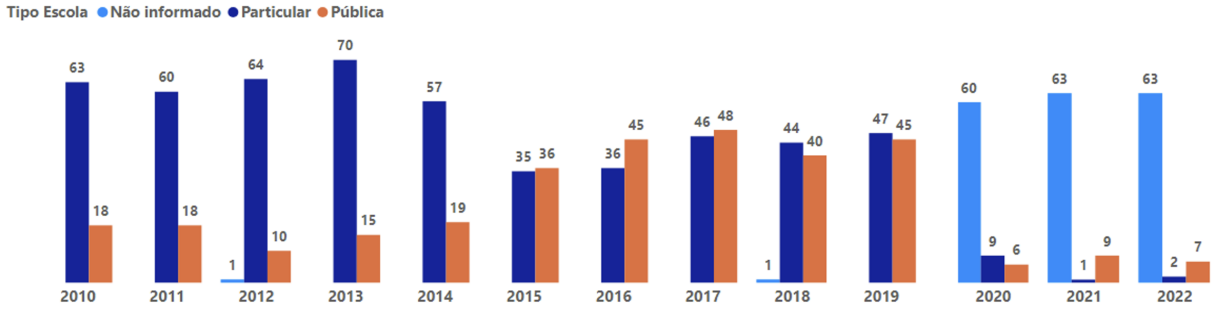

Figura 4.2: Ingressos por tipo de escola no curso de Engenharia de computação

Diante dos dados apresentados sobre os ingressos de alunos por tipo de escola, fica evidente que o contexto acadêmico do curso de Engenharia de Computação era predominantemente composto por alunos de escolas particulares. Entre os anos de 2010 e 2012, a UnB apresentou uma média de 19,61% de alunos provenientes de escolas públicas e 80,39% de alunos de escolas particulares. Especificamente, em 2010 e 2011, o percentual de alunos de escolas públicas foi de 22,22% e 23,08%, respectivamente, enquanto em 2012 houve uma queda para 13,51%. Neste mesmo contexto de desigualdade, ainda existem críticas governamentais à proposta das cotas e alegações negativas sobre a qualidade do ensino nas universidades públicas após a implementação da lei das cotas [9] [34].

No entanto, felizmente, esse contexto tem mudado significativamente nos últimos anos. Com a implementação da Lei de Cotas, observa-se um aumento considerável no número de alunos provenientes de escolas públicas e alunos PPI, o que tem impactado positivamente a realidade de muitos estudantes nas universidades públicas do país.

4.1.2 Entrada de Alunos por Tipo de Cota

Observa-se no gráfico da Figura 4.3 o primeiro registro de entrada de alunos beneficiados pela Lei 12.711/2012 em 2013, seguido por um crescimento gradual no número de alunos cotistas a partir desse período, evidenciando o impacto da Lei das Cotas.

Com a entrada da Lei de Cotas, o número de alunos PPI aumenta a partir de 2013, com um destaque evidente em 2016, onde ve-se um salto considerável de ingressos.

No mesmo período, observa-se que os alunos não PPI, especialmente os de baixa renda, também foram impactados pela Lei de Cotas, mas em menor proporção. Grupos como "Escola Pública Baixa Renda-Não PPI"e "Escola Pública Alta Renda-Não PPI"mantiveram uma participação moderada, mas não tiveram o mesmo aumento expressivo que os alunos PPI.

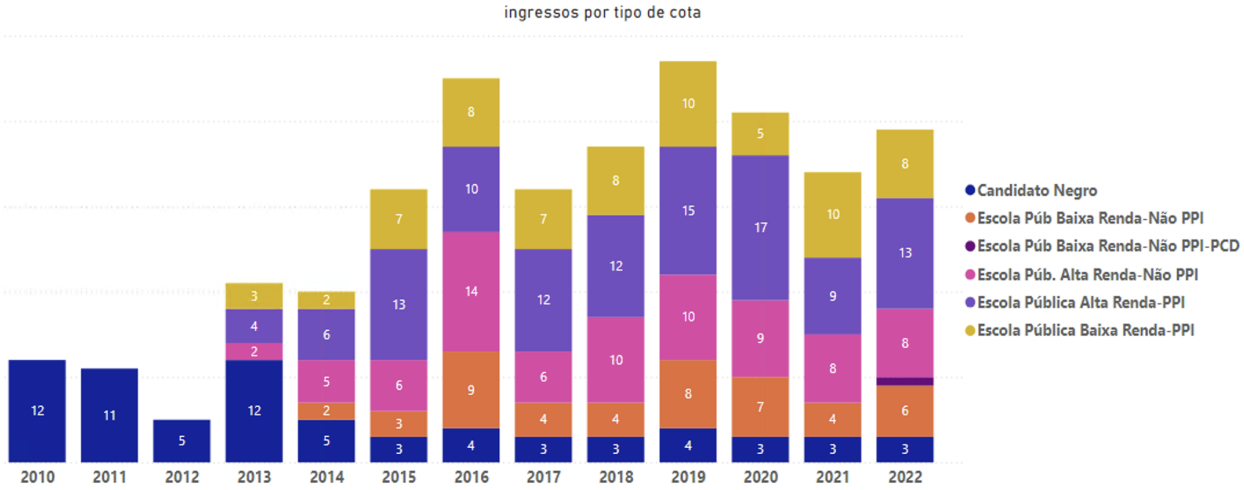

Figura 4.3: Ingressos por tipo de cota no curso de Engenharia de computação

O número de alunos PPI de alta renda se mantém relativamente baixo em comparação com os de baixa renda. Isso reflete a realidade socioeconômica, na qual a maioria dos estudantes PPI que acessam a universidade por meio das cotas são de baixa renda. O número de alunos não PPI de alta renda também apresenta uma participação pequena, mas constante ao longo dos anos.

Assim, o gráfico reflete claramente o efeito da Lei de Cotas no aumento da diversidade socioeconômica e racial no curso de Engenharia de Computação. A implementação das cotas resultou em um aumento expressivo de estudantes PPI e de baixa renda, especialmente os vindos de escolas públicas, o que demonstra a efetividade da lei em ampliar o acesso desses grupos ao ensino superior. Ainda que os alunos cotistas de alta renda, tanto PPI quanto não PPI, tenham uma presença menor, destaca-se a importância de politicas que promovam a equidade no acesso ao ensino superior a todos.

4.1.3 Entrada de Alunos por Sexo

Para analisar a entrada de alunos no curso considerando o sexo, levantamos dados referentes aos anos de 2010 a 2022.

Podemos observar na Figura 4.4 que, em sua maioria, a quantidade de ingressos dos alunos do sexo masculino é significativamente maior. Ao analisar os ingressos no curso de Engenharia de Computação entre 2010 e 2022, observa-se que, em média, 11,87% dos ingressantes foram alunas, enquanto 88,13% foram alunos. No ano de 2010, a porcentagem de alunas foi 18,52%, destacando-se como uma das mais altas do período, com 81,48% de alunos do sexo masculino. Já em 2015, o ano com o menor número de ingressantes do sexo feminino, a porcentagem foi de 8,11% para alunas e 91,89% para alunos, reforçando a predominância masculina ao longo dos anos.

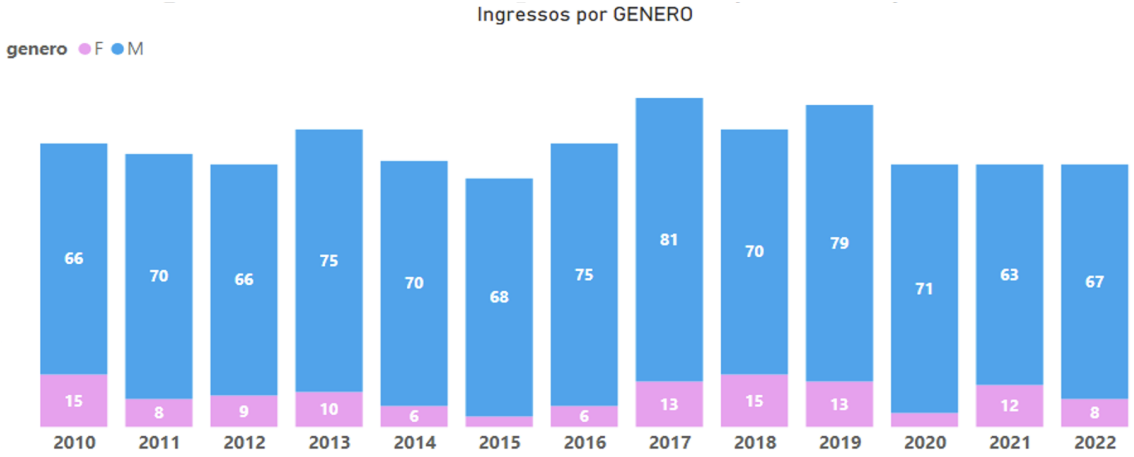

Figura 4.4: Ingressos por tipo de escola no curso de Engenharia de computação

4.2 Resultados e Análise do Desempenho do Aluno Durante o Curso

Para avaliar o desempenho dos alunos do curso de Engenharia de Computação, foi realizada uma análise dos índices de aprovação e reprovação nas disciplinas iniciais do curso, abrangendo os três primeiros semestres. Nesta seção, foram examinados os resultados dos alunos nesses semestres, com ênfase nos seguintes pontos:

- Desempenho dos alunos por sexo e por matéria (aprovações e reprovações): Para esta análise, foram destacados os dados a partir dos semestres de 2010. Dessa forma, foi possível identificar e analisar especificamente os dados do grupo alcançado pela Lei 12.711/2012. O mesmo período de tempo foi aplicado na análise do desempenho dos alunos na matéria Algoritmos e Programação de Computadores (Seção 4.3).
- Desempenho dos alunos por sistema de cotas: Para garantir um resultado devidamente equilibrado, especificamente nas seções referentes aos alunos cotistas e não cotistas, foram utilizados dados a partir do ano 2016, pois foi nesse período que o número de alunos cotistas e não cotistas se igualou em termos de ingresso na UnB.(Figura 4.1).

4.2.1 Análise de Desempenho nas Disciplinas Iniciais do Curso

Nos primeiros semestres do curso de Engenharia de Computação, os alunos cursam matérias nas áreas de matemática, física e programação. No primeiro semestre, são ofertadas seis disciplinas: Algoritmos e Programação de Computadores, Cálculo 1, Física 1, Física 1 - Experimental, Introdução à Álgebra Linear e Introdução à Engenharia de Computação.

No segundo semestre, cinco disciplinas são ofertadas: Cálculo 2, Probabilidade e Estatística, Física 2, Estrutura de Dados e Física 2 - Experimental. No terceiro semestre, os alunos estudam Cálculo 3, Sinais e Sistemas, Métodos de Programação e Sinais e Sistemas em Tempo Discreto.

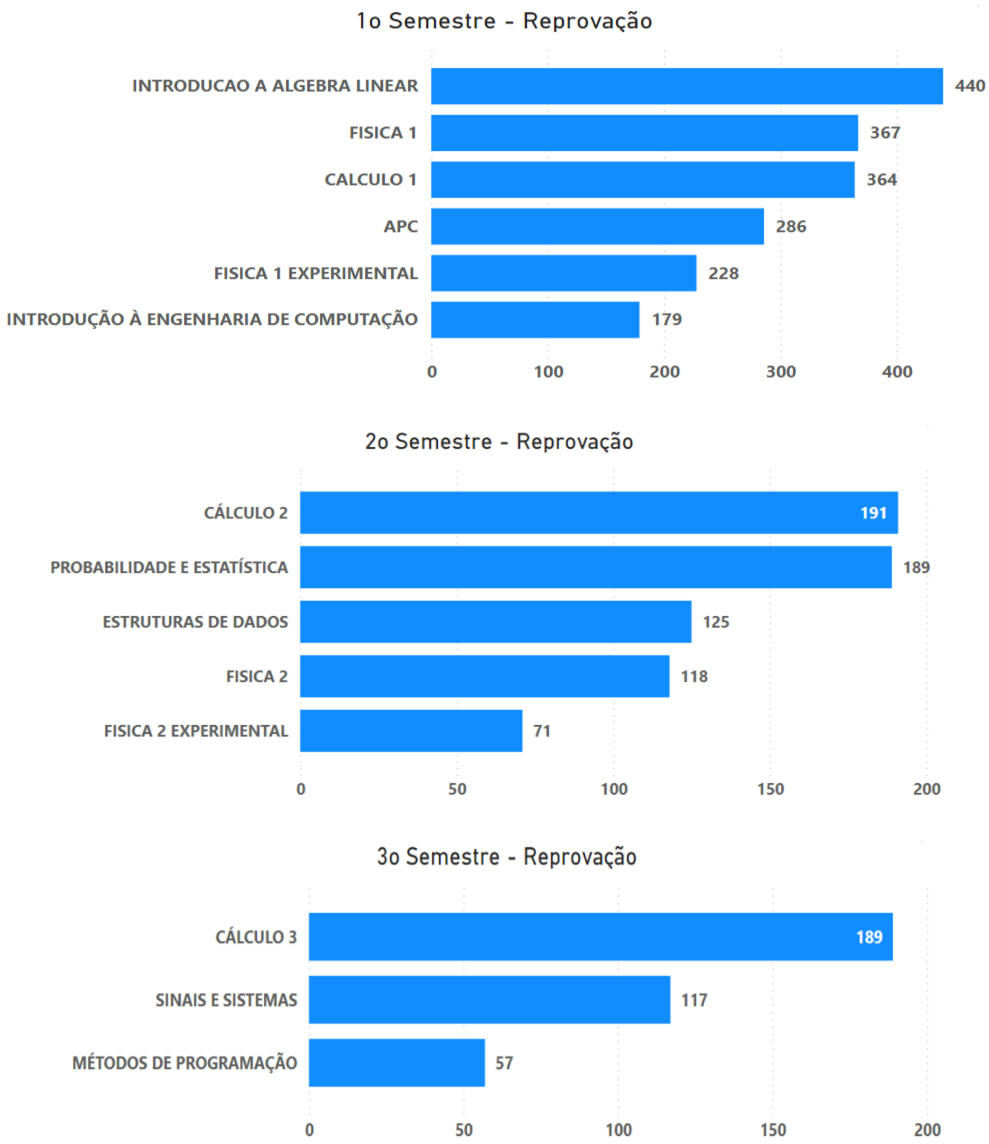

Figura 4.5: Quantidades de reprovações nos primeiros 3 semestres

Na UnB, o sistema de avaliação utiliza menções para refletir o desempenho dos alunos. As siglas representam diferentes níveis de desempenho: SS (Superior) indica o mais alto rendimento; MS (Médio Superior) para desempenho acima da média; MM (Médio) para o suficiente; MI (Médio Inferior) para um desempenho abaixo da média; II (Inferior) para insuficiente, sem aprovação; e SR (Sem Rendimento) para casos de abandono ou falta de comparecimento.

Para analisar o desempenho dos alunos nos primeiros semestres, e identificar as matérias com maiores números de reprovações, foram examinadas as disciplinas iniciais do curso, levando em consideração a quantidade total de aprovações e reprovações, sendo considerados a contagem de menções MM, MS e SS como aprovações e a contagem de menções SR, II e MI como reprovações. Retirada de matrícula e tracamento de semestre não foram considerados para as análises desta seção.

A Figura 4.5 percebe-se que a matéria Física 1 se destacou no contexto de reprovações, sendo a segunda disciplina que mais reprova os alunos do primeiro semestre, ficando atrás somente da disciplina Introdução a Álgebra Linear, que possui um número significativo de reprovações nesse período.

No contexto do segundo semestre, as disciplinas Cálculo 2 e Probabilidade e Estatística apresentam números de reprovações muito similares, indicando um provável nível de dificuldade alto de ambas as matérias e revelando o grande número de alunos reprovados nesse semestre.

De forma similar, a disciplina Cálculo 3 também exibe um grande número de reprovações em comparação com as outras duas disciplinas presentes no currículo do terceiro semestre do curso. Por possuir uma ementa extensa e complexa, Cálculo 3 apresenta quase o dobro de reprovações em relação a Sinais e Sistemas, que é a segunda disciplina com maior número de reprovações no terceiro semestre.

4.2.2 Análise de Desempenho por Tipo de Cota

Para analisar o desempenho dos alunos cotistas e não cotistas ao longo do curso, nesta seção foram considerados os alunos que ingressaram a partir do ano de 2016 (ano em que a Lei das Cotas atingiu 50% das vagas reservadas) e as matérias iniciais analisadas neste trabalho, que apresentaram os maiores índices de reprovação, conforme mostrado na Seção 4.2.1.

Na Figura 4.6 foi contabilizada a soma das menções MI, II e SR (média de nota final inferior a cinco pontos, considerada reprovação) referentes às matérias apresentadas no gráfico. Observando a Figura 4.6, pode-se notar que, na maioria dos casos, as reprovações são mais frequentes entre os alunos cotistas. Das matérias destacadas nesta imagem, apenas em 2 matérias (Sinais e Sistemas e Métodos de Programação) os alunos não cotistas apresentam mais reprovações que os alunos cotistas.

Em Física 2, pode-se observar que o número de reprovações foi o mesmo para ambos os grupos. Observado os valores de reprovações das outras matérias, inferimos que mesmo com o número semelhante de ingressos de alunos cotistas em 2016, no primeiro semestre, a dificuldade é maior para os alunos que entram por meio das cotas. Este grupo ainda apresenta dificuldades nas matérias do primeiro semestre, e somente no segundo semestre

esses alunos conseguem alcançar números menores de reprovações, semelhantes ao grupo de alunos não cotistas em matérias como Cálculo 2, Física 2 e Probabilidade e Estatística. Isso indica uma tendência positiva (quantidade menor de reprovações pode indicar um aumento na quantidade de aprovações) em relação ao desempenho de ambos os grupos, cotistas e não cotsitas, ao longo do curso.

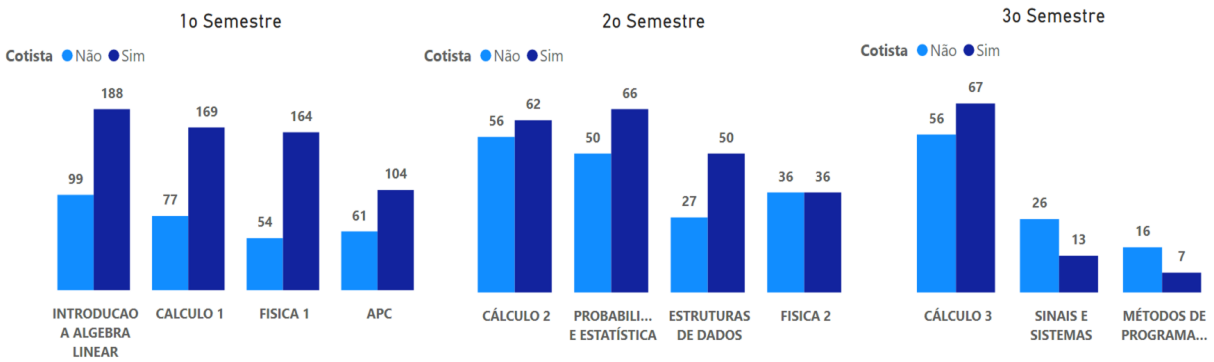

Figura 4.6: Matérias iniciais com maior número de reprovações

Por meio da análise da Figura 4.6, é possível observar que, no primeiro ano do curso, os alunos cotistas tendem a apresentar uma maior taxa de reprovação. No entanto, a partir do segundo semestre, o número de reprovações tende a se equiparar ao dos alunos não cotistas.

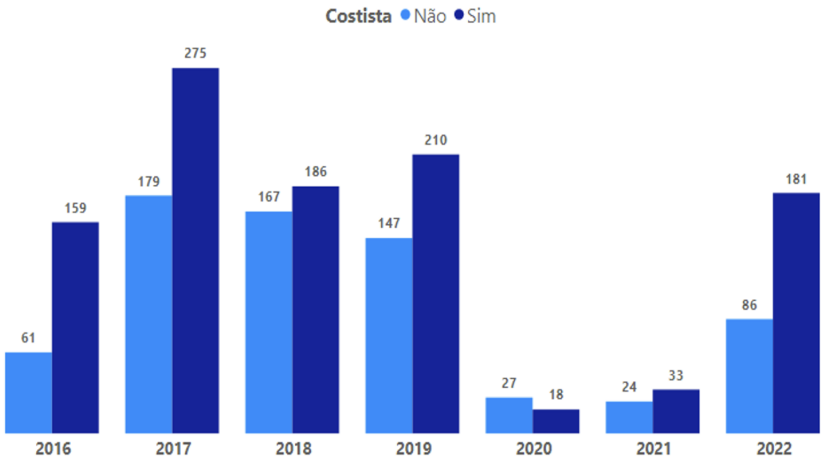

Figura 4.7: Quantidade de reprovações das matérias iniciais do curso

A Figura 4.7, apresenta a quantidade acumulada de reprovações dos alunos do curso de Engenharia de Computação nas matérias iniciais destacadas no início desta seção, com o período de análise começando em 2016. Nota-se uma diferença significativa no número de reprovações entre os alunos cotistas e não cotistas. Os alunos cotistas demonstram uma maior dificuldade nas matérias iniciais, com um número mais elevado de reprovações,

particularmente em 2017. Embora os alunos não cotistas também apresentem reprovações, os números são relativamente menores. O pico de reprovações ocorreu em 2017, enquanto o menor índice foi observado em 2021.

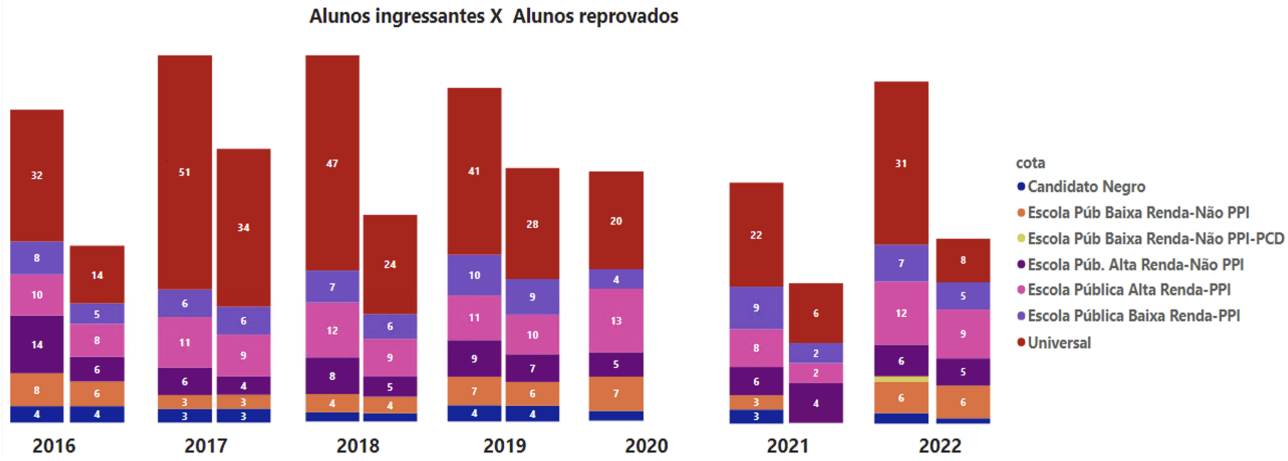

Figura 4.8: Quantidade de reprovações das mateiras do primeiro semestre

Na Figura 4.8, a barra da esquerda representa a quantidade de alunos que ingressaram no curso naquele semestre, enquanto a barra da direita mostra a quantidade de alunos que ingressaram e reprovaram em alguma disciplina durante o mesmo período. A análise é segmentada por tipo de cota. Observa-se que o grupo de alunos provenientes de escolas públicas de alta renda e PPI é o que possui o maior número de ingressantes, mas também o maior número de reprovações, cerca de 32% dos alunos. Em 2018 este grupo teve o maior registro de reprovações no período de 2016 a 2022, tendo 81% dos alunos reprovados, e em 2021 foi a menor taxa registrada, cerca de 25% de reprovações nas matérias iniciais. Já os alunos de escola pública de baixa renda e PPI também têm uma representatividade significativa, tanto em número de ingressantes quanto de reprovados, cerca de 22%, tendo destaque em 2016, onde 42% dos alunos reprovaram, sendo esta a menor taxa de reprovação deste grupo. Já em 2022 cerca de 83% dos alunos reprovaram. Alunos do grupo de Escolas Públicas de Baixa Renda e não PPI, apresentaram 100% de reprovações em 2017, 2018 e 2022, indicando a necessidade de uma maior atenção para este grupo, a menor taxa registrada foi em 2016 , 75% dos alunos reprovaram naquele ano. Outros grupos de alunos cotistas aparecem no gráfico, porém em menor quantidade. Para fins de comparação, os alunos das vagas universais também foram incluídos no gráfico, tendo destaque em 2022 com apenas 25% de reprovações contra 66% de reprovações em 2017.

Diferente da Figura 4.7, onde pode-se observar que houveram reprovacoes em 2020, na Figura 4.8, o gráfico mostra que não houve reprovações em 2020, isso quer dizer que dos alunos que entraram em 2020 nenhum foi reprovado em matérias naquele mesmo ano, as reprovações do indicados em 2020 na Figura 4.7 representam reprovações de alunos

que ingressaram em anos anteriores a 2020. Esse fenômeno pode ser atribuído às medidas adotadas pela UnB durante o período da pandemia que ocorreu durante os semestres de 2020. Nesse período, a UnB permitiu que os alunos retirassem a matrícula da disciplina a qualquer momento do semestre, o que reduziu drasticamente a quantidade de reprovações. Esse impacto pode ser observado como um grande vale no gráfico, indicando a baixa quantidade de alunos reprovados durante os anos da pandemia.

No entanto, após a pandemia, o número de reprovações em ambos os grupos de alunos cotistas e não cotistas subiu novamente. Esse fato sugere a possível existência de dificuldades enfrentadas pelos alunos no período pós-pandemia.

4.2.3 Análise de Desempenho por Sexo

No início deste capítulo, analisamos a quantidade de alunos que ingressam no curso e notamos uma grande diferença entre o número de alunos ingressantes do sexo masculino e feminino. Este fenômeno pode ocorrer devido a uma série de fatores econômicos e sociais presentes no contexto de vida de cada aluno. Tendo isso em vista, nesta seção analisaremos o desempenho acadêmico dos estudantes do curso de Engenharia de Computação considerando o sexo. Ressaltamos que, até o presente momento, no sistema de matrículas da Universidade de Brasília só é possível registrar os sexos masculino ou feminino.

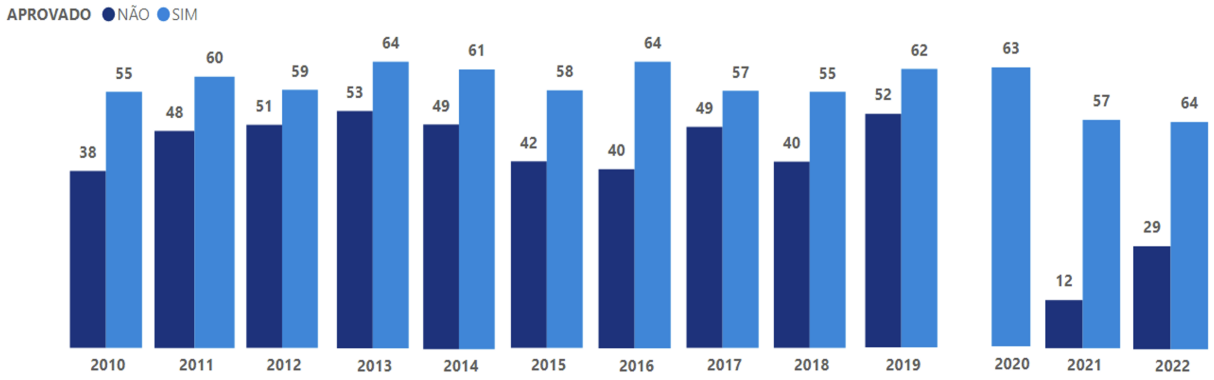

Figura 4.9: Resultado masculino de desempenho das mateiras do primeiro semestre

Analisando a Figura 4.9, observa-se que em todos os anos o número de aprovação nas matérias do primeiro semestre é maior que o número de reprovação. O gráfico de desempenho dos alunos do sexo masculino nas matérias do primeiro semestre de Engenharia de Computação mostra uma tendência geral de aprovação ao longo dos anos. Ao longo do período analisado a porcentagem média de aprovação foi de 60,26%, enquanto a média de reprovação foi de 39,74%. A maior porcentagem mensal de aprovação foi registrada em 2021, com 81,43%, enquanto a menor foi em 2012, com 51,72%. Já a maior porcentagem

de reprovação ocorreu em 2012, com 48,28%, e a menor em 2021, com 18,57%. Em 2020, devido a pandemia não houve reprovações.

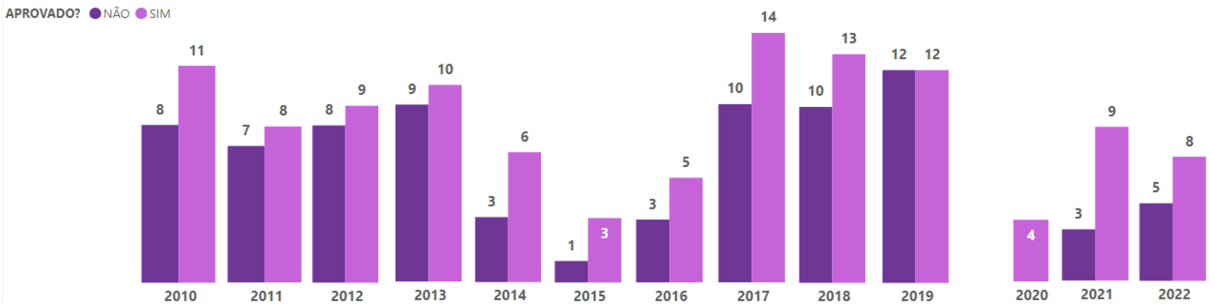

Figura 4.10: Resultado feminino de desempenho das mateiras do primeiro semestre

Nas análises da Figura 4.10, comparando com o gráfico dos alunos do sexo masculino, observa-se que as alunas têm uma distribuição mais equilibrada entre aprovações e reprovações. A análise do desempenho feminino das matérias do primeiro semestre ao longo dos anos no curso de Engenharia de Computação mostra uma média de aprovação é de 64,92%, enquanto a porcentagem de reprovação é de 35,08%.

Os dados analisados demonstram que as alunas tiveram um desempenho ligeiramente melhor ao longo do período analisado, com uma diferença de 4,66% a mais na taxa de aprovação em comparação aos alunos. Consequentemente, a taxa de reprovação das alunas foi 4,66% menor. Esses dados sugerem que, em média, as alunas tiveram mais sucesso nas disciplinas do que os alunos.

4.3 Resultados e Análise do Desempenho do Alunos em Algoritmos e Programação de Computadores

Nesta seção, a na análise de dados da disciplina Algoritmos e Programação de Computadores (APC) é apresentada, pois é através dela que muitos alunos têm o primeiro contato com a programação [35]. Esse contato inicial é crucial, pois pode influenciar significativamente o progresso no curso [36] [37]. Embora não seja a matéria com a maior taxa de reprovação no curso de Engenharia de Computação, APC é a quarta com mais reprovações, dentre as matérias analisadas neste trabalho. Por isso, é importante analisar as tendências positivas ou negativas que essa disciplina pode implicar ao longo do curso. Até 2014, a disciplina era chamada de Computação Básica; portanto, os dados nos gráficos desta seção referem-se a ambas as disciplinas.

Para esta análise geral de desempenho, foram destacados os dados a partir dos semestres de 2010. Para análise de desempenho dos alunos cotistas e não cotistas foram

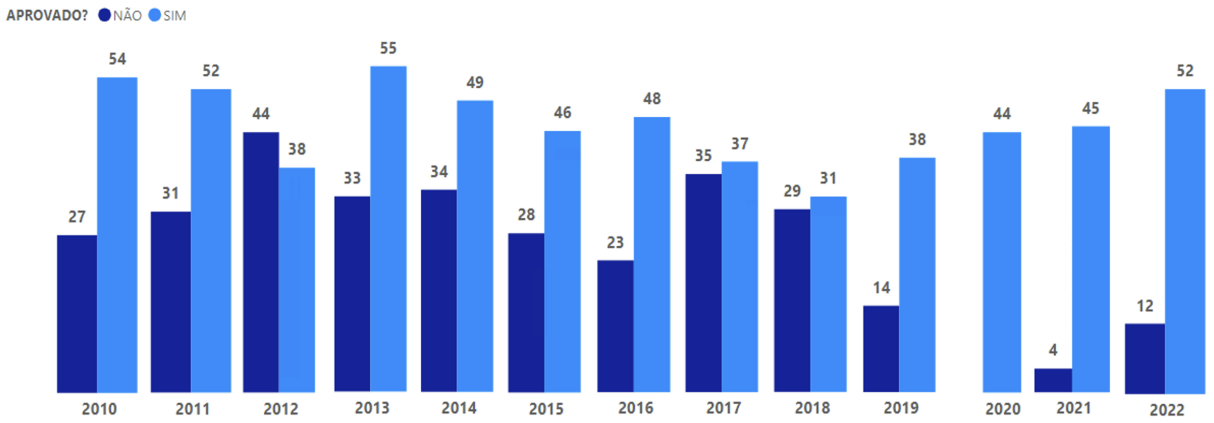

Figura 4.11: Aprovações e Reprovações dos alunos em APC

considerados somente os alunos que ingressaram a partir de 2016, ano em que a quantidade de alunos não cotistas se equiparou a quantidade de alunos não cotistas. A análise da Figura 4.11 focou na quantidade de aprovações e reprovações. Durante o período analisado, a porcentagem média de aprovação no curso de APC da Universidade de Brasília foi de 65%, enquanto a porcentagem de reprovação é de 35%. A maior porcentagem de aprovação na disciplina APC foi registrada em 2022, com 81,25%, enquanto a menor ocorreu em 2012, com 46,34%. Já a maior porcentagem de reprovação foi em 2012, com 53,66%, e a menor foi em 2022, com 18,75%.

Observa-se os efeitos da pandemia em 2020, com nenhum registro de reprovação. No entanto, a partir do primeiro semestre de 2022, os números voltaram a alcançar valores altos.

Nota-se uma diferença entre os números de aprovados e reprovados na disciplina em todos os anos analisados, e, de modo geral, os alunos tendem a ter mais facilidade em serem aprovados na matéria, visto que, na maioria dos anos, os valores referentes às aprovações são maiores. Esse comportamento indica um aspecto positivo no modo como a disciplina está sendo ministrada. Parece que a maior parte dos alunos tem um bom primeiro contato com a programação, e, por serem aprovados nessa disciplina de Introdução à Programação, têm uma perspectiva mais favorável de concluir o curso com êxito.

4.3.1 Análise de desempenho por tipo de cota

A fim de entender o desempenho dos alunos no contexto das cotas, repetimos a mesma análise aplicada nos gráficos anteriores, foram utilizados dados a partir do ano 2016, pois foi nesse período que o número de alunos ingressantes cotistas e não cotistas se igualou (Figura 4.1).

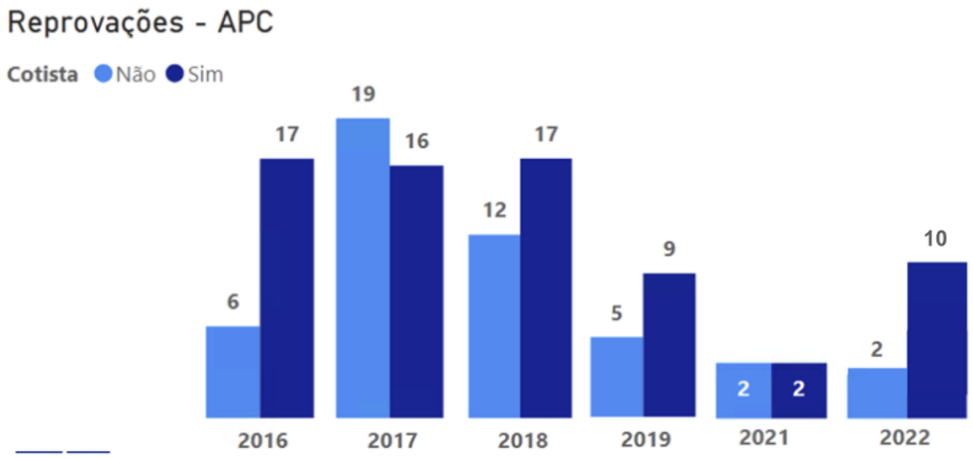

Figura 4.12: Reprovações em APC

De acordo com a Figura 4.12 os alunos cotistas aparentam manter uma quantidade de reprovações semelhante ao longo dos anos, enquanto os números referentes aos alunos não cotistas variam bastante de um ano para o outro. No geral, a diferença entre os dois grupos é pequena, com casos em que o número de reprovações foi o mesmo (como em 2021). Nota-se também uma diminuição acentuada no número de reprovações nos anos de 2020, 2021 e 2022. Esse fenômeno ocorreu por conta da pandemia de Covid-19, durante a qual a quantidade geral de reprovações dos alunos da Universidade de Brasília foi mínima.

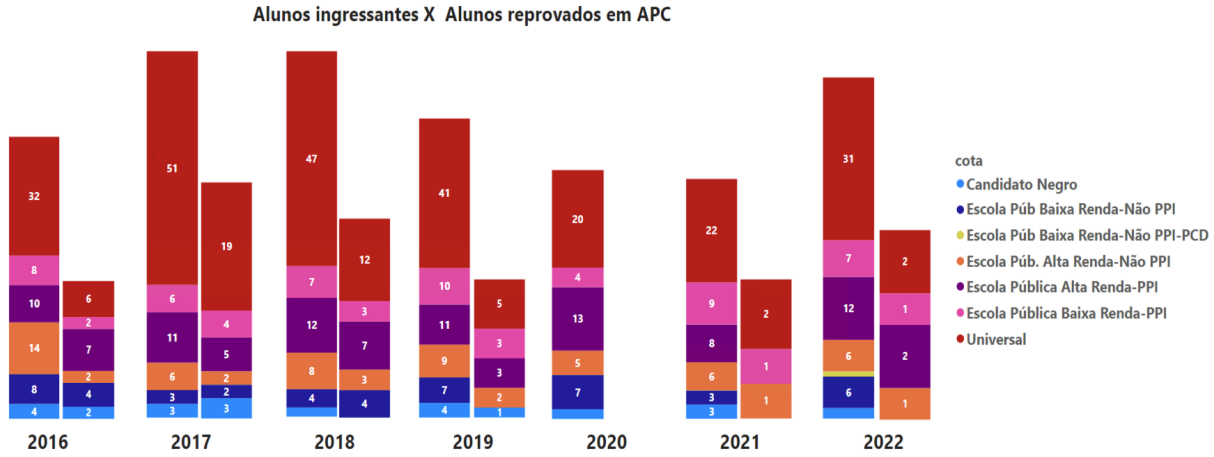

Figura 4.13: Reprovações por Tipo de Cota em APC

A análise da Figura 4.13 revela que no período de 2016 a 2022, dentre os alunos reprovados, o grupo de Escola Pública Alta Renda - Não PPI em 2016 onde apenas 14% dos aluno reprovaram em APC, porém em 2018 37% dos alunos foram reprovados. Os grupos de Escola Púb Baixa Renda - Não PPI e Escola Púb Alta Renda - PPI, possuem uma baixa taxa de reprovação de 16,67% tendo destaque em 2022 onde a taxa

de reprovação foi de 16% e 0% respectivamente. Por outro lado, em 2018 alunos baixa renda não PPI apresentaram 100% de reprovações, enquanto os alunos de alta renda PPI apresnetaram 70% de reprovações em 2016. Já o grupo Escola Pública Baixa Renda - PPI possui 21,79% da reprovação total dos alunos, sendo que em 2021 apenas 11% dos alunos reprovaram em APC e em 2017 66% dos alunos foram reprovados. O grupo com a menor porcentagem total de reprovações neste mesmo período foi o grupo de Candidatos Negros, que representa 11,54% das reprovações, tendo 0% de reprovações em 2018, 2020, 2021 e 2022. Contudo em 2017 teve 100% dos alunos reprovados em APC.

Academicamente, esses dados sugerem que exite uma diferença de desempenho entre alunos cotistas e não cotistas, indicando que os cotistas podem estar enfrentando dificuldades no aprendizado de programação, visto que alunos cotistas reprovam mais que os alunos não cotistas nas mateiras do primeiro semestre (Figura 4.6). Socialmente, isso pode refletir desigualdades preexistentes, como diferenças na qualidade do ensino fundamental e médio, acesso a recursos educacionais e suporte extracurricular, que impactam o desempenho acadêmico desses alunos no ensino superior. Esses dados ressaltam a necessidade de políticas públicas e práticas de apoio direcionadas aos alunos cotistas para diminuir essas diferenças e promover a igualdade de oportunidades dentro da universidade, reiterando a importância da Lei 12.711/2012.

4.3.2 Análise de Desempenho por Sexo

O gráfico da Figura 4.14 apresenta uma leve variação no desempenho dos alunos do sexo masculino na disciplina de APC. Observa-se que, na maior parte dos anos analisados, os alunos tendem a ser mais aprovados em APC, com os números de aprovados superando os de reprovados. A porcentagem média total de aprovação para os alunos do sexo masculino na disciplina APC é de 66,82%, enquanto a média de reprovação é de 33,18%.

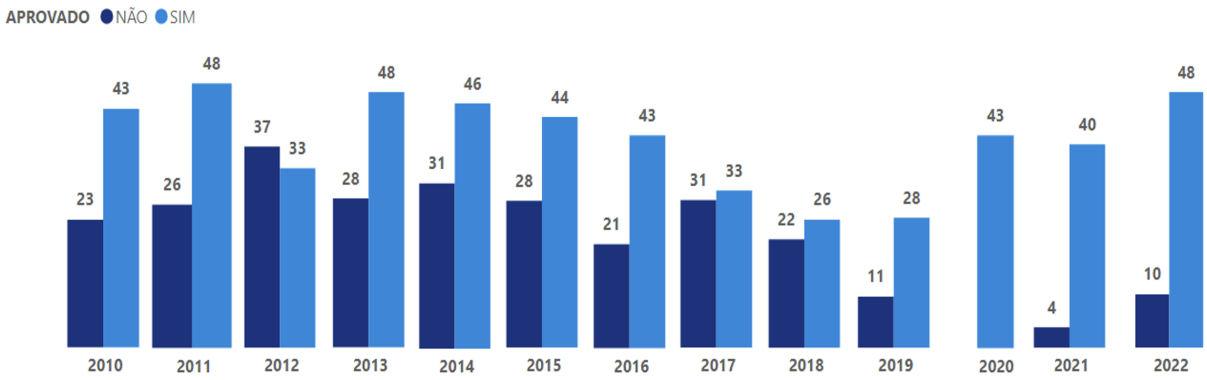

Figura 4.14: Resultado masculino de desempenho em APC

A análise dos dados em APC para alunos do sexo masculino revela que a maior porcentagem de aprovação foi de 82,76% em 2022, enquanto a menor foi de 47,14% em 2012. Em termos de reprovação, a maior porcentagem foi registrada em 2012, com 52,86%, e a menor ocorreu em 2022, com 17,24%.

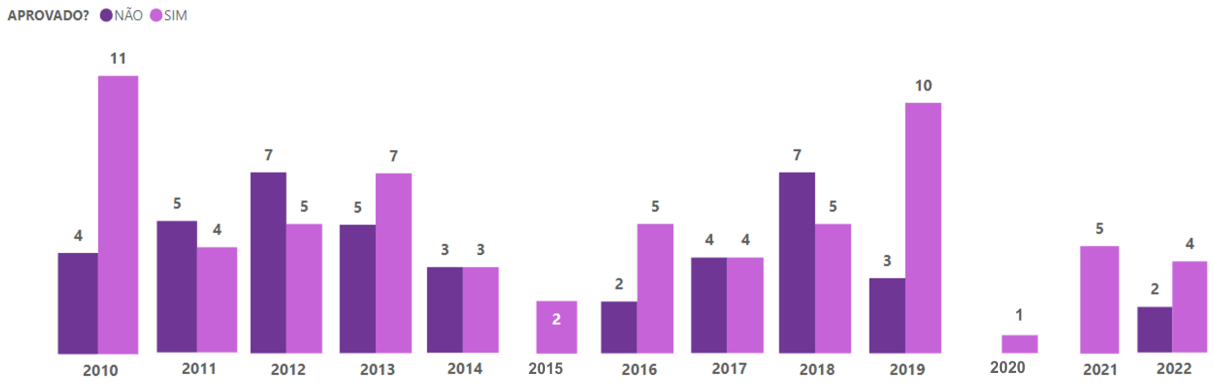

Figura 4.15: Resultado feminino de desempenho em APC

A Figura 4.15 que representa o desempenho das alunas também exibe um padrão de variação entre aprovações e reprovações. A porcentagem média total de aprovação para as alunas é de 61,11%, enquanto a média de reprovação é de 38,89%.

A comparação entre as porcentagens médias de aprovação e reprovação na disciplina APC para os sexos masculino e feminino mostra uma diferença. As alunas apresentaram uma média de aprovação de 61.11%, enquanto os alunos do sexo masculino tiveram 66,82%, uma diferença de 5.71% a favor dos alunos. Em relação à reprovação, os alunos tiveram uma média de 33,18%, ligeiramente inferior à das alunas, que foi de 38,89%.

4.4 Resultados e Análise Acerca da Forma de Saída do Aluno

Cabello e Chagas afirmam que a realidade da evasão nas universidades é uma preocupação significativa, especialmente em cursos com altos índices de reprovação nos primeiros anos. Esses cursos geralmente também apresentam as maiores taxas de evasão, o que destaca a importância de compreender o que ocorre durante essa fase crítica de adaptação dos estudantes ao ensino superior. A reprovação tem um impacto negativo substancial na taxa de evasão, especialmente para aqueles que enfrentam dificuldades logo no início do curso. Portanto, políticas públicas voltadas para reduzir a evasão devem priorizar os primeiros anos dos cursos, implementando medidas que ofereçam suporte acadêmico e emocional aos estudantes [38].

Da mesma forma, ingressar na universidade não garante automaticamente boas oportunidades aos alunos, especialmente se não houver condições favoráveis para sua permanência até a conclusão do curso. Segundo Lima a evasão muitas vezes está associada ao fracasso acadêmico, vulnerabilidade social e falta de suporte institucional adequado. No entanto, há também casos em que os estudantes abandonam seus cursos simplesmente porque não encontram sentido no tipo de ensino oferecido [39]. Portanto, é crucial analisar os índices de evasão e conclusão dos alunos do Curso de Engenharia de Computação da UnB para entender os motivos específicos que levam os estudantes a abandonar seus estudos.

4.4.1 Análise por Forma de Saída

Esta seção analisa a forma de saída dos alunos do curso de Engenharia de Computação. Neste trabalho evasão foi considerado os alunos que saíram do curso e não se formaram.

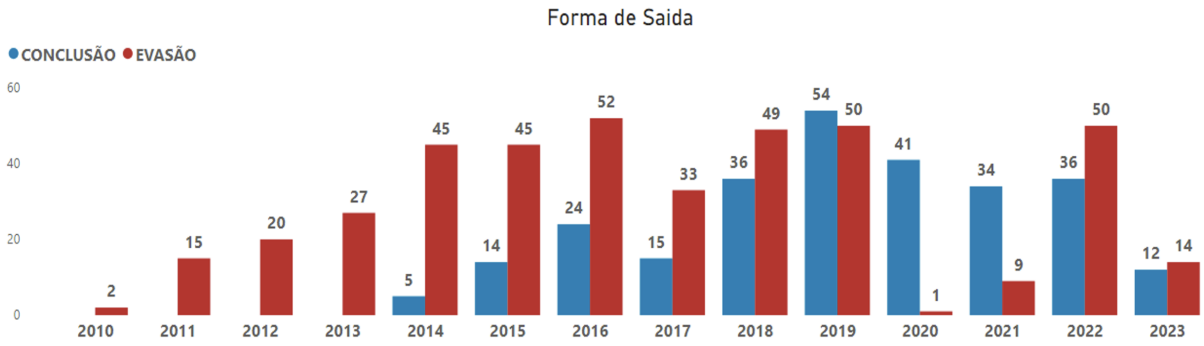

Figura 4.16: Resultado gerais de evasão e conclusão

Avaliando a Figura 4.16, é possível notar que, a porcentagem média de conclusão dos alunos no curso de Engenharia de Computação é de 39,68%, enquanto a média de evasão é de 60,32%. Nos primeiros cinco anos do curso de Engenharia de Computação, observa-se um número significativo de desistências e abandonos. Somente em 2014, cinco anos após o início do curso de Engenharia de Computação, formaram-se os 5 primeiros alunos. Ao longo da maior parte da história do curso, o número de evasões supera o de conclusões, com exceção dos anos de pandemia, quando essa tendência se altera significativamente. De forma geral, até 2020 os índices de evasão são consistentemente mais altos do que os de conclusão. No entanto, ao analisar os anos de 2019 e 2022, períodos pré e pós-pandemia, nota-se uma tendência de equilíbrio entre os números de evasão e conclusão, sugerindo que a pandemia influenciou diretamente o comportamento de evasão dos alunos. A flexibilização do trancamento de disciplinas neste período pode ter facilitado a conclusão para alunos que já estavam próximos de finalizar o curso, além de reduzir o número de reprovações e, consequentemente, de evasões durante esse período.

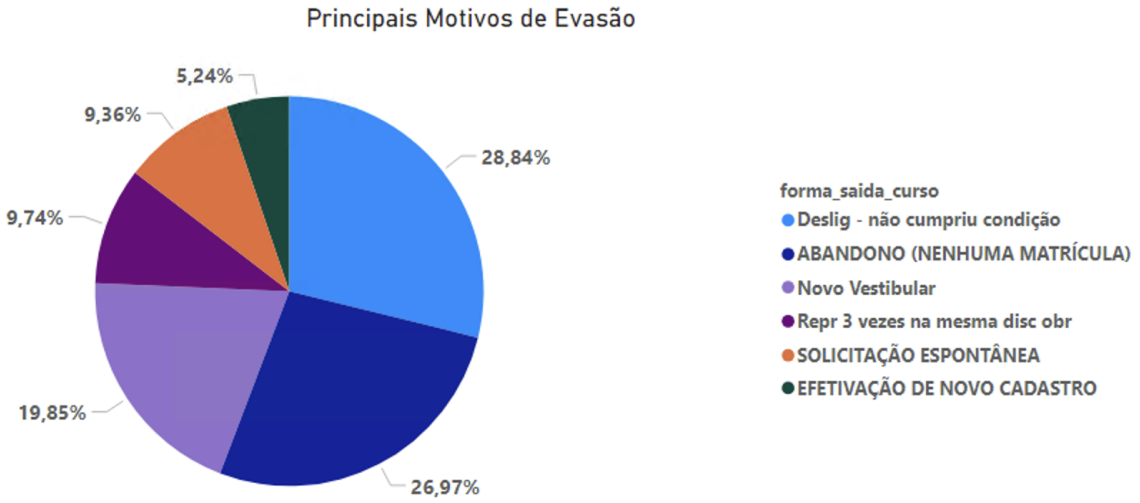

Figura 4.17: Principais motivos de evasão do curso

De acordo com o SIGAA os principais motivos são: desligamento , abandono de curso, realização de um novo vestibular como alternativa entre outros (Figura 4.17). Os motivos como Mudança de Curso, Falecimento e Desligamento por Falta de Documento foram omitidos por representatem menos de 1% dos dados.

Apesar de existirem motivos de evasão que não significa que o aluno nescessáriamente abandonou o curso, como por exemplo a realização de um novo vestibular, Cabello e Chagas em [38], levantam a preocupação de que a alta taxa de evasão também pode indicar insatisfação com o curso ou o desejo de mudar de área, eles observaram que aproximadamente 80% dos alunos que evadiram o fizeram por essas mesmas razões.

Vários motivos podem levar os alunos a evadirem. Por exemplo, a alta taxa de desligamento por não cumprimento de condições e reprovações repetidas (3 vezes) sugere dificuldades significativas relacionadas ao desempenho acadêmico dos alunos. De qualquer maneira, é incontestável a necessidade urgente de políticas focadas na retenção dos alunos, especialmente nos primeiros anos do curso.

4.4.2 Análise de Evasão e Conclusão por Tipo de Cota

Com o objetivo de analisar a trajetória dos alunos cotistas e não cotistas no curso, foram contabilizados os alunos que ingressaram a partir de 2016, ano no qual o número de alunos ingressantes cotistas e não cotistas mais se equiparou. Na Figura 4.18 observa-se que, quatro anos após o ingresso, o primeiro aluno a se formar, em 2019, foi um aluno cotista. Nos anos subsequentes, o número de alunos cotistas formados cresceu pouco, apresentando uma diferença significativa em relação aos não cotistas.

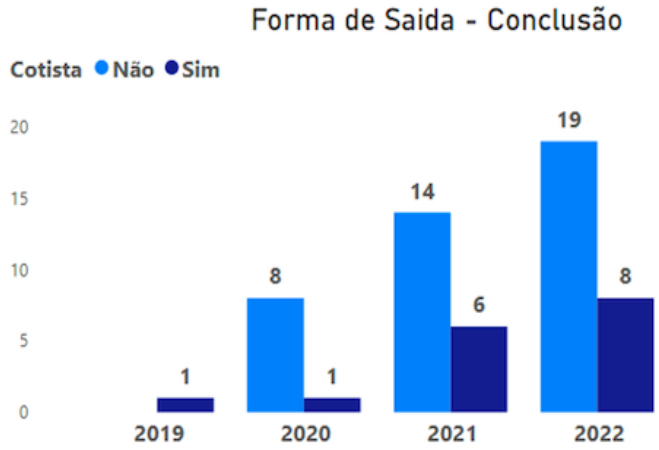

Figura 4.18: Resultado gerais de conclusão

Apesar dessa tendência de crescimento, o número de conclusões de curso por parte de alunos não cotistas continua maior do que o de cotistas. Um ponto preocupante é que, mesmo com a distribuição equilibrada de vagas (50% reservadas para cotistas), em 2016, os cotistas ainda não se formavam na mesma proporção que os não cotistas. Por outro lado, o gráfico evidencia um aspecto positivo: o crescimento gradual do número de formandos cotistas ao longo dos anos, ainda que de forma lenta.

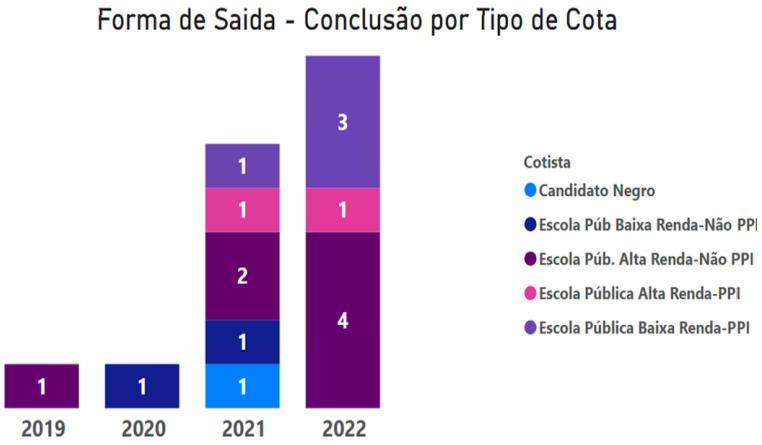

Figura 4.19: Resultado de Conclusão por Tipo de Cota

Na figura Figura 4.19 pode-se observar como os alunos de diferentes tipos de cota concluiram o curso. Nota-se que os alunos de escola pública de alta renda não PPI tende a concluir mais o curso do que os outros tipos de cota, cerca de 11% dos alunos deste grupo de cotas concluíram o curso no periodo de 2016 a 2022. A quantidade de alunos de escola pública de baixa renda PPI também se mostrou em maior número em relação a conclusão do curso. Indicando que alunos de escolas públicas e alta renda PPI apresentam uma menor tendência em se formar.

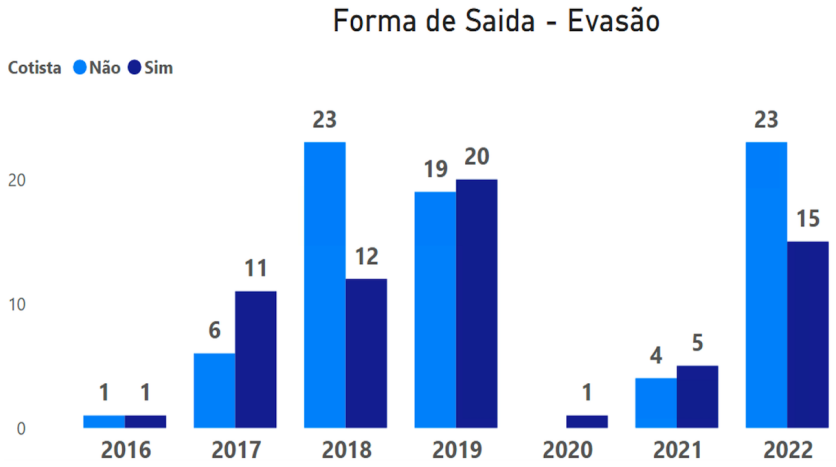

Figura 4.20: Resultado gerais de evasão

A Figura 4.20 apresenta os números de alunos cotistas e não cotistas que evadiram o curso, a análise das porcentagens médias de evasão entre cotistas e não cotistas mostra que as evasões foram levemente equilibradas. A média de evasão dos alunos cotistas foi de 45,86%, enquanto a média de evasão dos não cotistas foi de 54,14%.

Essa diferença de 9,34% sugere que há uma discrepância pouco significativa entre os grupos em termos de evasão. A proporção de alunos que evadiram o curso parece ser bastante similar, independentemente de serem cotistas ou não.

Analisando as evasões em relação ao tipo de cota, nota-se pela Figura 4.21 que alunos de escola pública de alta renda PPI tem uma grande representatividade nos números de alunos que evadiram o curso, sendo o grupo de cotas que mais evade no decorrer dos anos. Os grupos escolas públicas baixa renda PPI e alta renda não PPI, também apresentaram recorrência nas quantidades de alunos que evadiram o curso no período analisado. Alunos negros e baixa renda não PPI apresentaram uma menor tendência a evasão.

A partir da análise da Figura 4.23 e Figura 4.22 dos alunos cotistas e não cotistas que ingressaram no curso a partir de 2016, é possível entender os motivos que levaram os alunos a evadirem o curso.

A Figura 4.22 ilustra que a maior parcela dos alunos não cotistas do curso de Engenharia de Computação, correspondendo a 34,12%, abandonou o curso. Seguindo essa tendência, cerca de 20% dos alunos evadiram devido à incapacidade de atender às condições para permanência no curso. Adicionalmente, um terceiro grupo, representando aproximadamente 14% dos alunos, optou por evadir o curso com o intuito de prestar um novo vestibular.

Da mesma maneira, analisando a Figura 4.23 observa-se que os alunos cotistas enfrentam dificuldades semelhantes às dos não cotistas. Entretanto, o principal motivo de evasão entre os cotistas é o não cumprimento das condições de permanência da univer-

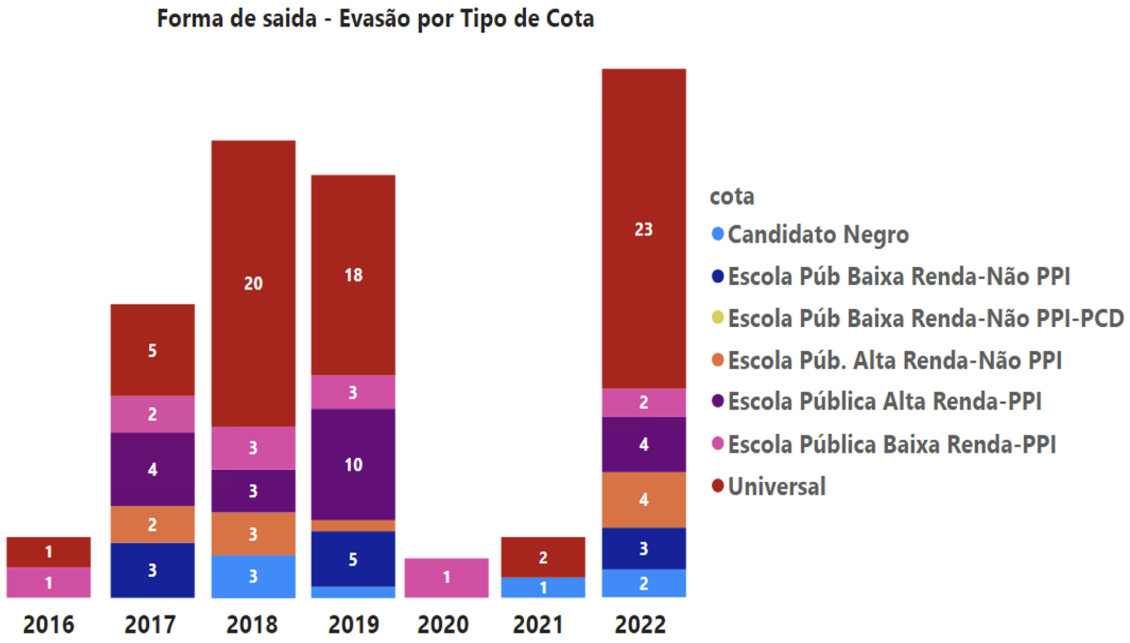

Figura 4.21: Resultado gerais de Evasão por Tipo de Cota

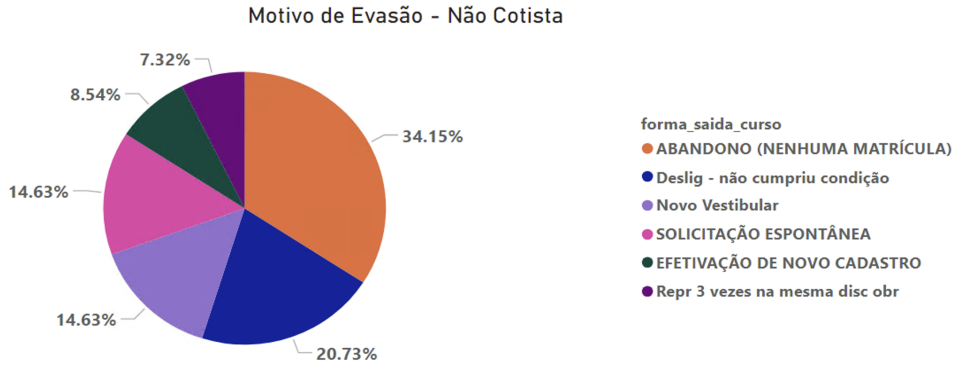

Figura 4.22: Motivos de Evasões - Alunos Não Cotistas

sidade, representando 28% dos casos. Em seguida, uma quantidade similar de alunos deixou o curso por abandono, o que sugere tanto a exigência do curso quanto possíveis fatores externos que desmotivam a continuidade. Por fim, 18% dos alunos optaram por evadir o curso com o objetivo de prestar um novo vestibular.

De modo geral, os alunos de ambos os grupos apresentam comportamentos semelhantes em relação à evasão, com os mesmos motivos predominantes para a saída do curso.

4.4.3 Análise de Evasão e Conclusão por Sexo

Na Figura 4.24 observa-se a contagem de alunos do sexo masculino que evadiram e concluíram o curso de Engenharia da Computação de 2010 a 2022. Vemos que dentre os alunos

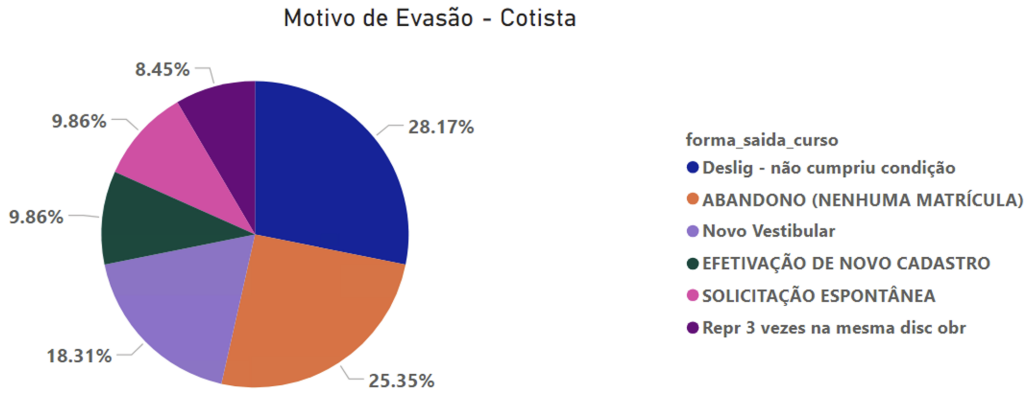

Figura 4.23: Motivos de Evasões - Alunos Cotistas

que ingressaram do curso em 2010, 4 deles se formaram em 2014, nos anos anteriores é possível perceber uma grande quantidade de evasão inicial do curso.

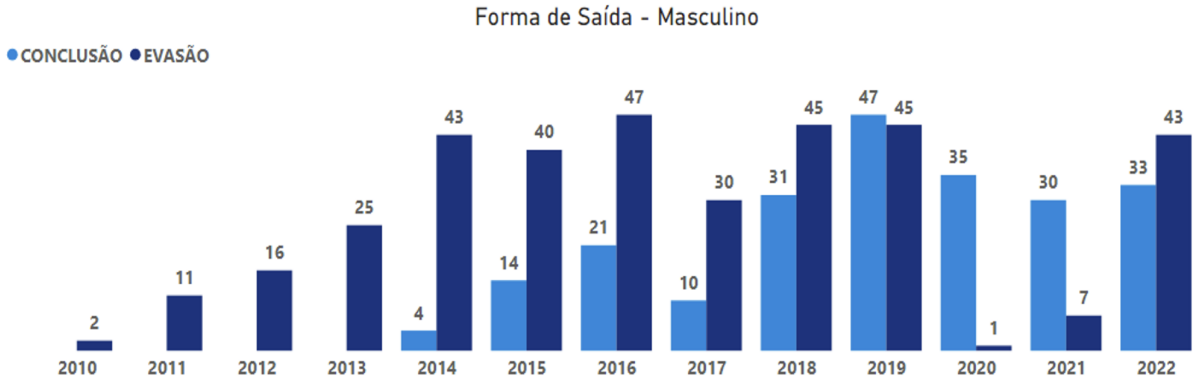

Figura 4.24: Resultado de evasão e conclusão masculino

Até o ano de 2016 vemos que o gráfico segue um padrão de crescimento, apresentando o número de evasão sempre maior que o de conclusão. A porcentagem média de conclusão dos alunos do curso é de 38,79%, enquanto a porcentagem média total de evasão é de 61,21%.

A Figura 4.25 apresenta a contagem de alunas que concluíram ou evadiram do curso de Engenharia da Computação ao longo dos anos. No geral o gráfico revela o comportamento da taxa de evasão e conclusão das alunas, destacando a necessidade de apoio e inclusão para melhorar a taxa de permanência e o sucesso das mulheres na Engenharia de Computação, especialmente considerando os desafios e a realidade de um ambiente predominantemente masculino. A porcentagem média total de conclusão das alunas no curso de Engenharia de Computação é de 44,16%, enquanto a porcentagem média total de evasão é de 55,84%

Podemos observar que, em 2014, ocorreu a primeira formatura da turma iniciada em 2010, enquanto outras duas alunas evadiram. Em comparação, as alunas apresentaram

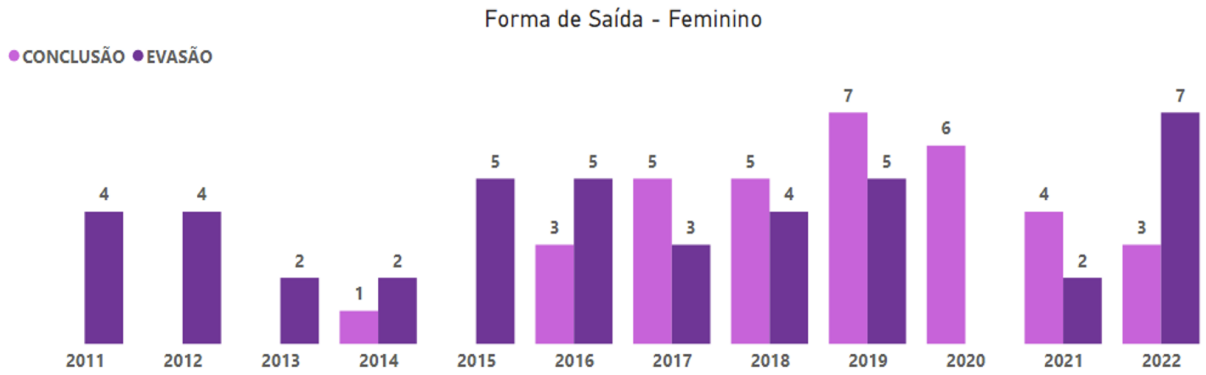

Figura 4.25: Resultado de evasão e conclusão feminino

uma média de 44,16% de conclusão, enquanto os alunos tiveram 38,79%, o que indica que as alunas concluíram o curso em maior proporção. Já em relação à evasão, os alunos tiveram uma média mais alta, de 61,21%, comparada aos 55,84% de evasão entre as alunas.

A pandemia de COVID-19 impactou fortemente esses números. Em 2020, houve 6 conclusões, devido às flexibilizações da UnB mencionadas neste trabalho. Em 2021, o número de conclusões ainda superou as evasões, sugerindo uma adaptação às novas condições de ensino.

As razões para a grande diferença entre os números masculinos e femininos podem ser várias, incluindo fatores socioeconômicos e ambiente acadêmico. No entanto, é fato que a presença de mulheres na computação traz benefícios como diversidade de pensamento, inovação e melhor representação de diferentes grupos de usuários na criação de tecnologias. Políticas públicas e iniciativas de inclusão podem ajudar a aumentar o número de conclusão de alunas nos cursos relacionados a tecnologia.

Pode-se observar que a pandemia também interferiu significativamente nos números dos gráficos dos alunos do curso. Houve um aumento na taxa de conclusão e uma grande diminuição na quantidade de evasões, como observado em 2020, quando nenhum aluno ou aluna evadiu. Em 2022, no período pós-pandemia, os números começaram a retornar gradativamente ao comportamento observado nos anos anteriores à pandemia.

4.5 Análise Preditiva com Algoritmos de *Machine Learning*

Nesta seção são apresentados os resultados da análise dos principais fatores que podem contribuir com a desistência dos alunos do curso de Engenharia de Computação e o grau de relevância de cada fato. Os dados foram separados em classes, ou *features*, para analisar as tendências do comportamento dos alunos presentes no *dataset*, obtendo assim

os indicadores de maior importância em relação à desistência dos estudantes. Para o treinamento do algoritmos as classes principais foram: *Id_pessoa*, *ira*, *gênero*, *idade*, *estado_nascimento*, *cota*, *segundo_grau_tipo_escola*, *curso*, *forma_ingresso_unb*, *APC* (vezes que cursou a matéria).

Por serema modelos bem conhecidos por sua capacidade de classificação e boa precisão, os algoritmos utilizados foram *Gradient Boosting Machine* [24], *Random Forest* [23], e *Support Vector Machine* [1]. Essas classes foram informadas aos algoritmos, que com base na forma de saída dos alunos previu qual delas podem ou não influenciar mais na classe *forma_saida_curso*, informando ao final o grau de importância da classe em relação a forma de saída do aluno. Como o objetivo era prever o padrão do comportamento da saída dos alunos, as quantidades dos alunos ativos não foram utilizadas para a realização da predição, a base final possuía um total de 358 alunos do curso de Engenharia de Computação, a organização dos dados informados aos algoritmos esta na Tabela 4.1.

Tabela 4.1: Quantidade de Evasão e Conclusão.

Classe	Quantidade
Evasão	229
Conclusão	129

A Tabela 4.2 se refere as quantidades de aluno que concluíram e evadiram o curso, e mostra a forma final do *Dataset* utilizado para treinamento dos algoritmos.

4.5.1 Resultados Algoritmo Gradient Boosting Machine

Para prever a influência de cada *feature* no conjunto de dados, após o tratamento dos dados, 20% dos dados foram utilizados como conjunto de teste, enquanto os outros 80% foram usados para treinamento. Dessa forma, o algoritmo alcançou uma precisão de 87.41%, Esta métrica indica a precisão média do modelo ao prever a evasão ou conclusão do aluno, baseada na validação cruzada (Figura 2.4). A porcentagem alcançada nesse teste sugere que o modelo é eficaz em suas previsões. O número de ocorrências de cada classe no conjunto de testes foi coletado pelo algoritmo e consideradas relativamente balanceadas, com 30 das ocorrências tendendo para a conclusão do curso e 42 tendendo para a evasão, como pode ser visto na Tabela 4.3. Como era esperado, as métricas de precisão, revocação e F1-score sugerem que o modelo é mais eficiente em prever evasão (com precisão e F1-score mais altos) do que conclusão, justificando a alta acuracidade apresentada.

Ao final da predição, o resultado de importância por *feature* pode ser observado na Figura 4.26. A importância de uma característica na previsão de evasão dos alunos refere-se a quanto essa variável contribui para a precisão do modelo de *machine learning* ao prever

Tabela 4.2: Qtd Con: quantidade de Conclusões; Qtd Ev: quantidade de Evasões; % Ev: Porcent Evasão; % Con: Porcent Conclusões

Classe	Descrição	Qtd Con	Qtd Ev	% Ev	% Con
Gênero	Masculino	116	208	64.20%	35.80%
	Feminino	13	21	61.76%	38.24%
Idade	Até 17	19	18	48.65%	51.35%
	Entre 18 e 20	102	169	62.36%	37.64%
	Entre 21 e 25	7	25	78.13%	21.88%
	Entre 26 e 30	0	9	100.00%	0.00%
	Mais de 30	1	8	88.89%	11.11%
Estado	DF	84	158	65.29%	34.71%
	Fora DF	45	71	61.21%	38.79%
Cota	Candidato Negro	3	21	87.50%	12.50%
	Pública Alta Renda	10	57	85.07%	14.93%
	Pública Baixa Renda	6	32	84.21%	15.79%
	Universal	110	119	51.95%	48.05%
Tipo Escola	Pública	23	116	83.45%	16.55%
	Particular	106	113	51.58%	48.42%
Forma Ingresso	ENEM	16	6	27.27%	72.73%
	PAS	40	68	62.96%	37.04%
	Vestibular	59	103	63.56%	36.44%
	Outros	14	52	78.79%	21.21%
APC	1 vez	36	125	77.65%	22.35%
	2 vezes	77	74	52.03%	47.97%
	3 vezes	4	26	86.67%	13.33%
	4 vezes	12	3	20.00%	80.00%

se um aluno vai ou não abandonar o curso, as características com maior importância são aquelas que fornecem mais informações úteis para distinguir entre alunos que vão evadir e os que vão permanecer, dessa forma foi destacado apenas os primeiros três motivos mais importantes mostrados no gráfico.

Na Figura 4.26, o F Score é a métrica que representa a importância das características, o IRA tem o F Score mais alto (408), indicando que esta característica é a mais determinante para prever se um aluno vai abandonar o curso. Isso sugere que o desempenho acadêmico dos alunos é um forte indicador de evasão. Alunos com IRA baixo têm uma maior probabilidade de evadir.

Com um F Score de 89, a Forma de Ingresso na UnB é a segunda característica mais importante, diferentes formas de ingresso podem estar associadas a diferentes níveis de preparação e adaptação, influenciando a probabilidade de evasão. Para ser mais preciso, as formas de ingresso fornecidas ao modelo de treinamento foram agrupadas, de forma que, para essa análise não é possível identificar especificamente qual e a foram de ingresso que mais contribui para a evasão do aluno.

Tabela 4.3: Resultados do Algoritmo GBM

Classe	**Precision**	**Recall**	**F1-Score**	**Support**
CONCLUÍDO	0.75	0.80	0.77	30
EVASÃO	0.85	0.81	0.83	42
Macro Avg	0.80	0.80	0.80	72
Weighted Avg	0.81	0.81	0.81	72
Accuracy	0.8741			

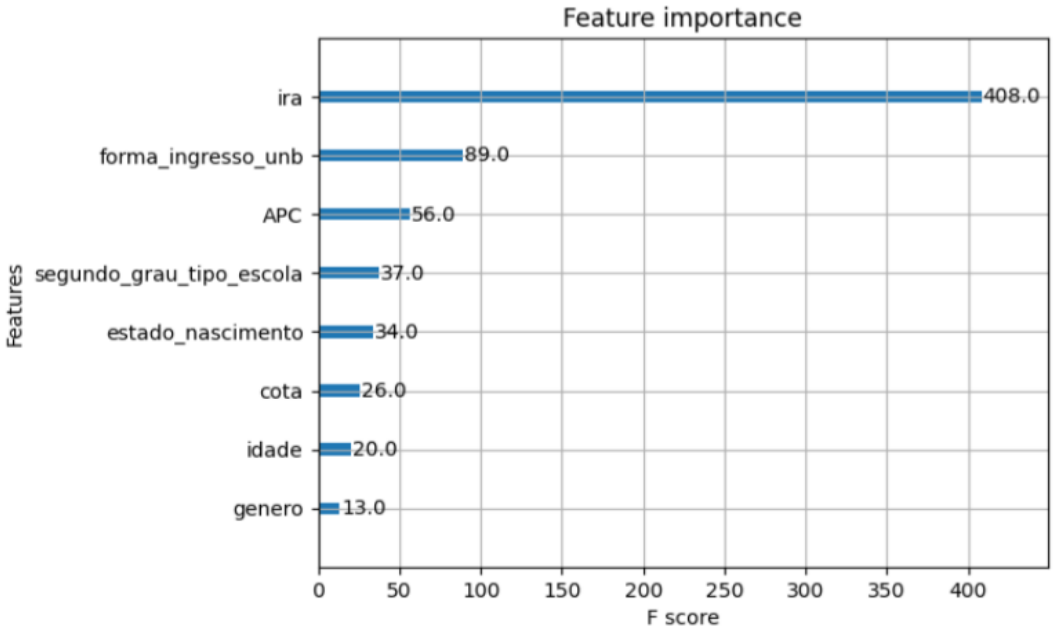

Figura 4.26: Resultado de predição do modelo de dados - Algoritmo GBM

Algoritmos e Programação de Computadores é a terceira *feature* mais importante na evasão dos alunos, o que é esperado, dada a dificuldade da disciplina e seu impacto significativo na decisão dos alunos de continuar ou não no curso após cursá-la. Apesar de estar em terceiro lugar no *ranking* de importância, é crucial prestar atenção nessa disciplina, pois, conforme o gráfico sugere, quanto mais vezes os alunos precisam refazê-la, maiores são as chances de evasão.

Seguindo a ordem de importância, temos as seguintes *features*: tipo de escola, estado de nascimento, cota, idade e gênero. Cada fator exerce uma influência significativa na decisão de evasão dos alunos. Esses dados permitem compreender melhor o peso de cada motivo que contribui para as taxas de evasão observadas atualmente no curso de Engenharia de Computação.

4.5.2 Resultados Algoritmo Support Vector Machine

A fim de obter uma melhor análise dos dados educacionais utilizamos também o classificador SVM, esse modelo tem um bom desempenho na detecção de evasões, com uma boa precisão e F1-score, o que é importante para minimizar falsos negativos, ou seja, quando algumas conclusões são incorretamente classificadas como evasões [26].

Tabela 4.4: Resultados do Algoritmo SVM

Classe	**Precision**	**Recall**	**F1-Score**	**Support**
CONCLUÍDO	0.76	0.85	0.80	41
EVASÃO	0.90	0.84	0.87	67
Macro Avg	0.83	0.84	0.84	108
Weighted Avg	0.85	0.84	0.84	108
Accuracy	0.8426			

Como mostra a Figura 4.4, a precisão geral do modelo é de 0.8426, indicando que 84.26% das previsões estão corretas. A métrica de Support indica que foram encontrados 41 casos verdadeiros de conclusão no conjunto de dados e 67 casos verdadeiros de evasão (Figura 2.5). No geral a precisão do algoritmo indicou que 76% das previsões de conclusão estão corretas e 90% das previsões de evasão estão corretas.

Dessa forma, foi analisado os resultados dos principais coeficientes previstos pelo algoritmo demonstrado na Tabela 4.5. Para compreender os resultados do algoritmo é necessário saber que os valores positivos na imagem indicam que um aumento na característica está associado a um aumento na probabilidade de evasão, e valores negativos indicam que um aumento na característica está associado a uma diminuição na probabilidade de evasão.

- IRA (Índice de Rendimento Acadêmico): Um IRA mais alto está bastante associado a uma menor probabilidade de evasão. Isso confirma que alunos com melhor desempenho acadêmico tendem a permanecer no curso.

- APC: a aprovação do aluno na disciplina Algoritmos e Programação de Computadores está associada a uma menor probabilidade de evasão.

- Idade: Alunos mais jovens (até 17 anos) têm menor probabilidade de evasão, enquanto a faixa de idade entre 21 e 25 anos tem maior probabilidade de evasão. Idades acima de 26 anos não têm impacto significativo.

- Cotas: Valores de coeficiente negativo (Figura **??**), indicam que a caracteristica está associada a uma diminuição na probabilidade de evasão, dessa forma, alunos cotistas negros têm uma probabilidade ligeiramente menor de evadir, enquanto cotistas de escolas públicas de alta renda têm maior probabilidade.

Tabela 4.5: Resultado de predição do classificador SVM

Característica	**Coeficiente**
ira	-2.745509e+00
APC	-1.115957e+00
genero_F	2.260243e-01
genero_M	-2.260243e-01
idade_Até 17	-7.771643e-01
idade_Entre 18 e 20	2.047676e-01
idade_Entre 21 e 25	5.723967e-01
idade_Entre 26 e 30	0.000000e+00
idade_Mais de 30	0.000000e+00
estado_nascimento_DF	-2.418626e-02
estado_nascimento_Fora do DF	2.418626e-02
cota_Candidato Negro	-2.593047e-01
cota_Escola Pública Alta Renda	2.877067e-01
cota_Escola Pública Baixa Renda	3.368097e-02
cota_Universal	-6.208294e-02
segundo_grau_tipo_escola_Particular	-4.467676e-01
segundo_grau_tipo_escola_Pública	4.467676e-01
forma_ingresso_unb_ENEM	-1.372047e-01
forma_ingresso_unb_OUTROS	-2.047676e-01
forma_ingresso_unb_PAS	-7.778702e-02
forma_ingresso_unb_VESTIBULAR	4.197592e-01

- Tipo de Escola: Alunos de escolas particulares têm menor probabilidade de evasão comparados a alunos de escolas públicas.

- Forma de Ingresso na UnB: Alunos que ingressaram via ENEM, PAS ou Vestibular têm menor probabilidade de evasão, enquanto aqueles que ingressaram por vestibular têm maior probabilidade de evadir do curso.

4.5.3 Resultados Algoritmo Random Forest

Já no algoritmo de classificação Randon Forest (Tabela 4.6), foi observado que o modelo tem um bom desempenho no geral, com precisão e recall altos, especialmente para a classe de evasão, o que foi importante para identificar corretamente os alunos que podem deixar o curso. O equilíbrio entre precisão e recall nas duas classes sugere que o modelo faz um bom trabalho em minimizar tanto falsos positivos quanto falsos negativos.

A Acurácia do modelo é de 0.8704, indicando que cerca de 87.04% das previsões estão corretas. A precisão indica que 81% das previsões de conclusão estão corretas e 91% das previsões de evasão estão corretas. Pela métrica Support vemos que o algoritmo

Tabela 4.6: Resultados do Modelo Random Forest

Classe	**Precision**	**Recall**	**F1-Score**	**Support**
CONCLUÍDO	0.81	0.85	0.83	41
EVASÃO	0.91	0.88	0.89	67
Macro Avg	0.86	0.87	0.86	108
Weighted Avg	0.87	0.87	0.87	108
Accuracy	0.8704			

identificou 41 casos verdadeiros de conclusão no conjunto de dados e 67 casos verdadeiros de evasão.

Tabela 4.7: Importâncias das Features no Modelo RF

Feature	**Importance**
ira	0.619856
APC	0.100668
cota_Universal	0.051400
segundo_grau_tipo_escola_Particular	0.027644
segundo_grau_tipo_escola_Pública	0.026220
forma_ingresso_unb_ENEM	0.023195
cota_Escola Pública Alta Renda	0.015664
estado_nascimento_Fora do DF	0.014550
forma_ingresso_unb_OUTROS	0.013768
forma_ingresso_unb_VESTIBULAR	0.013373
estado_nascimento_DF	0.013139
idade_Até 17	0.012249
forma_ingresso_unb_PAS	0.011350
idade_Entre 18 e 20	0.010907
idade_Entre 21 e 25	0.010469
cota_Escola Pública Baixa Renda	0.009144
genero_M	0.008105
cota_Candidato Negro	0.007211
genero_F	0.006872
idade_Entre 26 e 30	0.003287
idade_Mais de 30	0.000931

Os resultados do Randon Forest, podem ser observados na Tabela 4.7, onde os valores indicam a importância das features no algortmo e representam o peso de cada característica na decisão do modelo. Valores mais altos indicam que a *feature* tem um impacto maior nas previsões do modelo. Por exemplo, um valor de 0.619856 para IRA significa que essa característica contribui significativamente mais para as decisões do modelo em comparação com outras. Já valores menores indicam que a *feature* tem menos influência nas previsões.

Destacando as importâncias das principais *feature* vemos que o IRA é a mais influente, com uma importância de 0.619856, indicando que o desempenho acadêmico é o principal fator na evasão dos alunos. A disciplina APC vem em seguida, com 0.100668, sugerindo que dificuldades nessa disciplina também são significativas. As categorias de cota, tipo de escola do segundo grau, e forma de ingresso também têm papéis importantes, mas em menor escala.

Ao comparar os resultados dos modelos preditivos Random Forest, SVM e Gradient Boosting, é possível observar variações importantes na precisão e na relevância das características para prever a evasão ou conclusão dos alunos. O GB apresentou a maior precisão, com 87,41%, e destacou o IRA como a característica mais determinante, com um F Score de 408, sugerindo que o desempenho acadêmico é o principal indicador para prever a evasão. Além disso, a forma de ingresso na universidade também se mostrou relevante, com um F Score de 89, indicando que diferentes processos seletivos influenciam diretamente a probabilidade de conclusão ou abandono.

No modelo SVM, a precisão geral foi 84,26%, com uma diferença entre a capacidade de prever corretamente a conclusão (76%) e a evasão (90%). Isso sugere que o SVM tem melhor desempenho na identificação de casos de evasão. Já o RF apresentou uma acurácia similar ao GB, de 87,04%, com 81% de precisão nas previsões de conclusão e 91% nas previsões de evasão. O GB destaca-se por sua precisão e pela relevância do IRA, enquanto o RF equilibra uma boa acurácia e uma maior sensibilidade à evasão. O SVM, teve bom desempenho na previsão de conclusões, mas se sobressaiu na detecção de evasões.

Capítulo 5

Conclusão

Após uma ampla análise das bases de dados do SIGRA e SIGAA disponibilizados pela UnB, e investigação do histórico dos alunos do curso de Engenharia de Computação, dividida em períodos de início do curso, desenvolvimento do aluno durante o curso, conclusão e evasão, é possível chegar a uma conclusão sobre os itens levantados na introdução deste trabalho.

Constatou-se o claro efeito positivo que a Lei das Cotas exerceu sobre o contexto academico da UnB, pois ao final quase 10 anos após o decreto da lei, a presença de outros grupos sociais é de fato comprovada. Observa-se que, no período anterior à Lei 12.711/2012, havia uma grande diferença entre o número de cotistas e não cotistas, especialmente entre alunos provenientes de escolas públicas e particulares. Era nítida a desigualdade presente no contexto acadêmico. Felizmente, essa realidade mudou e, paralelamente, percebe-se que o principal objetivo da implementação da Lei vem sendo alcançado: a promoção da diversidade e igualdade racial nas universidades.

Por ser um curso que possui uma ementa abrangendo matérias de engenharia elétrica, computação, matemática e física, pode-se considerar que o nível de dificuldade dos primeiros semestres do curso de Engenharia de Computação é bastante elevado. As disciplinas iniciais, como Cálculo 1 e 2, Álgebra Linear e Física 1 e 2, têm altos índices de reprovação, tornando a fase introdutória do curso um grande desafio para os novos alunos. Conforme observado neste trabalho, o IRA (Índice de Rendimento Acadêmico) é o indicador mais relacionado com a evasão. Ou seja, se o aluno reprova frequentemente e possui um IRA baixo, ele tem uma grande probabilidade de evadir. Esse cenário é particularmente observado nos semestres iniciais, levando à conclusão de que a grande quantidade de reprovações no início do curso justifica uma parte significativa das evasões apresentadas neste trabalho.

A dificuldade das matérias do curso também levanta dúvidas acerca do desempenho dos alunos cotistas que ingressam no curso. Contudo, observou-se que os resultados ten-

dem a se igualar com o passar dos anos, indicando que alunos cotistas e não cotistas possuem dificuldades similares, alcançando números de reprovações e aprovações muito semelhantes. Este fato ajuda a entender que o lugar de origem do aluno não afeta diretamente a probabilidade de evasão, reforçando assim, o impacto positivo da Lei das Cotas na universidade.

Também foi observada a grande necessidade de ações voltadas ao incentivo de meninas no curso de Engenharia de Computação. Notou-se ainda que existem fatores internos e externos à universidade que influenciam negativamente a presença de mulheres no meio acadêmico tecnológico. Os números indicam que essa influência é significativa, destacando a urgência de implementar medidas para promover a inclusão e retenção de mulheres nesse curso.

As análises realizadas evidenciam o impacto dessa política pública na promoção de diversidade e inclusão no ambiente acadêmico. Além disso, essas observações contribuem para identificar melhorias tanto na estrutura da universidade quanto na metodologia e na ministração das disciplinas. Dessa forma, o estudo visa contribuir para o aprimoramento do curso, almejando alcançar uma realidade acadêmica que, de maneira igualitária, beneficie todos os grupos sociais atendidos pela Universidade de Brasília.

Para trabalhos futuros indica-se avaliar também como é o desempenho dos alunos nas materias finais do curso de Engenharia de computação. Podendo realizar também a mesma análise em materias consideradas chave no curso como Organização e Arquitetura de Computadores, Estrutura de Dados e Sistemas Operacionais.Além disso, um estudo focado na disciplina de APC (Algoritmos e Programação de Computadores) seria interessante, uma vez que essa matéria demonstrou ter uma forte influência na predição de evasão dos alunos, destacando a necessidade de uma atenção especial a esse aspecto.

Referências

[1] Yan Shi, Tianxu Zhang: *Feature analysis: support vector machine approaches*, 2001. ix, 12, 13, 46

[2] Shahhosseini, Mohsen e Guiping Hu: *Improved Weighted Random Forest for Classification Problems*, página 42–56. Springer International Publishing, 2021, ISBN 9783030665012. http://dx.doi.org/10.1007/978-3-030-66501-2_4. ix, 14

[3] Cunha, Eglaisa: *Sistema universal e sistema de cotas para negros na unb: Um estudo de desempenho*. 2006. 1

[4] Carvalho, José Jorge de: *Usos e abusos da antropologia em um contexto de tensão racial: o caso das cotas para negros na unb*. Horizontes antropológicos, 11(23):237–246, 2005, ISSN 0104-7183. 1

[5] Ketulhe, Kailany, Maristela Holanda, Alice Lima, Alice Borges, Aleteia Araujo, Carla Castanho, Carla Koike e Roberta B Oliveira: *Análise do desempenho acadêmico das alunas cotistas na primeira disciplina de programação da universidade de brasília*. Em *Anais do XVI Women in Information Technology*, páginas 1–11. SBC, 2022. 1, 3

[6] Velloso, Jacques: *Vestibular com cotas para negros na unb, candidatos e aprovados nos exames*. (23), 2005. 1

[7] CARVALHO, J. J.de.; SEGATO, R.L: *Plano de metas para a integraçãosocial, étnica e racial da universidade de brasília*. Mimeo, (3), 2002. 1, 2, 20, 25

[8] Assunção, Amanda Vanessa Pereira, Catarina de Almeida Santos e Danielle Xabrega Pamplona Nogueira: *Política de cotas raciais na unb: um estudo sobre o acesso de negros na universidade durante o período 2004 a 2012*. Revista HISTEDBR on-line, 18(1):212–233, 2018, ISSN 1676-2584. 2, 20, 24

[9] Santos, Claudia Regina dos e João Hagenbeck Parizzi: *Dilemas raciais brasileiros: o racismo estrutural e os limites e as perspectivas da lei nº 12.711/2012 / brazilian racial dilemmas: the structural racism and the limits and perspectives of law 12.711/2012*. Revista Educação e Políticas em Debate, 9(Especial), 2020, ISSN 2238-8346. 2, 26

[10] Holanda, Maristela, Marília Dantas, Gustavo Couto, Jan Correa, Aleteia Patrícia de Araújo e Maria Emília Walter: *Perfil das alunas no departamento de computação da universidade de brasília*. Em *Anais do XI Women in Information Technology*,

páginas 1208–1212, Porto Alegre, RS, Brasil, 2017. SBC. https://sol.sbc.org.br/index.php/wit/article/view/3408. 3

[11] Araujo, Aleteia, Maristela Holanda, Carla Koike, Roberta Oliveira, Carla Castanho *et al.*: *Meninas. comp: trabalhando a diversidade de gênero na região central do brasil*. https://gredos.usal.es/handle/10366/152006, 2023. 3

[12] Figueiredo, Rejane M da C, Luiz CM Ribeiro Jr, Cristiane S Ramos e Edna Dias: *Graduação em engenharia de software versus graduação em engenharia de computação: uma reflexão*. FEES-Fórum em Educação de Engenharia de Software. SBES-Simpósio Brasileiro de Engenharia de Software (SBES), 2010. 3

[13] Encyclopaedia Britannica, The Editors of: *Data analysis*. Encyclopaedia Britannica, 2024. https://www.britannica.com/science/data-analysis, Acesso: Agosto de 2024. 7

[14] *Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective*. SN Computer Science, 2024. https://link.springer.com/article/10.1007/s42979-020-00326-1, Acesso: Agosto de 2024. 7

[15] Method, Research: *Data analysis - process, methods and types*, 2024. https://researchmethod.net/data-analysis-process-methods-and-types/, Acesso: Agosto de 2024. 7

[16] Simitsis, Alkis e Panos Vassiliadis: *From conceptual design to performance optimization of etl workflows: current state of research and open problems*. The VLDB Journal, 17(4):200–220, 2008. 7

[17] Vassiliadis, Panos e Alkis Simitsis: *Extraction, transformation, and loading*. Em *Handbook of Data Warehousing*, páginas 200–223. Springer, 2010. 8

[18] Silva, Fabiano Couto Corrêa da: *Visualização de dados: passado, presente e futuro | data vizualization: past, present and future*, 2019. 8

[19] Hansen, Luiza A.: *Análise visual de dados educacionais: Um estudo de caso das disciplinas introdutórias de programação da unb*. Tese de Mestrado, Programa de Pós-graduação em Informática Universidade de Brasília, 2021. 8

[20] Hebert Ramos, Vladimir Diniz: *Alternativas locacionais de empreendimentos utilizando aprendizado de máquina*, 2022. 9

[21] Lindholm, Andreas, Niklas Wahlström, Fredrik Lindsten e Thomas B. Schön: *Machine Learning - A First Course for Engineers and Scientists*. Cambridge University Press, 2022. https://smlbook.org. 9

[22] *Scikit learn user guide*, Acesso em: Julho, 2024. https://scikit-learn.org/stable/user_guide.html. 10

[23] Linlan Liu, Mingxiao Niu, Chao Zhang Jian Shu: *Light gradient boosting machine-based link quality prediction for wireless sensor networks*, 2022. 10, 46

[24] *Ensembles: Gradient boosting, random forests, bagging, voting, stacking*, Acesso em: Julho, 2024. https://scikit-learn.org/stable/modules/ensemble.html#gradient-boosting. 11, 13, 46

[25] *Cross-validation: evaluating estimator performance*, Acesso em: Julho, 2024. https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation. 11

[26] *Plot classification boundaries with different svm kernels*, Acesso em: Julho, 2024. https://scikit-learn.org/stable/auto_examples/svm/plot_svm_kernels.html#sphx-glr-auto-examples-svm-plot-svm-kernels-py. 13, 49

[27] Cheng Xu, Jing Wang, Tianlong Zheng Yue Cao Fan Ye: *Prediction of prognosis and survival of patients with gastric cancer by weighted improved random forest model*, 2021. 13

[28] Rosane Rossato Binotto, Marcus Vinícius Maltempi, Rogério Aparecido Batista da Silva: *Potencialidades da programação em python para o desenvolvimento do pensamento criativo em matemática*, 2023. 18

[29] Beatriz Martins Pereira, Camila Solange Moreno Maldonado Godoi, Thayna Barros Viana Rafael Medeiros Hespanhol: *Estudo comparativo da aplicaÇÃo dos programas python e orange para a anÁlise aprofundada de bancos de dados*, 2021. 18

[30] Bernardo Rurik Aparecido Gomes, Jhon Cruz, Valéria Jordani de Oliveira Silva Carlos da Cunha Sena Francisco Nairon Monteiro Júnior: *Power bi para tomada de decisÕes estratÉgicas: AnÁlise de indicadores-chave de desempenho (kpis)*, 2023. 19

[31] Veselina Naneva, Kremena Stefanova: *Programming with dax*, 2022. 19

[32] Vanegas, Claudia Elena Durango, Juan Camilo Giraldo Mejía, Fabio Alberto Vargas Agudelo e Dario Enrique Soto Duran: *A representation based on essence for the crisp-dm methodology.* Computación y Sistemas (CyS), 27(3), 2023. 19

[33] *Brasília - darcy ribeiro*, Acesso em: Setembro, 2024. https://www.unb.br/campi/darcy-ribeiro. 19

[34] Moreira, Jacqueline de Oliveira Moreira, Leandro Bento da Silva, Karinne Vieira de Jesus e Rodrigo Goes e Lima: *Do parcial das cotas ao inteiro das polÍticas afirmativas: Uma leitura psicanalÍtica sobre a atualidade da lei 12.711/2012.* Psicologia em revista, 27(2):570–589, 2021, ISSN 1677-1168. 26

[35] Stephenson, Chris, Alison Derbenwick Miller, Christine Alvarado, Lecia Barker, Valerie Barr, Tracy Camp, Carol Frieze, Colleen Lewis, Erin Cannon Mindell, Lee Limbird *et al.*: *Retention in computer science undergraduate programs in the us: Data challenges and promising interventions.* ACM, 2018. 34

[36] Luxton-Reilly, Andrew, Ibrahim Albluwi, Brett A Becker, Michail Giannakos, Amruth N Kumar, Linda Ott, James Paterson, Michael James Scott, Judy Sheard e

Claudia Szabo: *Introductory programming: a systematic literature review.* Em *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, páginas 55–106, 2018. 34

[37] Becker, Brett A e Keith Quille: *50 years of cs1 at sigcse: A review of the evolution of introductory programming education research.* Em *Proceedings of the 50th acm technical symposium on computer science education*, páginas 338–344, 2019. 34

[38] Andrea Felippe Cabello, Tiago Medina Chagas: *ReprovaÇÕes e evasÃo: Uma anÁlise com base na metodologia do inep*, 2021. 38, 40

[39] Paulo Lima, Cynthia Bisinoto, Nilce Santos de Melo Mauro Luiz Rabelo: *Taxas longitudinais de retenção e evasão: uma metodologia para estudo da trajetória dos estudantes na educação superior*, 2019. 39