



UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE - UERN
UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO - UFERSA
PROGRAMA DE PÓS-GRADUAÇÃO
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

CAMYLE MICAELY SILVEIRA PEIXOTO

**EVASÃO ACADÊMICA NAS INSTITUIÇÕES PÚBLICAS: UM ESTUDO
DE CASO NO CAMPUS DA UFERSA-ANGICOS**

MOSSORÓ

2024

CAMYLE MICAELY SILVEIRA PEIXOTO

**EVASÃO ACADÊMICA NAS INSTITUIÇÕES PÚBLICAS: UM ESTUDO
DE CASO NO CAMPUS DA UFERSA-ANGICOS**

Dissertação de Mestrado Acadêmico
apresentado ao Programa de Pós-Graduação
em Ciência da Computação – PPgCC da
Universidade Federal Rural do Semi-Árido e
Universidade do Estado do Rio Grande do
Norte como requisito para obtenção do título
de Mestre em Ciência da Computação.

Linha de Pesquisa: Otimização e Inteligência
Computacional

Orientador: Sebastião Emidio Alves Filho.
Prof. Dr. UERN

MOSSORÓ

2024

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Central Orlando Teixeira (BCOT)
Setor de Informação e Referência (SIR)

A large empty rectangular box with a black border, intended for the CIP data.

Bibliotecário-Documentalista
Nome do profissional, Bib. Me. (CRB-15/10.000)

CAMYLA MICAELY SILVEIRA PEIXOTO

**EVASÃO ACADÊMICA NAS INSTITUIÇÕES PÚBLICAS: UM
ESTUDO DE CASO NO CAMPUS DA UFERSA-ANGICOS**

Dissertação de Mestrado Acadêmico
apresentado ao Programa de
Pós-Graduação em Ciência da
Computação – PPgCC da Universidade
Federal Rural do Semi-Árido e
Universidade do Estado do Rio Grande
do Norte como requisito para obtenção
do título de Mestre em Ciência da
Computação.

Linha de Pesquisa: Otimização e
Inteligência Computacional

Defendida em: 19 / 12 / 2024.

BANCA EXAMINADORA

Sebastião Emidio Alves Filho, Prof. Dr. (UERN)
Presidente

Angélica Félix de Castro, Profa. Dra. (UFERSA)
Membro Examinador Interno

Lenardo Chaves e Silva, Prof. Dr. (UFERSA)
Membro Examinador Interno

Ismenia Blavatsky de Magalhães, Profa. Dra. (UFRN)
Membro Examinador Externo

Dedico este trabalho, que representa a conclusão de mais uma etapa importante da minha vida, à professora Ana Lúcia dos Santos (In Memoriam), carinhosamente conhecida como Lucinha, que me ensinou a superar as limitações da deficiência e a acreditar no meu potencial. Sua defesa pela inclusão e educação de qualidade foi um marco em minha jornada, e ao concluir esta dissertação, honro seu legado.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pela força e inspiração que Seu nome representa em minha vida, sustentando-me em cada conquista diária.

A minha família meu mais profundo reconhecimento pelo apoio incondicional em todas as minhas decisões, pelo incentivo constante e encorajamento, que sempre renovam minha força de vontade para alcançar meus objetivos.

Expresso minha gratidão aos professores da graduação, que foram fundamentais na minha formação e me prepararam para alcançar este título. Um agradecimento especial à professora Windiz Ferreira, por seu auxílio nas correções do projeto de seleção do mestrado, contribuindo de forma significativa para o meu ingresso no programa.

Ao meu orientador, registro minha mais sincera gratidão pelos ensinamentos, paciência e dedicação ao longo de toda a minha trajetória acadêmica no mestrado. À professora Ceres por ter me acolhido no período do estágio docência, contribuindo para o desenvolvimento das atividades com autonomia e leveza.

Agradeço à Banca Examinadora, por sua análise criteriosa, contribuições valiosas e por todo o aprendizado que proporcionaram neste momento tão significativo.

Aos amigos que estiveram ao meu lado durante este período, especialmente Ronaldo e Aureliano, colegas de mestrado, que me auxiliaram nas disciplinas e compartilharam experiências enriquecedoras. A Ednilson que contribuiu bastante no percurso final do mestrado.

Um agradecimento especial à minha amiga Micaele Cavalcante, cuja assistência na inscrição para a seleção foi crucial. Mesmo diante de dificuldades técnicas, ela não desistiu e sempre me motivou a seguir em frente. Minha gratidão também a João Rodrigues, por estar sempre à disposição para me ajudar, e à família de Lucinha, por ter aceito ser citada na dedicatória.

Ensinar não é transferir conhecimento e sim
criar as possibilidades de apreensão.
Paulo Freire

RESUMO

O Ensino Superior pode proporcionar uma gama de oportunidades na vida e na carreira dos discentes. Entretanto, nem todos que iniciam seus estudos chegam a concluí-lo, pelas mais diversas causas. É papel dos gestores universitários desenvolver políticas de permanência estudantil para evitar, de forma eficiente e eficaz, o abandono de curso. Contudo, cada estudante tem seus desafios particulares, que podem variar de acordo com o local em que fazem seus estudos, e isso deve ser levado em consideração para tentar prever quais discentes têm maior chance de evadir. A presente pesquisa tem como objetivo analisar os principais fatores causadores da evasão universitária por uma perspectiva mais pessoal do olhar do discente e seus desafios cotidianos, com foco no Campus da Ufersa em Angicos/RN. Foram estudados possíveis fatores associados à evasão, destacando problemas financeiros, conciliação entre trabalho e estudo, o deslocamento até a universidade e questões relacionadas à inclusão. O estudo desenvolveu uma abordagem qualitativa, integrando análise documental, aplicação de questionários aos discentes e o emprego de técnicas de Aprendizagem de Máquinas e Mineração de Dados para a análise e previsão das causas da evasão. Os resultados evidenciaram que a evasão pode ser influenciada por fatores externos, como a dificuldade para se deslocar até a sede do Campus, a necessidade de trabalhar para se sustentar, barreiras de acessibilidade, entre outros. Com base nas conclusões obtidas, foi desenvolvido um modelo computacional com acurácia de 100%, para alguns casos, podendo auxiliar os gestores na identificação e previsão das causas de evasão no ensino superior. Além disso, o estudo aponta a necessidade de investigações futuras, aprofundando a compreensão sobre os fatores que impactam a permanência no ensino superior.

Palavras-chave: Evasão acadêmica. Ensino superior. Aprendizagem de Máquinas. Mineração de Dados. Análise Exploratória de Dados.

ABSTRACT

Higher education can provide a range of opportunities in the lives and careers of students. However, not everyone who begins their studies ends up completing them, for a variety of reasons. It is the role of university administrators to develop student retention policies to efficiently and effectively prevent students from dropping out of courses. However, each student has their own particular challenges, which may vary according to where they study, and this should be taken into account in order to try to predict which students are most likely to drop out. This research aims to analyze the main factors that cause university dropout from a more personal perspective, from the perspective of students and their daily challenges, focusing on the UFERSA Campus in Angicos/RN. Possible factors associated with dropout were studied, highlighting financial problems, balancing work and study, commuting to university, and issues related to inclusion. The study developed a qualitative approach, integrating document analysis, application of questionnaires to students, and the use of Machine Learning and Data Mining techniques to analyze and predict the causes of dropout. The results showed that dropout rates can be influenced by external factors, such as difficulty in getting to the campus, the need to work to support oneself, accessibility barriers, among others. Based on the conclusions obtained, a computational model was developed with 100% accuracy in some cases, which can help managers identify and predict the causes of dropout rates in higher education. In addition, the study points to the need for future research to deepen the understanding of the factors that impact retention in higher education.

Keywords: Academic dropout rates. Higher education. Machine learning. Data mining. Exploratory data analysis.

SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 MOTIVAÇÃO.....	15
1.2 OBJETIVOS.....	17
1.2.2 Objetivo geral.....	17
1.2.3 Objetivos específicos.....	17
1.3 QUESTÕES DE PESQUISA.....	17
1.4 HIPÓTESES.....	18
1.5 ESTRUTURA DO DOCUMENTO.....	19
2 FUNDAMENTAÇÃO TEÓRICA.....	20
2.1 EVASÃO NO ENSINO SUPERIOR.....	20
2.2 MACHINE LEARNING.....	22
2.3 MINERAÇÃO DE DADOS.....	26
2.4 MÉTRICAS PARA AVALIAÇÃO DE MODELOS DE MACHINE LEARNING e MINERAÇÃO DE DADOS.....	29
3 REVISÃO SISTEMÁTICA DE LITERATURA.....	32
3.1 PLANEJAMENTO.....	32
3.1.1 Questões de Pesquisa.....	32
3.1.2 Processo de Busca.....	33
3.1.3 Critérios para Seleção e Avaliação.....	34
3.2 RESULTADO DA REVISÃO.....	35
3.2.1 Descrição dos Trabalhos Seleccionados.....	36
3.2.2 Respostas às Questões de Pesquisa.....	42
4 METODOLOGIA.....	54
4.1 CONTEXTUALIZAÇÃO.....	54
4.2 ELABORAÇÃO DO QUESTIONÁRIO.....	57
4.3 APLICAÇÃO DOS QUESTIONÁRIOS.....	61
5 RESULTADOS.....	62
5.1 DADOS COLETADOS.....	62
5.2 ANÁLISE SOBRE MORADIA E TRANSPORTE.....	63
5.3 ANÁLISE SOBRE INCLUSÃO, PRECONCEITOS E VOCAÇÃO.....	66
5.4 ANÁLISE SOBRE EMPREGO, RENDA E TEMPO DE DEDICAÇÃO.....	68
5.5 ANÁLISE SOBRE A INTENÇÃO DE SE EVADIR.....	71
5.6 MINERAÇÃO DOS DADOS DOS QUESTIONÁRIOS.....	74

5.6.1. Tratamento e Limpeza de Dados.....	75
5.6.2. Aplicação de Modelos de Machine Learning.....	78
5.7 DISCUSSÃO ACERCA DA QUESTÃO GERAL DE PESQUISA E HIPÓTESES	
81	
6. CONCLUSÃO.....	83
7. REFERÊNCIAS.....	85
APÊNDICE A - QUESTIONÁRIO APLICADO COM OS PARTICIPANTES.....	94

LISTA DE ILUSTRAÇÕES

Figura 1: Evolução Média dos Indicadores de Trajetória dos Ingressantes em Cursos de Graduação no Brasil - 2011 a 2020.....	15
Figura 2: Comparação entre os dados de Ingresso e Saída de discentes do Campus da UFRSA-Angicos.....	16
Figura 3: Exemplo de aplicação do algoritmo KNN.....	23
Figura 4: Processo introdutório de uma árvore de decisão.....	24
Figura 5: Processo do KDD.....	28
Figura 6. Quantidade de abandonos de curso no Campus Angicos - UFRSA.....	58
Figura 7: Relação entre curso e semestre de ingresso.....	62
Figura 8: Relação entre a moradia e o transporte se é um problema na permanência do curso.....	65
Figura 9: Boxplot do tempo médio no deslocamento até a universidade.....	66
Figura 10 : Análise de experiências: preconceito e violência no contexto educacional...	67
Figura 11: Percentuais dos estudantes trabalham.....	69
Figura 12: Jornada de trabalho dos estudantes.....	69
Figura 13 : Contribuição na renda familiar pelos estudantes.....	70
Figura 14: Discentes casados e não casados.....	70
Figura 15: Estudantes com e sem filhos.....	71
Figura 16: As principais dificuldades para permanecer no curso.....	72
Figura 17: Visão Geral das Dificuldades Acadêmicas e Fatores de Evasão.....	79

LISTA DE QUADROS

Quadro 1: Características das aprendizagem supervisionada, não supervisionada e por reforços.....	25
Quadro 2: Métricas de avaliação nos modelos de ML e MD baseados nas técnicas de classificação.....	30
Quadro 3: Termos de busca para RSL.....	33
Quadro 4: As causas de evasão das universidades públicas encontradas nos trabalhos selecionados.....	42
Quadro 5: As bases de dados usadas para obter resultados da evasão acadêmica dos últimos anos encontradas nos trabalhos selecionados.....	44
Quadro 6: Aplicações para prever os riscos de evasão por cada período dos cursos de Ensino Superior encontrados nos trabalhos selecionados.....	46
Quadro 7: Modelos mais rápidos e eficazes para investigar os dados de evasão acadêmica encontrados nos trabalhos selecionados.....	47
Quadro 8: Os tipos de dados de evasão encontrados nos trabalhos selecionados.....	49
Quadro 9: Resultados informados aos stakeholders pelos trabalhos selecionados.....	51
Quadro 10: Perguntas e Respostas sobre a Análise dos Dados Encontrados.....	76
Quadro 11: Resultado do PyCaret com a variável tempo de deslocamento.....	81
Quadro 12: Resultado do PyCaret sem a variável tempo de deslocamento.....	82

1 INTRODUÇÃO

A Educação Superior refere-se à especialização do ser humano. A busca por um curso de ensino superior se dá, sobremaneira, pelo destaque nas redes de trabalho e a possibilidade da aquisição de diversas competências para a área de atuação desejada. Nessa percepção, de acordo com a Lei de Diretrizes e Bases, n.º 9394 de dezembro de 1996 (Brasil, 1996).

Art. 43. A Educação Superior tem por finalidade:

I - estimular a criação cultural e o desenvolvimento do espírito científico e do pensamento reflexivo;

II - formar diplomados nas diferentes áreas de conhecimento, aptos para a inserção em setores profissionais e para a participação no desenvolvimento da sociedade brasileira, e colaborar na sua formação contínua;

III - incentivar o trabalho de pesquisa e investigação científica, visando o desenvolvimento da ciência e da tecnologia e da criação e difusão da cultura, e, desse modo, desenvolver o entendimento do homem e do meio em que vive [...].

Contudo, dados de 2019 mostravam que 74% da população com 25 anos ou mais de idade possuía ensino básico completo ou equivalente, mas apenas 17,4% com o Ensino Superior completo (IBGE, 2019).

Na vida acadêmica os estudantes têm um compromisso elevado em relação à integração aos estudos, pois a responsabilidade é maior, onde, muitas vezes pela falta da disponibilidade ou recursos financeiros, acabam sendo desestimulados para concluírem os cursos (Silva, 2020). Para agravar essa situação, durante o percurso acadêmico surgem vários obstáculos que podem levar à evasão, caracterizando-se como um fenômeno de alunos que interrompem a trajetória institucional antes do prazo previsto.

No Ensino Superior o abandono ocorre frequentemente tanto nas redes públicas, como nas privadas. Quando as desistências advêm de estudantes das Instituições de Ensino Superior (IES) públicas, os recursos investidos não são retornados para a sociedade. Já no setor privado, isso é tido como uma perda de tempo e recursos financeiros para as famílias dos desistentes, além de, muitas vezes, também envolver recursos públicos. Em ambos os casos, a evasão é uma fonte de inércia de professores, funcionários, equipamentos e espaço físico que impacta no desenvolvimento do país como um todo (Silva Filho *et al.*, 2007).

O abandono dos estudos não se deve exclusivamente a dificuldades de compreensão de conteúdos, é necessário levar em consideração os fatores externos relacionados ao aluno, tais como: vocação, aspectos socioeconômicos e problemas de ordem pessoal (Paredes, 1994). Entender suas principais causas tem sido o enfoque de diversas pesquisas (Oliveira, Bezerra

Torres, 2021; Paredes, 1994; Rolim e Almeida, 2021; Saccaro, França e Jacinto, 2019; Silva Filho, 2007; Silva, 2021; Silva e Freitas, 2022; Wegner, 2022).

Após toda investigação outros autores foram mais a fundo, utilizando de softwares a fim de sintetizar dados que auxiliassem na compreensão da evasão de estudantes como os trabalhos de Lopes, Carvalho e Lahm (2016) e Santos *et al.*, (2019), que analisaram as causa de evasão promovendo modelos computacionais como os modelos de regressão logística múltipla e a implantação de um *Data Warehouse* para buscar uma solução capaz de auxiliar a gestão acadêmica na identificação de padrões que impactam na evasão no Ensino Superior.

Pela sua importância, isso se tornou uma preocupação para as políticas públicas na área da Educação. Há anos, políticas voltadas para a ampliação do acesso e permanência no ensino superior vêm sendo efetivadas, em destaque as iniciativas como o Plano de Reestruturação e Expansão das Universidades Federais (REUNI) em 2007, o Sistema Nacional de Avaliação da Educação Superior (SINAES) em 2004 e o Plano Nacional de Assistência Estudantil (PNAES) em 2010.

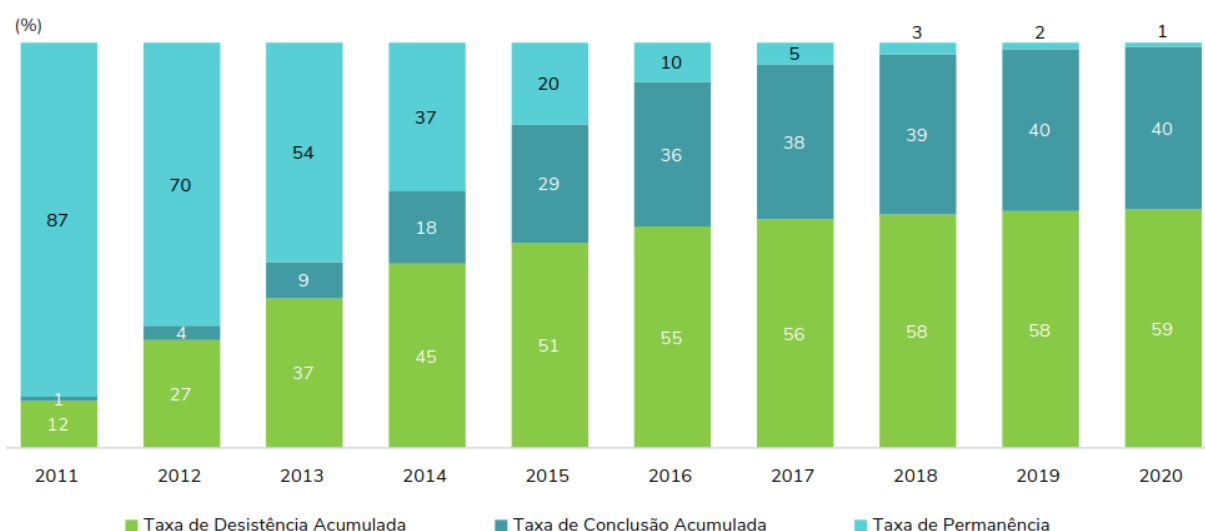
As IES também têm se modificado para proporcionar um ambiente mais igualitário para a formação humana e transformação em uma sociedade mais inclusiva, que venha a diminuir preconceitos, bullying e racismo pelo mundo. Para Dias Sobrinho (2010), entre as principais ações inclusivas que o MEC (Ministério da Educação) brasileiro impulsiona, o mais questionado é o programa de cotas que busca dar um olhar diferenciado aos pobres, indígenas, negros, pardos, quilombolas e Pessoas Com Deficiência (PCD). A Lei n. 12.711, de 29 de agosto de 2012 (Brasil, 2012), busca amenizar a desigualdade entre as minorias no país, que gradativamente foram excluídas da comunidade acadêmica realizando o que é chamado de reparação histórica. Ou seja, é uma forma de realizar a reparação das desigualdades práticas da vida social e estudantil, permitindo a atuação igualitária das diversas partes da sociedade, em total respeito às suas diferenças. Tomim (2016) comenta sobre a importância do professor de ter o compromisso educacional e social, semelhante aos atos de Paulo Freire, fazendo o máximo para que seus alunos não abandonem as novas oportunidades de se empoderar dos benefícios de uma educação transformadora.

Porém, é percebido que a situação vem se agravando nos últimos anos. De acordo com os dados publicados pela Universidade Federal da Integração Latino-Americana (UNILA,

2019)¹, em 2019 as universidades federais tinham 18,3 milhões de discentes, e contavam com a média de evasão de 14,57%. Em pesquisa realizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) foi constatado que no ano de 2020 mais de 8,6 milhões de discentes se matricularam no Ensino Superior, porém, nos últimos anos, apenas 40% (quarenta por cento) dos ingressantes concluíram seus estudos (INEP, 2022). Os outros entraram nas estatísticas do abandono universitário, conforme dados mostrados na Figura 1.

O presente estudo será direcionado à Universidade Federal Rural do Semi-Árido (UFERSA), sediada na cidade de Mossoró-RN desde 1967. Além do campus Central, a referida universidade é composta pelos campi de Caraúbas, Pau dos Ferros e Angicos. O foco desta dissertação, pelo fato da autora ser egressa de lá, é no campus Angicos.

Figura 1: Evolução Média dos Indicadores de Trajetória dos Ingressantes em Cursos de Graduação no Brasil - 2011 a 2020.



Fonte: (INEP, 2021)

1.1 MOTIVAÇÃO

O campus UFERSA Angicos² é uma unidade que agrega estudantes de vários setores regionais. Hoje ele oferece os cursos de graduação de: Bacharelado em Ciência e Tecnologia; Bacharelado em Sistemas de Informação; Engenharia Civil; Engenharia de Produção;

¹ Disponível em:

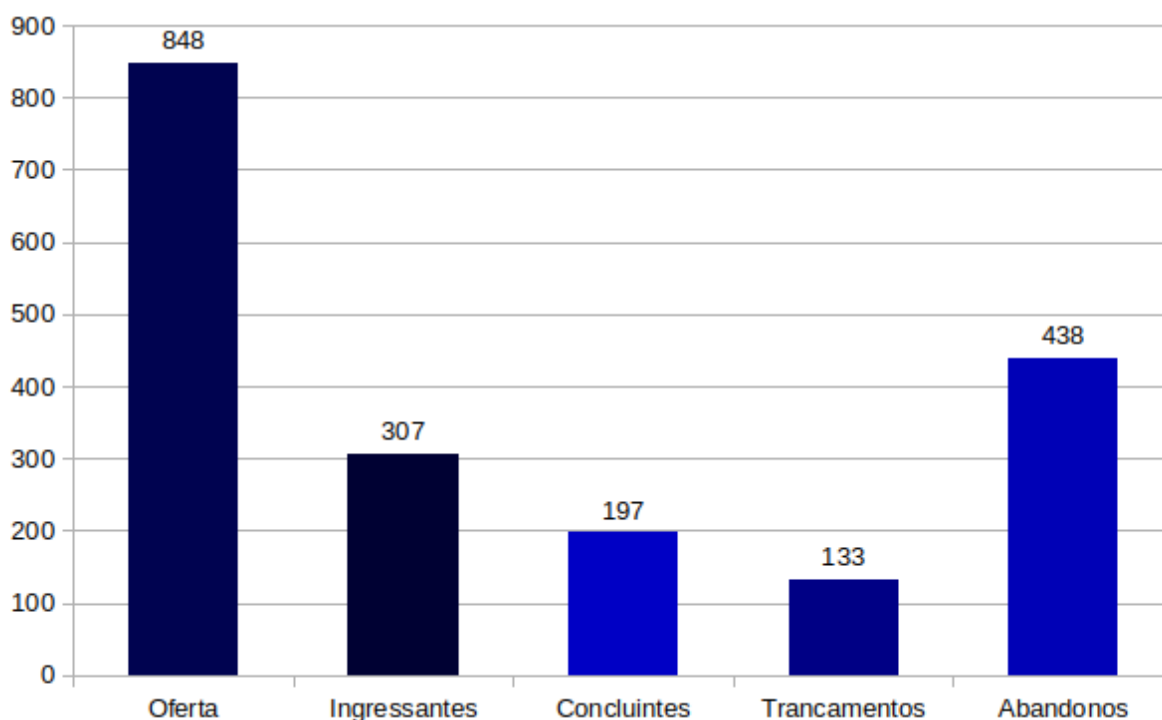
<https://portal.unila.edu.br/reitoria/espaco-reitoria/dados-sobre-a-taxa-de-evacao-nas-universidades-federais>. Acesso em: 16 set. 2023.

² Página eletrônica da UFERSA Campus Angicos: <https://angicos.ufersa.edu.br/>. Acesso em: 16 set. 2023.

Licenciatura em Computação e Informática; e Pedagogia. A universidade disponibiliza algumas assistências necessárias para os discentes, mas mesmo assim a evasão educacional continua presente por meio de vários fatores. O campus está localizado em uma região urbana pequena em que, geralmente, o público estudantil não reside na cidade em que o campus se encontra.

Com base nesse conhecimento prévio sobre a universidade, a presente pesquisa também se debruçou em verificar os registros da UFERSA. Apenas para exemplificar o problema, os seguintes dados foram extraídos por meio da Divisão de Registro Acadêmico-DRA (UFERSA, 2023). Como amostra apresentada na Figura 2, o ano de 2023 (incluindo os semestres 2023.1 e 2023.2) apresentou a oferta de 848 vagas, das quais apenas 307 foram ocupadas. Em contrapartida, apenas 197 discentes concluíram seus cursos, enquanto 133 realizaram o trancamento e 438 abandonaram seus estudos. Isto significa que, neste ano, mais discentes desistiram dos cursos que se interessaram por fazê-los no Campus Angicos, o que é uma situação bastante preocupante.

Figura 2: Comparação entre os dados de Ingresso e Saída de discentes do Campus da UFERSA-Angicos.



Fonte: Divisão de Registros Acadêmicos da UFERSA- Campus Angicos (2023)

Diante desse cenário, o motivador principal é a contribuição no entendimento sobre a evasão no ensino superior ao estudar as causas que levam o aluno optar pelo cancelamento do seu curso. O ponto de partida foi a experiência pessoal da mestranda, uma ex-aluna do

campus que residia em um município vizinho. Ao vivenciar as barreiras enfrentadas pelos estudantes, a mestranda decidiu realizar uma análise detalhada para identificar os fatores reais que contribuem para a evasão. A experiência da autora deste trabalho enquanto discente revela que alguns motivos que podem influenciar a permanência dos discentes são: a adaptação inicial e o não se identificar com o curso superior, somados com a existência de obstáculos externos, como o deslocamento, comportamento social e comunicação. Portanto, a intenção deste trabalho é contribuir com a redução da evasão universitária na instituição através da compreensão das suas causas e, assim, construir subsídios que possam inspirar a promoção de ações de permanência estudantil.

1.2 OBJETIVOS

1.2.2 Objetivo geral

A presente pesquisa tem como objetivo geral identificar, através da análise de questionários aplicados aos discentes e da criação de um modelo preditivo de *Machine Learning*, perfis de estudantes que tenham maior probabilidade de se evadirem dos cursos da UFERSA - Campus Angicos, contribuindo com os gestores na tomada de decisão para combater este problema.

1.2.3 Objetivos específicos

- Compreender, por meio da literatura, os fatores que influenciam a evasão do ensino superior que estão presentes na UFERSA Campus - Angicos;
- Avaliar as necessidades e dificuldades dos discentes, assim como as ações de permanência que o campus realiza;
- Utilizar a metodologia computacional de Mineração de Dados como ferramenta para análise das respostas dos questionários utilizados na pesquisa, com foco na evasão universitária.

1.3 QUESTÕES DE PESQUISA

Em meio às motivações apresentadas, foram aplicadas as metodologias descritiva e explicativa, que são utilizadas para identificar as características de um determinado contexto. A descritiva usa o nível aprofundado sobre o conhecimento, fazendo a observação sistemática e empírica no interesse da pesquisa escolhida. A explicativa é aquela que envolve análise por um indivíduo ou grupo de indivíduos podendo usar métodos de entrevistas (Gomes; Gomes, 2019). Utilizando a abordagem *Design Science* (Wieringa, 2014), foi elaborado um conjunto de questões de pesquisa a ser respondido com a execução do trabalho. Diante disso, a questão de pesquisa que norteia este trabalho é:

Questão Geral de Pesquisa (QGP): Como identificar os fatores que causam a evasão de forma precoce dos discentes presentes nos cursos de graduação do Campus de Angicos da UFRSA e utilizá-los para definir estratégias de prevenção para esse fenômeno?

Em conformidade com o problema de pesquisa, a Questão Geral de Pesquisa pode-se fragmentar em mais três grupos de perguntas: Questões Conceituais (QC), Questões Inovadoras (QI) e Questões Práticas (QP). As QC apresentam as necessidades da problemática e o porquê dos fatos ocorrem dessa maneira. As QI pontuam as inovações aplicadas depois do uso das tecnologias, fundamentando em propor um modelo de predição em virtude da evasão. Por fim, as QP têm a finalidade de conhecer as práticas adequadas para usar no cotidiano atual. Tais questões são descritas a seguir.

Questões Conceituais (QC):

- **QC1** – Quais são as causas mencionadas na literatura em relação ao que ocorre na evasão acadêmica?
- **QC2** – Quais as fontes dos dados (as bases) e em que formatos estão disponíveis?

Questões Inovadoras(QI):

- **QI1** - O que pode ser aplicado para predizer os discentes com maior risco de evasão a cada período?
- **QI2** – Quais modelos computacionais mais rápidos e eficazes usados?

Questões Práticas(QP):

- **QP1** – Quais os tipos de dados (sociais, financeiros, acadêmicos) que melhor explicam esse fenômeno da evasão acadêmica?

- **QP2** – Como os resultados são apresentados aos *stakeholders*?

1.4 HIPÓTESES

As hipóteses a serem investigadas neste trabalho são:

1. O maior índice de evasão ocorre nos primeiros períodos;
2. Os discentes com dificuldade em se locomover até o/ou no campus têm tendência a desistir dos seus estudos;
3. Discentes que são responsáveis por famílias têm propensão a se evadirem;
4. Discentes de cursos das áreas de exatas e tecnológicas têm maior risco de abandonarem seus cursos;
5. Discentes que apresentam alguma dificuldade de conciliar os estudos com o trabalho estão mais propensos à evasão.

1.5 ESTRUTURA DO DOCUMENTO

O trabalho está estruturado da seguinte maneira:

Capítulo 2 - Referencial Teórico: norteia a respeito da temática da evasão estudantil, caracterizando os principais trabalhos encontrados na literatura tratando sobre os atributos da evasão institucional, como também adentrar nos conceitos das ferramentas, os algoritmos que trabalham com Mineração de Dados (MD) acompanhado do *Machine Learning* (ML), método de avaliação, o grau e fatores utilizados para destacar as causas de evasão.

Capítulo 3 - Trabalhos Relacionados: são listados trabalhos que discutem a mesma temática desta pesquisa, através de uma revisão sistemática de literatura, avaliando os resultados encontrados e identificando lacunas e tendências no campo de estudo.

Capítulo 4 - Metodologia: apresenta as etapas, os dados encontrados, técnicas e ferramentas utilizadas na pesquisa, desde a coleta e tratamento dos dados até a análise.

Capítulo 5 - Resultados: são apresentados os resultados obtidos nesta pesquisa, com o intuito de explorar a discussão relacionada aos experimentos que podem ser realizados nos cursos oferecidos pela UFERSA/Angicos.

Capítulo 6 - Conclusão: são apresentadas as considerações finais sobre os principais resultados de pesquisa, suas implicações para os cursos da UFERSA/Angicos e sugestões para reduzir a evasão acadêmica.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados conteúdos que fundamentam a proposta, revelando-se os aspectos relacionados à Evasão no Ensino Superior, *Machine Learning* e Mineração de Dados.

2.1 EVASÃO NO ENSINO SUPERIOR

A evasão no ensino superior ainda é um problema nas universidades brasileiras, principalmente nas instituições públicas. É um fenômeno complexo, por se tratar de uma questão dos discentes que está além das competências de ensino, pesquisa e extensão (Wegner, 2022). Essa situação, muitas vezes, não é levada em consideração nos planejamentos dos departamentos acadêmicos. Apenas em setores específicos de atendimento aos alunos.

Para discorrer sobre a evasão no ensino superior no contexto da evasão, é preciso compreender suas dimensões diante da educação brasileira. Podem haver várias causas que implicam na desistência na formação dos discentes. Para Dias Sobrinho (2010), a evasão está mais relacionada a questões pessoais, do que vinculada à qualidade do ensino ofertado naquele determinado ambiente.

A problemática muitas vezes é conduzida pela questão do desconhecimento do mundo universitário, de não se identificarem com a área, pelos recursos financeiros e pela sobrecarga fora do âmbito acadêmico (Lopes, Carvalho e Lahm, 2016). Diante desse contexto, foram criados o Programa de Bolsa Permanência do Ministério da Educação e as bolsas de pesquisa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ). Todos esses incentivos visam oportunizar uma garantia de renda mínima aos estudantes para sua permanência no ensino superior.

Com as criações do Fundo de Financiamento ao Estudante do Ensino Superior (FIES), o Programa Universidade para Todos (PROUNI) e o Sistema de Seleção Unificada (SISU) a quantidade de matriculados nos cursos passou de 4,6 milhões mas, ainda assim, no Brasil o número de graduados ainda é considerado baixo comparado a outros países (Saccaro *et al.*, 2019; Silva e Freitas, 2022).

No decorrer dos anos a evasão vem ocorrendo em áreas onde as demandas são inovações tecnológicas e aumento de produtividade. Moraes Filho (2021) associa que isso acontece principalmente nas turmas iniciais, relacionadas aos desafios perante as novas didáticas dos cursos que exigem mudança de hábito, submetendo os alunos a um bloqueio. Souza, Ramos e Serafim (2021) apontam que, nos últimos anos, o quadro de evasão tem mudado em meio à COVID-19, identificando um problema a mais, aplicado no baixo letramento digital, no qual alimenta a evasão educacional.

Em meio ao contexto vivenciado, surgiram incertezas, dúvidas, medo de como iria ser essa formação, aos docentes provocou a necessidade inequívoca de se reinventar e inovar as estratégias pedagógicas, preservando, ao mesmo tempo, a qualidade do ensino para que seus educandos não sentissem motivados a desistir (Santos *et al.*, 2020). Note-se que aumenta a preocupação com este tema.

No Plano Nacional de Educação (PNE) edição 2014-2024, as metas 12 e 13 correspondem ao Ensino Superior. São elas:

Meta 12: “Elevar a taxa bruta de matrícula na Educação Superior para 50% (cinquenta por cento) e a taxa líquida para 33% (trinta e três por cento) da população de 18 (dezoito) a 24 (vinte e quatro) anos, assegurada a qualidade da oferta e expansão para, pelo menos, 40% (quarenta por cento) das novas matrículas, no segmento público.”

Meta 13: “Elevar a qualidade da Educação Superior e ampliar a proporção de mestres e doutores do corpo docente em efetivo exercício no conjunto do sistema de Educação Superior para 75% (setenta e cinco por cento), sendo, do total, no mínimo, 35% (trinta e cinco por cento) doutores.” (BRASIL, 2014)

Contudo, alcançar essas metas 12 e 13 requererá um grande esforço no planejamento de todas as etapas da educação brasileira. O projeto de expansão e interiorização das universidades brasileiras, assim como os programas de permanência estudantil sofreram cortes contínuos (PROIF FEDERAÇÃO, 2023). Esta realidade dificultou ainda mais o acesso ao ensino superior e a continuação dos estudantes nos cursos (Santos, *et al.*, 2020).

Por meio dos relatos dos autores, destaca-se que mesmo com as buscas inovadoras, a evasão continua vigente. Para que esses problemas diminuam ou sejam eliminados, este estudo foi realizado com a intenção de sugerir soluções que possam resolver os obstáculos apresentados.

Diante da persistência da evasão no ensino superior, torna-se necessário buscar soluções inovadoras que integrem tecnologias avançadas ao planejamento educacional. Nesse

contexto, o uso de ferramentas tecnológicas, como o *Machine Learning*, surge como uma abordagem promissora para analisar dados, identificar padrões e propor estratégias de intervenção mais eficazes. A próxima seção apresenta os fundamentos do *Machine Learning* e sua aplicabilidade no enfrentamento desses desafios.

2.2 MACHINE LEARNING

Machine Learning (aprendizado de máquina em português), é considerada como uma subárea da Inteligência Artificial que lida com criações automatizadas no modelo por meio de um aprendizado individual ou dedutivo (Santos, 2024). O surgimento do *Machine Learning* (ML) se deu através da evolução da Inteligência Artificial (IA). Pesquisadores vinham se interessando em construir máquinas que pudessem desenvolver algoritmos capazes de aprender com os erros e realizar uma predição em meio ao conjunto de dados que está sendo disponibilizado para fazer experiências (Machado Segundo, 2022).

Há tipos de aprendizagem no subcampo ML que se divide em:

- **Aprendizagem supervisionada:** Nesse tipo de aprendizagem, os algoritmos são aplicados por meio de observações rotuladas, ou seja, a partir de dados de entrada e saída (NAU *et al.*, 2020). A título de exemplo, são apresentados dois algoritmos bem conhecidos:
 - **KNN:** Significando *K-Nearest Neighbors*, é um dos algoritmos que pode ser usado para classificação e regressão na aprendizagem supervisionada. Trabalhando com vizinhos mais próximos, classifica de acordo com as suas classes, no qual possibilita realizar comparação ponderada em peso. Em resumo, o algoritmo é trabalhado com amostras que estão posicionadas nos espaços de atributos e a distância, mediante às demais amostras de treinamento que são calculadas (Falqueto *et al.*, 2019). Para compreender melhor essa classificação, é demonstrado um exemplo na Figura 3, em que as instâncias já estão classificadas (azul e vermelho) e uma nova amostra de teste é inserida para verificar a qual categoria ela pertence de acordo com os *K* vizinhos mais próximos.

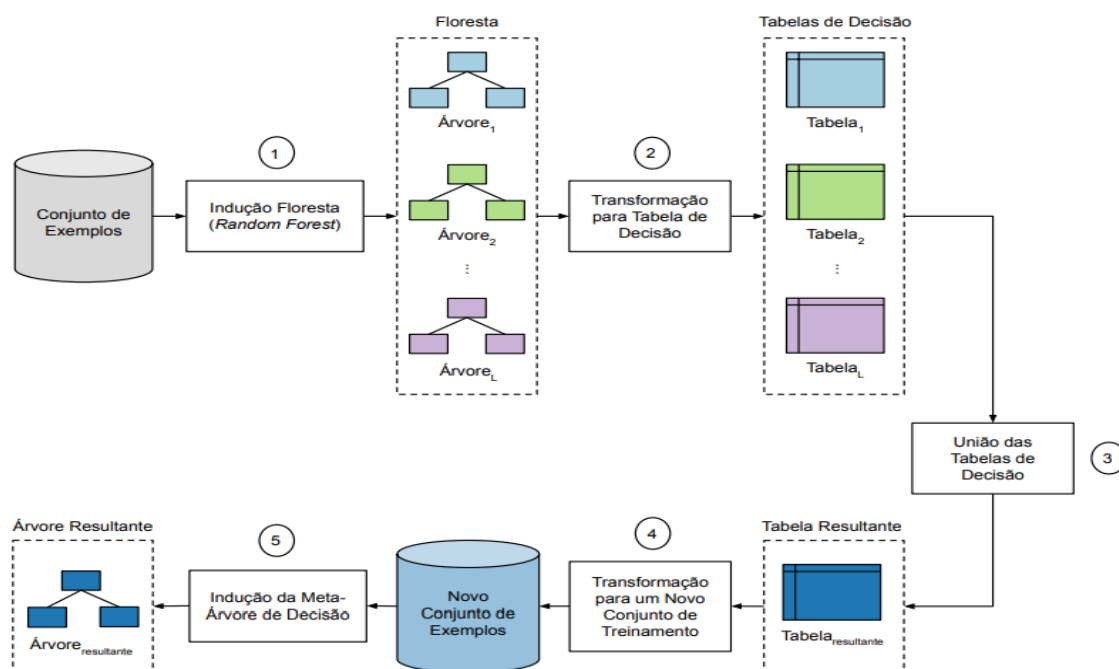
Figura 3: Exemplo de aplicação do algoritmo KNN.



Fonte: Falqueto *et al.*, (2019).

- Random Forest (Floresta Aleatória):** Trata-se de um algoritmo de classificação, onde agrega vários modelos de árvore de decisão. Tem como finalidade estruturar uma base de dados de forma aleatória, por meio da distribuição de diversos algoritmos de árvore de decisão e utilizando um subconjunto de atributos relacionais. Esse modelo consiste na intenção de fazer a agregação dos classificadores sob árvore de decisão, sendo construído de forma para que a estrutura seja aplicada de maneira aleatória (Lorenzetti & Telöcken, 2016). Na Figura 4 é ilustrado um processo introdutório de uma árvore de decisão que é conduzido pelo *Random Forest*. Nele, pode-se notar que o algoritmo parte de um conjunto de árvores (floresta) que são treinadas isoladamente e, após isso, têm seus modelos reunidos em uma tabela de decisão única que resulta em um modelo com uma única meta-árvore.

Figura 4: Processo introdutório de uma árvore de decisão.



Fonte: Ferreira, (2022).

- **Aprendizagem não supervisionada:** Essa categoria tem como objetivo aplicar algoritmos com apenas dados de entrada, sem nenhum tipo de rótulo. Com isso, os dados são agrupados por base de familiaridade, permitindo encontrar padrões de comportamento, detectando situações de interesse. O trabalho de Oliveira *et al.*, (2022), por exemplo, mostra o uso de algoritmos de mineração de dados educacionais (EDM) para agrupar os alunos de um curso à distância de biologia em três diferentes perfis de engajamento (desengajado, mediano e engajado) de acordo com suas interações no ambiente virtual de aprendizagem.
- **Aprendizagem por reforços:** É uma técnica de aprendizagem que consiste em aprender a partir de uma série de reforços, ou seja, ela trabalha por base de treinamento, selecionando as ações disponíveis naquele determinado estado para obter respostas concretas definindo qualidade nas ações futuras (Silva, 2019).

O Quadro 1 apresenta algumas características que definem a aprendizagem supervisionada, a não supervisionada e a por reforços, baseadas no trabalho de Nascimento (2022). Neste trabalho utilizaremos apenas a aprendizagem supervisionada.

Quadro 1: Características das aprendizagem supervisionada, não supervisionada e por reforços

Aprendizagem Supervisionada	Aprendizagem não supervisionada	Aprendizagem por reforços
Realizar treinamentos com os dados rotulados.	Realizar treinamentos com os dados não rotulados.	Realizar treinamento em frequência.
Identificar o <i>feedback</i> para verificar se está obtendo a saída desejada.	Não recebe nenhum <i>feedback</i> de verificação.	<i>Feedback</i> usado para ajustar os pesos para a próxima iteração.
Prevê os resultados de saída.	Obtém os dados de padrões ocultos.	Obtém os dados das últimas iterações.
Estabelece padrões nos modelos de saída	São obtidos somente os dados de entrada	Padrões definidos nos pesos.
Os dados de supervisão têm como objetivo treinar o modelo para fornecer novos dados.	Têm o objetivo de encontrar padrões ocultos e <i>insights</i> úteis no conjunto de dados.	O erro é usado para obter o padrão para as próximas entradas.

Fonte: Autoria própria (2024).

É utilizado para a problemática deste trabalho um modelo de aprendizado supervisionado. Em uma perspectiva de fazer uma avaliação nos algoritmos que são adotados para realizar análise, as principais classes de algoritmos utilizadas são: Classificação e Regressão.

- **Classificação:** pode ser entendida como uma forma de solucionar problemas se enquadrando no aprendizado supervisionado, com o objetivo de fazer o treinamento do modelo com dados onde é informado a que categoria em cada observação pertence e, após isso, classificar as novas observações nas classes corretas. O formato desse tipo de algoritmo tem a finalidade de prever se os dados presentes estão adequados para a classificação pertencente (Netto & Maciel, 2021).
- **Regressão:** também é ligada ao aprendizado supervisionado, entretanto, a diferença da classificação é que o objetivo de identificar a categoria é de forma contínua, podendo ser uma probabilidade, em meio da previsão de valores numéricos. A regressão pode

ser utilizada em um ou mais atributos com variáveis contínuas e dependentes (Netto & Maciel, 2021).

Para atingir o resultado desejado é preciso conhecer como esses dados são aplicados, entender os atributos utilizados no qual os conjuntos de dados são selecionados e divididos em dois tipos: quantitativos e qualitativos. O quantitativo consiste em medidas numéricas, já o qualitativo não considera as entradas numéricas como rótulo ou atributo, ele está mais ligado a classificar categorias, ou seja, nominais ou ordinárias (Günther, 2006).

2.3 MINERAÇÃO DE DADOS

Na seção anterior foi discutida a importância da classificação e da regressão de dados, utilizando os métodos de aprendizagem, no qual possibilitará verificar as circunstâncias que os estudantes evadem-se à universidade. Nesta seção é abordada a Mineração de Dados (MD), mostrando seus conceitos básicos fundamentais.

Para Gualhano e Vasconcelos (2020), a Mineração de Dados é um tratamento para analisar e extrair padrões de informações de grandes bases de dados. No trabalho deles, a grande parte desses dados estão relacionados a dados sobre evasão. Em Carrano *et al.*, (2019), a Mineração de Dados é utilizada em diferentes táticas para compreender as causas da evasão institucional. O crescimento é rápido dos volumes das bases que estão sendo trabalhadas em tamanho e dimensão, criando a necessidade de extrair conhecimentos sobre elas.

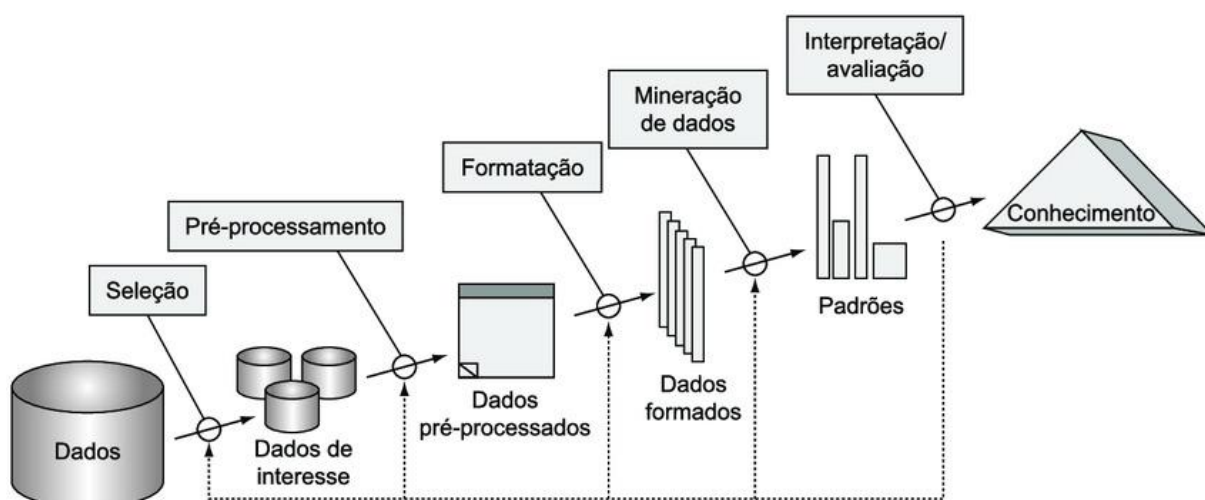
Uma das metodologias mais utilizadas é a descoberta de conhecimento em bancos de dados, o KDD (no inglês *Knowledge-Discovery in Databases*). Ela agrega um conjunto de técnicas que tem o objetivo de analisar padrões de informações pertencentes à base de dados de grande porte (Lopes, Carvalho e Lahm, 2016). Ele também é definido como um processo de dados que tem o intuito de compreender as etapas que estão presentes no momento, com o objetivo de facilitar a análise dos padrões descobertos. Ele foi criado em 1989, com o objetivo de usar processo para descoberta de informações úteis sobre os dados, focando em aplicações de alto nível. Os sistemas de KDD são vistos como ferramenta que podem ser interativas, no qual não é considerado como um sistema de análise automático (Collazos, Barreto e Pellegrini, 2016).

Especificando melhor, o KDD é um processo que cria relações de interesse que não são observáveis pelos analisadores do assunto. Esse processo é composto por cinco etapas (Fayyad *et al.*, 1996): Seleção dos dados; Pré-processamento e limpeza de dados; Transformação dos dados; Mineração de dados; Interpretação de avaliação sobre os dados, conforme processo ilustrado na Figura 5 .

Na primeira etapa, os dados brutos coletados são selecionados de modo que apenas as características mais relevantes para o estudo são utilizadas, tentando também manter um balanceamento entre as diferentes classes e faixas de valores. Esses dados passam por um pré-processamento, no qual há a correção e/ou remoção de observações com dados ausentes e/ou inconsistentes, objetivando evitar vieses ou distorções. Os dados pré-processados são, então, transformados de modo a poderem ser melhor visualizados e tratados pelos algoritmos de ML. São utilizadas técnicas como conversão de dados categóricos para numéricos usando algum processo de codificação (como o *Label Encoding* e o *One Hot Encoding*), padronização, conversão de formatos (em datas, por exemplo), entre outros.

Os dados transformados passam, então, pelo processo de MD, onde podem ser feitas análises exploratórias, estatísticas e previsões com métodos de ML em busca de padrões existentes nos dados. Finalmente, na etapa de avaliação, os padrões e informações encontradas são apresentadas aos interessados no processo (*stakeholders*), resultando em conhecimento que pode ser utilizado para resolução de problemas e apoio na tomada de decisão.

Figura 5: Processo do KDD.



Fonte: Fayyad *et al.*, (1996)

Em Gualhano, Vasconcelos (2020), as etapas do KDD são utilizadas na área da Educação, apresentando demandas diante dos dados com ferramentas e metodologias, dando a possibilidade de realizar uma análise e uma interpretação sobre os problemas de ensino e aprendizado.

Quando adaptada para a investigação de dados educacionais, como a evasão dos discentes nos cursos, a técnica é chamada de Mineração de Dados Educacionais (do inglês *Educational Data Mining* - EDM), em que se torna mais amplo o desenvolvimento em algoritmos, na concepção de conhecer e coletar dados que estão disponíveis nos fenômenos educacionais (Castro e Ferrari, 2016). A respeito da evasão dos discentes, a EDM vem sendo utilizada com modelos preditivos e de correlação para possibilitar os achados dos motivos de evasão de instituições de ensino, desenvolvendo uma tarefa fundamental para a análise de dados e possibilitando a extração de resultados.

2.4 MÉTRICAS PARA AVALIAÇÃO DE MODELOS DE *MACHINE LEARNING* e MINERAÇÃO DE DADOS

Os algoritmos de ML e MD são conhecidos como tecnologias que vêm sendo eficientes para a previsão dos dados. Com base nisto, surge a necessidade de utilizar as métricas para avaliação que possibilitem uma comparação no desempenho, como também suas limitações. Deste modo, destacam-se algumas métricas de avaliação, constantemente utilizadas pelas comunidades de ML e MD para fazer comparação entre dois algoritmos ou mais que poderão ser aplicados neste estudo.

As Métricas de Classificação podem ser adaptadas tanto para classificação binária como nas múltiplas classes. As tarefas de classificação são trabalhadas para prever qual é a categoria de amostra pertencente. Algumas métricas que podem ser citadas são:

- ***Confusion Matrix* ou Matriz de Confusão:** é uma métrica que fornece objeto visual que são calculados através de matrizes que descrevem erros sistemáticos nas previsões dos modelos de classificação, ou seja, é uma matriz que sumariza os resultados que estão em análise (Netto & Maciel, 2021). Por exemplo, nos resultados de uma turma de disciplina, a matriz de confusão pode possibilitar a visualização dos números de aprovados e reprovados obtidos por meio dos algoritmos de aprendizagem de máquina

que são realizados para prever a quais grupos pertencem. Na matriz de confusão são considerados quatro tipos de valores, de acordo com o que o algoritmo prevê que irá ocorrer (positivo) ou não (negativo), e se ele acerta a previsão (verdadeiro) ou não (falso):

- VP - verdadeiro positivo (*true positive*): percentual predito de maneira positivamente correta no total que será positivo;
- VN - verdadeiro negativo (*true negative*): percentual predito negativo de maneira corretamente no total que será negativo.
- FP - falso positivo (*false positive*): percentual que é predito positivo diante do total que se encontrava negativo.
- FN - falso negativo (*false negative*): percentual que é predito negativo diante do total que estava positivo.

Algumas métricas são derivadas a partir dos resultados da matriz de confusão, sendo a acurácia a mais conhecida e simples de compreender pois dá o percentual de previsões corretas em relação ao total. Contudo, outras métricas, como a precisão (*precision*) e a revocação (*recall*), também podem ser derivadas (Netto & Maciel, 2021). No Quadro 2 são apresentadas estas métricas juntamente com as fórmulas e descrições.

Quadro 2: Métricas de avaliação nos modelos de ML e MD baseados nas técnicas de classificação

Métrica	Fórmula	Descrição
<i>Precision</i>	$VP / (VP + FP)$	A precisão de modo geral é usada para medir valores previstos corretamente de modo positivo (VP), considerando a relação entre os verdadeiros positivos (VP) e a soma de todos os valores previstos como positivos, incluindo os falsos positivos (FP).
<i>Recall</i>	$VP / (VP + FN)$	Usado para medir modelos de frações utilizando amostra para detecção de valores verdadeiros positivos (VP) em relação a todas as previsões corretas, verdadeiros positivos (VP) e verdadeiros negativos (VN).

Fonte: Adaptado de Nascimento (2022)

Na avaliação de desempenho com modelos de regressão, algumas métricas usadas, por exemplo, são:

- Erro Quadrático Médio (*Mean Squared Error - MSE*): é uma métrica utilizada por ML, que realiza cálculos em forma de raiz quadrática média com a função de medir os erros dos valores observáveis através da Equação 1:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 \quad (1)$$

- Erro Médio Absoluto (*Mean Absolute Error - MAE*): é uma métrica trabalhada em ML, semelhante a métrica anterior onde leva a diferença em prever o modelo em valor real ao quadrado, tornando no final um valor absoluto como mostrado na Equação 2:

$$MAE(y, p) = \frac{1}{N} \sum_{i=1}^N |y_i - p_i| \quad (2)$$

Entretanto, além das métricas apresentadas, existem outras que podem ser aplicadas de acordo com cada circunstância.

3 REVISÃO SISTEMÁTICA DE LITERATURA

As tendências conceituais e técnicas das pesquisas deste trabalho partem de uma Revisão Sistemática da Literatura (RSL). A RSL se caracteriza em delimitar uma visão geral, avaliando e interpretando a problemática que está sendo discutida (Pereira, Cysneiros e Aguiar, 2019). Com isso, neste capítulo é apresentada a RSL no intuito de identificar na literatura trabalhos desenvolvidos relacionados à temática da evasão.

Por se tratar de um estudo sistemático, a RSL deve seguir por um protocolo tanto planejado como fundamentado, de modo que possibilite o processo ser replicado e livre de viés. Nesta RSL, o protocolo empregado foi baseado no trabalho de Oliveira, Bezerra e Torres (2021) que define a sua realização em seis etapas: i) planejamento; ii) definição as fontes de busca; iii) seleção de trabalhos relevantes; iv) avaliação dos trabalhos; v) extração dos dados; vi) síntese e redação para responder às questões de pesquisa.

3.1 PLANEJAMENTO

Nesta seção serão especificadas as questões de pesquisa, as fontes de busca e os critérios que estão envolvidos no processo de seleção dos trabalhos.

3.1.1 Questões de Pesquisa

Baseado na descrição da Questão Geral de Pesquisa definida na Seção 1.3, a Questão de Pesquisa (QP) desta RSL se caracteriza em: “Como identificar os fatores que causam a evasão de forma precoce e que soluções foram utilizadas para combatê-lá?”. Para que a RSL possa contribuir com este trabalho, para responder a sua questão de pesquisa foram utilizadas as questões também definidas na Seção 1.3:

Questões Conceituais:

- **QC1** – Quais são as causas mencionadas na literatura em relação ao que ocorre evasão acadêmica?
- **QC2** – Quais as fontes dos dados (as bases) e em que formatos estão disponíveis?

Questões Inovadoras:

- **QI1** - O que pode ser aplicado para prever os discentes com maior risco de evasão a cada período?
- **QI2** – Quais modelos computacionais mais rápidos e eficazes usados?

Questões Práticas:

- **QP1** – Quais os tipos de dados (sociais, financeiros, acadêmicos) que melhor explicam esse fenômeno da evasão acadêmica?
- **QP2** – Como os resultados são apresentados aos *stakeholders*?

3.1.2 Processo de Busca

A pesquisa envolveu trabalhos publicados nos últimos dez anos, uma vez que os dados em análise nos estudos citados neste trabalho utilizam de informações para fins de comparação com os dias atuais e os desafios contemporâneos. A busca foi realizada em três fontes relevantes na área de tecnologia educacional, a citar:

- *IEEE Xplore Digital Library* (www.ieeexplore.ieee.org);
- SOL (<https://sol.sbc.org.br/index.php/indice>);
- Scielo (<https://www.scielo.br/>).

Estas fontes foram escolhidas com base na: i) disponibilidade em acessar o conteúdo dos artigos; ii) importância e consolidação das fontes, e iii) indexação de artigos publicados em periódicos, eventos e conferências voltadas para o objetivo desta pesquisa.

Com isso, para selecionar os trabalhos destas fontes com as expressões centrais que direcionam aos trabalhos buscados nesta pesquisa, foram definidos alguns termos mediante as áreas da Educação e Tecnologia, com ênfase nas estratégias dos estudos e na formação dos discentes do ensino superior. Os termos utilizados deveriam estar nos idiomas português ou inglês, levando em consideração que as fontes de pesquisa predominam os artigos científicos escritos em tais idiomas. Os termos definidos são apresentados no Quadro 3.

Quadro 3: Termos de busca para RSL.

Área	Termos
Educação	Evasão; Reprovação; Abandono.
Tecnologia	Algoritmo; Estatística; Dados.
Estratégias	Análise; Modelo; Predição.
Formação	Estudantes Universitários; Estudantes de Graduação; Estudantes de Ensino Superior.

Fonte: Autoria própria (2024).

Para realizar esse processo, uma *string* de busca foi definida com base nos termos especificados no Quadro 3, resultando na seguinte *string* de busca:

Em virtude das especificidades das fontes de busca, a *string* foi traduzida para o idioma inglês. Portanto, a *string* de busca no idioma inglês ficou da seguinte forma: ***(Evasion OR Reproach OR dropout) AND (Machine Learning OR Artificial Intelligence OR Data Mining OR Data Science) AND (Analysis OR Model OR Predicting) AND (University students OR Graduation Student OR Higher Education Students)***.

3.1.3 Critérios para Seleção e Avaliação

Na realização desta seleção, foram definidos dois conjuntos, os Critérios de Inclusão (CI) e os Critérios de Exclusão (CE), de forma que os trabalhos escolhidos pudessem responder às perguntas supracitadas.

Critério de Inclusão:

CI1) Pesquisa com foco na temática da evasão em cursos presenciais.

CI2) Publicações dos últimos dez anos.

CI3) Publicações nos idiomas português e inglês.

CI4) Pesquisa que aponte as causas da evasão.

CI5) Pesquisa que aborda um estudo analítico ou preditivo com estatística, mineração ou aprendizagem de máquina.

Critérios de Exclusão:

CE1) Trabalhos duplicados em mais de uma fonte de pesquisa.

CE2) Trabalhos que não sejam artigos completos ou capítulos de livro

CE3) Pesquisa que não fornece dados suficientes para responder às questões de pesquisa.

CE4) Pesquisa que não aponta as fontes dos dados.

3.2 RESULTADO DA REVISÃO

A RSL realizou-se em quatro fases. Os resultados foram organizados em um arquivo de planilhas³, adicionando quatro fases divididas por seis planilhas, onde foi distribuída cada etapa com duas seções, uma para os critérios incluídos e outra para os excluídos. A Fase I, norteou-se pela pesquisa dos artigos referentes à temática, fazendo uma pré-seleção dos trabalhos achados nas fontes de pesquisa, a partir dos retornos da execução da *string* de busca. A pré-seleção se concretizou através da leitura dos títulos que correspondiam aos Critérios de Inclusão CI1, CI2 e CI3, e foram descartados os trabalhos que apresentavam os Critérios de Exclusão CE1 e CE2, atribuindo os achados em uma única Tabela da Fase I.

Na Fase II, com os trabalhos selecionados na Fase I, além de ler os títulos, foram lidos os resumos, selecionados os estudos que atendeu ao CI4 e descartados os que satisfizeram o CE3. Observados os trabalhos que correspondiam aos critérios foram organizadas tabelas separadas, Tabela de Inclusão (TI2) e Tabela de Exclusão (TE2). Na execução final da Fase II, obteve 30 trabalhos selecionados.

Na Fase III, efetuou-se uma leitura rápida dos artigos, verificando as metodologias aplicadas por cada trabalho. Nessa fase, os que contribuíram com o CI5 foram adicionados na Tabela de Inclusão (TI3), já os estudos do CE4 foram movidos para Tabela de Exclusão (TE3). No final dessa Fase III, 23 trabalhos foram selecionados.

Na Fase IV, revisou-se os trabalhos selecionados da TI3. Deste modo, foi realizada uma leitura completa dos artigos e uma revisão geral sob os critérios de inclusão e exclusão. Concluídas essas fases, os trabalhos que corresponderam aos critérios satisfatórios foram direcionados para responder às QPs. da Subseção 3.1.1.

³ https://docs.google.com/spreadsheets/d/10qHA9VCI2fYWSyNISJSKFYogaZghUP5H_dxqNhixnKk/

3.2.1 Descrição dos Trabalhos Seleccionados

Nos trabalhos seleccionados foram extraídas informações que fortaleceram a pesquisa. Os autores contribuíram com essa pesquisa apresentando os elementos relacionados à evasão no ensino superior. As informações descritas nos artigos se constituíram da seguinte forma.

O estudo de Murillo, Costa e Vásquez (2023) propõe um modelo preditivo que combina Análise de Sobrevivência, Árvores de Decisão e Random Forest para diagnosticar precocemente possíveis fatores de evasão em estudantes universitários. A metodologia consiste em três fases: estimar a probabilidade de permanência do aluno, estabelecer padrões de sobrevivência com base nas variáveis consideradas e identificar variáveis críticas usando Random Forest. Embora desenvolvido em uma universidade chilena, o modelo pode ser adaptado a diferentes contextos acadêmicos, considerando as particularidades de cada instituição.

O artigo mediado por Salas e Caldas (2024) apresenta uma abordagem de aprendizado de máquina para prever o desempenho acadêmico de estudantes universitários no Peru. Utilizando dados institucionais, os autores aplicam modelos preditivos para antecipar o rendimento dos alunos, com foco na melhoria de estratégias educacionais e redução da evasão. A pesquisa demonstra que essas ferramentas podem auxiliar na tomada de decisões mais assertivas pelas instituições.

Osorio e Santacoloma (2023), apresenta um sistema de alerta precoce para identificar automaticamente estudantes do primeiro semestre com alto risco de evasão. Utilizando um modelo de aprendizado de máquina treinado com dados históricos de estudantes, o sistema alcançou uma sensibilidade de 61,97% na predição de alunos em risco, permitindo intervenções antecipadas para reduzir a taxa de evasão.

O estudo de Teodoro e Kappel (2020) aplicou técnicas de aprendizado de máquina aos dados do INEP para prever a evasão em instituições públicas de ensino superior no Brasil. A técnica Random Forest alcançou aproximadamente 80% de precisão, destacando idade, participação em atividades extracurriculares e carga horária total do curso como fatores determinantes na evasão.

O objetivo de Assis e Marcolino (2024) foi identificar estudantes com risco de evasão no curso de Licenciatura em Computação, especialmente em áreas não metropolitanas. Para

isso, foram analisados dados acadêmicos de alunos utilizando 15 algoritmos de aprendizado de máquina, buscando determinar quais modelos previam com maior precisão os casos de evasão. Além disso, o estudo identificou as variáveis mais influentes associadas ao abandono do curso. Os resultados demonstraram a eficácia da abordagem proposta na identificação de estudantes propensos à evasão, permitindo a implementação de intervenções preventivas.

Hortêncio Filho, Vinuto e Leal (2020) analisam a evasão escolar nos campi com maiores índices do Instituto Federal do Ceará. Utilizando técnicas de aprendizagem de máquina, o estudo define a evasão como um problema de classificação, aplicando modelos preditivos. A proposta visa fornecer subsídios para ações preventivas por parte da equipe pedagógica e institucional.

O estudo realizado por Souza, Komati e Andrade (2022), aborda a problemática da evasão no ensino superior. Utilizando técnicas de análise de sobrevivência e modelos preditivos, os autores investigaram os principais fatores que influenciam a evasão no curso de Sistemas de Informação de uma instituição pública brasileira. Os resultados indicam que o risco de evasão é mais elevado nos semestres iniciais, com destaque para as disciplinas dos dois primeiros semestres, responsáveis por reter cerca de 50% dos estudantes. Essas conclusões ressaltam a importância de intervenções precoces para mitigar a evasão e promover a permanência dos alunos.

Martins *et al.* (2023) analisam a evasão estudantil nos estágios finais da graduação em uma universidade pública, utilizando dados de 30.000 alunos. Por meio de algoritmos de aprendizado de máquina supervisionados, o estudo identifica fatores acadêmicos e socioeconômicos relevantes para prever a evasão tardia. Os resultados mostram que é possível antecipar o risco de abandono mesmo nas fases finais do curso. A pesquisa reforça a importância de políticas de intervenção específicas para garantir a permanência dos estudantes até a conclusão da formação.

Ramos e Araújo (2022) realizaram uma investigação sobre os principais fatores que levam à evasão de mulheres em cursos de computação na Região Metropolitana de Belém. A pesquisa utilizou questionários respondidos por alunas, analisados por meio de ferramentas como IRAMUTEQ e WEKA. Os dados revelaram que a falta de incentivo, o ambiente masculino predominante e episódios de discriminação influenciam fortemente na decisão de

abandono. O estudo ressalta a importância de políticas institucionais voltadas à permanência e inclusão feminina na área de tecnologia.

Na temática de Santo e Goya (2020) investiga a aplicação de técnicas de aprendizado de máquina para prever a evasão de estudantes em cursos de Sistemas de Informação. Utilizando dados acadêmicos e socioeconômicos, os autores desenvolveram modelos preditivos que identificam alunos com maior risco de abandono. Os resultados destacam que fatores como desempenho acadêmico, frequência às aulas e situação financeira são determinantes na previsão da evasão. A pesquisa enfatiza a importância de intervenções precoces e políticas institucionais direcionadas para mitigar a evasão e promover a retenção estudantil.

Santos, Martins e Plastino (2021), investigam a possibilidade de prever a evasão de estudantes no ensino superior utilizando exclusivamente o desempenho acadêmico. Aplicando técnicas de Mineração de Dados, especialmente classificadores, foram desenvolvidos modelos preditivos que, ao final de cada semestre, avaliam a probabilidade de um aluno evadir ou concluir o curso. Os experimentos, realizados com dados reais de uma universidade brasileira, alcançaram acurácias entre 79,31% e 98,25%, indicando que o desempenho acadêmico é um forte indicador para prever a evasão estudantil.

Barbosa *et al.* (2023) trata sobre a evasão escolar em uma instituição de ensino técnico federal por meio de técnicas de análise de dados e aprendizado de máquina. O estudo utiliza informações acadêmicas e socioeconômicas dos alunos para treinar modelos preditivos. A *Random Forest* foi o classificador com melhor desempenho. Os resultados indicam a viabilidade de intervenções preventivas baseadas nas previsões. A pesquisa contribui para a gestão educacional com foco na permanência estudantil.

O estudo de Viana, Santana e Rabêlo (2022), aborda a problemática da alta taxa de evasão em Instituições Federais de Ensino Superior (IFES). Utilizando técnicas de Mineração de Dados Educacionais, os autores propõem a classificação de alunos evadidos e graduados por meio de modelos treinados para cada janela semestral. A pesquisa destaca a importância de estratégias preditivas para identificar e mitigar o risco de evasão, auxiliando docentes e gestores educacionais na tomada de decisões.

Rafiq, Rabbi e Ahammad (2021) aplicam técnicas de ciência de dados para prever o risco de evasão semestral entre estudantes universitários em Bangladesh. Utilizando algoritmos como Random Forest e Regressão Logística, o estudo analisa dados acadêmicos

para identificar alunos propensos a abandonar o curso. Os resultados destacam a eficácia dessas abordagens preditivas na identificação precoce de estudantes em risco, permitindo intervenções oportunas para reduzir as taxas de evasão. O estudo enfatiza a importância de estratégias baseadas em dados para melhorar a retenção estudantil no contexto universitário de Bangladesh.

Santos *et al.* (2019) analisam o uso de algoritmos de aprendizado supervisionado no contexto de mineração de dados educacionais para prever a evasão de estudantes universitários. O estudo utiliza bases de dados institucionais e aplica classificadores como Random Forest e SVM para identificar alunos com maior risco de abandono. Os resultados mostraram boa acurácia dos modelos, possibilitando ações preventivas por parte das instituições. A pesquisa reforça a relevância do uso de inteligência artificial na tomada de decisões educacionais.

Aguirre e Pérez (2020) aplicam técnicas de análise preditiva de dados para identificar estudantes com maior risco de abandono no ensino superior. Utilizando algoritmos de aprendizado de máquina, o estudo analisa dados acadêmicos e demográficos. Os modelos permitem intervenções mais eficazes e precoces. Os resultados mostram alto desempenho na identificação de padrões de evasão. A proposta contribui para estratégias institucionais de permanência.

Deb *et al.* (2024) desenvolveram um modelo preditivo baseado em aprendizado de máquina para identificar alunos com risco de evasão escolar. Utilizando dados históricos e variáveis educacionais, o estudo testa diferentes algoritmos para melhorar a precisão. Os resultados demonstram que técnicas supervisionadas são eficazes na previsão precoce da evasão, permitindo intervenções preventivas por parte das instituições de ensino. A pesquisa destaca o potencial da inteligência artificial na formulação de políticas educacionais voltadas à retenção estudantil.

O estudo de Serrano *et al.* (2024) explora o uso de modelos de aprendizado de máquina para prever a evasão acadêmica em estudantes universitários. A pesquisa utiliza dados educacionais para treinar classificadores capazes de identificar alunos em risco de abandono. Os autores analisam diferentes algoritmos para comparar acurácia e eficiência. O modelo proposto visa auxiliar instituições no planejamento de intervenções preventivas. A abordagem oferece suporte à gestão educacional com foco na permanência estudantil.

O estudo de Tariq *et al.* (2023) compara os algoritmos KNN e Árvore de Decisão na previsão de abandono precoce de estudantes universitários. Utilizando dados acadêmicos, os

autores analisam qual modelo oferece melhor desempenho na identificação de alunos em risco. Os resultados mostram variações na precisão dos métodos conforme o conjunto de dados. A proposta visa permitir intervenções educacionais mais eficazes. A pesquisa contribui para estratégias de retenção no ensino superior.

O estudo de Fernández-García *et al.* (2020) apresenta um sistema de recomendação voltado para auxiliar estudantes do ensino superior na escolha de disciplinas. Utilizando dados do histórico acadêmico, o sistema sugere matrículas alinhadas ao perfil do aluno. A proposta busca melhorar a tomada de decisão, evitar reprovações e aumentar a permanência estudantil por meio de técnicas de aprendizado de máquina.

O estudo de Hossain *et al.* (2022) analisa o impacto da pandemia de Covid-19 na taxa de abandono escolar em Bangladesh, utilizando técnicas de análise preditiva. A pesquisa aplica algoritmos de aprendizado de máquina para identificar fatores de risco associados à evasão. Os resultados evidenciam a importância de intervenções educacionais durante crises para reduzir o abandono escolar e orientar políticas públicas mais eficazes.

O estudo de Santos *et al.* (2019) apresenta uma metodologia baseada em aprendizado de máquina para prever a evasão de estudantes universitários, especialmente em instituições públicas. O modelo combina árvores de decisão, algoritmos genéticos e amostragem estratificada por conglomerados, alcançando precisão superior a 95%. A pesquisa destaca que estudantes com média abaixo de 5,79 e mais de um ano de matrícula têm maior risco de evasão. Além disso, cerca de um terço dos abandonos ocorrem no primeiro ano. Os resultados podem subsidiar ações de monitoramento e políticas preventivas.

O estudo de Gopalakrishnan *et al.* (2017) utilizou técnicas de mineração de dados para identificar fatores que influenciam a persistência e a graduação de estudantes universitários. Foram aplicadas análises visuais, seleção de atributos e algoritmos de previsão. A pontuação em matemática de entrada (ELM) foi um dos principais preditores, e intervenções como cursos de recuperação foram recomendadas. O Naïve Bayes foi mais eficaz para prever graduação, enquanto AdaBoost e SVM destacaram-se na previsão de persistência.

3.2.2 Respostas às Questões de Pesquisa

Nesta seção são apresentados os resultados que foram encontrados nos trabalhos selecionados na RSL em relação às questões de pesquisa. Eles estão nos Quadros 4, 5, 6, 7, 8 e 9, cada um correspondendo a uma questão. É importante enfatizar que nem todos os estudos

estão relacionados a todas as questões, então alguns podem não aparecer em todos os quadros, de acordo com a temática tratada.

QC1 – Quais são as causas mencionadas na literatura em relação ao que ocorre evasão acadêmica?

Quadro 4: As causas de evasão das universidades públicas encontradas nos trabalhos selecionados.

TRABALHO	RESULTADO
Murillo <i>et al.</i> (2023)	Mal gerenciamento de tempo para estudos.
Salas e Caldas (2024)	Desfavorecimento socioeconômico.
Osorio e Santacoloma (2023)	No aspecto demográfico, são considerados elementos como idade em que se inicia a faculdade e gênero. A nível individual, entram em questão o contexto social, desempenho escolar anterior, o psicológico que envolve a motivação e as atitudes dos estudantes. Qualidade do ensino e do ambiente de aprendizagem, além das políticas de financiamento.
Teodoro e Kappel (2020)	principais aspectos psicológicos que motivam a decisão são descritos por 7 categorias: motivos institucionais, pessoais, relacionados à falta de suporte, relacionados à carreira, relacionados ao desempenho acadêmico, interpessoais e relacionados à autonomia.
Assis e Marcolino (2024)	Baixo desempenho acadêmico.
Hortêncio Filho <i>et al.</i> (2020)	Fatores institucionais como defasagens no acervo bibliográfico, ausência de locais adequados para estudo e convivência e ausência ou mal-estado de conservação dos equipamentos para práticas. Fatores pessoais como conjuntura socioeconômica do país no momento, desemprego, baixo rendimento escolar e a não identificação com o curso escolhido.
Martins <i>et al.</i> (2023)	fatores internos da instituição: infraestrutura deficitária, acervo desatualizado, métodos de avaliação docente e deficiência didático pedagógica dos professores; e fatores externos: dificuldades financeiras, escolha equivocada do curso, falta de base para acompanhar o curso escolhido, fato de ter sido admitido em um curso que não foi sua primeira opção.
Ramos e Araújo (2022)	Histórico social e cultural sobre as mulheres que as limitam.

Santo e Goya (2020)	Questões vocacionais, financeiras e sociodemográficas.
Barbosa <i>et al.</i> (2023)	Baixo desempenho acadêmico.
Rafiq, Rabbi e Ahammad (2021)	Desafios acadêmicos, problemas familiares/pessoais, posicionamento político.
Santos <i>et al.</i> (2019)	Falta de suporte educacional.
Aguirre e Pérez (2020)	Problemas de adaptação e incorporação do aluno no meio acadêmico, social e familiares.
Deb <i>et al.</i> (2024)	Dificuldades acadêmicas, questões financeiras, problemas pessoais.
Serrano <i>et al.</i> (2024)	Problemas pessoais e de adaptação.
Tariq <i>et al.</i> (2023)	Comportamento social, instabilidade financeira e distúrbios psicológicos.
Hossain <i>et al.</i> (2022)	Problemas de saúde e ansiedade.
Gopalakrishnan <i>et al.</i> (2017)	Histórico acadêmico ruim, falta de suporte financeiro e baixa escolaridade dos pais.

Fonte: Autoria própria (2024).

QC2 – De onde vêm os dados (as bases) e os formatos?

Quadro 5: As bases de dados usadas para obter resultados da evasão acadêmica dos últimos anos encontradas nos trabalhos selecionados.

TRABALHO	RESULTADO
Murillo <i>et al.</i> (2023)	Registros institucionais.
Salas e Caldas (2024)	Registros institucionais e dados dos censos do Ministério da Educação do Peru (MINEDU).
Osorio e Santacoloma (2023)	Registros institucionais tanto acadêmicos quanto de saúde.
Teodoro e Kappel (2020)	Banco de Dados da educação superior do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira).
Assis e Marcolino (2024)	Registros acadêmicos individuais dos alunos.

Hortêncio Filho <i>et al.</i> (2020)	Dados da plataforma IFCE em Números.
Souza <i>et al.</i> (2022)	Banco de dados institucionais com informações acadêmicas, demográficas e socioeconômicas.
Martins <i>et al.</i> (2023)	Os microdados de estudantes foram obtidos através de uma solicitação de informação no portal governamental FalaBR ² , com o número 23546.051594/2020-15.
Ramos e Araújo (2022)	Formulário eletrônico via Google Forms.
Santos e Goya (2020)	Censo da Educação Superior disponibilizado pelo INEP.
Santos <i>et al.</i> (2021)	Dos registros acadêmicos.
Barbosa <i>et al.</i> (2023)	Dados institucionais do IFPE - Campus Jaboatão dos Guararapes.
Viana, Santana e Rabêlo (2022)	Registros acadêmicos da UFPI.
Rafiq, Rabbi e Ahammad (2021)	Diversos bancos de dados institucionais.
Santos <i>et al.</i> (2019)	Banco de dados da Universidade Federal do Sergipe (UFS).
Aguirre e Pérez (2020)	Registros acadêmicos.
Deb <i>et al.</i> (2024)	Registros acadêmicos.
Serrano <i>et al.</i> (2024)	Diversos bancos de dados institucionais.
Tariq <i>et al.</i> (2023)	Banco de Dados do Portal da Universidade do Sul da Ásia.
García <i>et al.</i> (2020)	Registros acadêmicos.
Hossain <i>et al.</i> (2022)	Questionário aplicado via Google Forms.
Santos <i>et al.</i> (2020)	Registros acadêmicos com informações de notas e sociais.
Gopalakrishnan <i>et al.</i> (2017)	Registros institucionais.

Fonte: Autoria própria (2024).

Q11 - O que pode ser aplicado para prever os discentes com maior risco de evasão a cada período?

Quadro 6: Aplicações para prever os riscos de evasão por cada período dos cursos de Ensino Superior encontrados nos trabalhos selecionados.

TRABALHO	RESULTADO
Murillo <i>et al.</i> (2023)	Métodos de predição usando árvores de decisão.
Salas e Caldas (2024)	Uso de modelos de regressão logística (LR).
Osorio e Santacoloma (2023)	Métodos de aprendizado de máquina e modelos preditivos.
Teodoro e Kappel (2020)	Aplicações de técnicas de Aprendizado de Máquina (ML), Árvores de Decisão e Redes Neurais.
Assis e Marcolino (2024)	Aprendizado de Máquina e modelos de predição.
Hortêncio Filho <i>et al.</i> (2020)	Métodos de aprendizado de máquina e modelos preditivos.
Souza <i>et al.</i> (2022)	Técnica de análise de sobrevivência combinada com a construção de modelos preditivos e mineração de dados.
Martins <i>et al.</i> (2023)	Algoritmos de aprendizado de máquina supervisionados.
Ramos e Araújo (2022)	Modelos de aplicação de Mineração de Dados.
Santos e Goya (2020)	Modelos de aprendizado de máquina e regressão logística.
Santos <i>et al.</i> (2021)	Técnicas de aprendizado de máquinas e algoritmos de classificação.
Barbosa <i>et al.</i> (2023)	Modelos de aprendizado de máquina e Inteligência Artificial (IA).
Viana, Santana e Rabêlo (2022)	Aplicação de Mineração de Dados Educacionais.
Rafiq, Rabbi e Ahammad (2021)	Técnicas de aprendizado de máquina, modelos de sobrevivência e predição, regressão logística (One vs Rest).
Santos <i>et al.</i> (2019)	Modelos de predição, mineração de dados educacionais, árvores de decisão e redes neurais.

Aguirre e Pérez (2020)	Aplicação de técnicas de aprendizado de máquina, modelos preditivos e regressão polinomial.
Deb <i>et al.</i> (2024)	Modelos de aprendizado de máquina, árvores de decisão e regressão linear.
Serrano <i>et al.</i> (2024)	Modelos preditivos, árvores de decisão e técnicas de aprendizado de máquina
Tariq <i>et al.</i> (2023)	Algoritmos de aprendizado de máquina e árvores de decisão.
García <i>et al.</i> (2020)	Técnicas de mineração de dados, redes neurais, regressão logística e sistemas de recomendações.
Hossain <i>et al.</i> (2022)	Modelos de predição e árvores de decisão.
Santos <i>et al.</i> (2020)	Técnicas de aprendizado de máquina, árvores de decisão e algoritmos genéticos.
Gopalakrishnan <i>et al.</i> (2017)	Técnicas de Mineração de dados e regressão linear.

Fonte: Autoria própria (2024).

Q12 – Quais modelos computacionais usados para obter os dados mais rápido e eficaz?

Quadro 7: Modelos mais rápidos e eficazes para investigar os dados de evasão acadêmica encontrados nos trabalhos selecionados.

TRABALHO	RESULTADO
Murillo <i>et al.</i> (2023)	Software R, utilizando o estimador de Kaplan-Meier.
Salas e Caldas (2024)	Ridge, Lasso, Random Forest e Extreme Gradient Boosting.
Osorio e Santacoloma (2023)	Algoritmo <i>Synthetic Minority Oversampling Technique</i> (SMOTE).
Teodoro e Kappel (2020)	Naive Bayes, K-Nearest Neighbors, Random Forest.
Assis e Marcolino (2024)	PyCaret.
Hortêncio Filho <i>et al.</i> (2020)	Software Python, Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) e gradient boosting.

Souza <i>et al.</i> (2022)	<i>gradient boosting (GB) e XGBoost.</i>
Martins <i>et al.</i> (2023)	Modelos de Random Forest e K-Nearest Neighbors (KNN).
Ramos e Araújo (2022)	<i>Softwares IRAMUTEQ e WEKA.</i>
Santos e Goya (2020)	<i>AdaBoost, Random Forest e K-Nearest Neighbor (KNN).</i>
Barbosa <i>et al.</i> (2023)	<i>Decision Tree Classifier, Random Forest e XGBoost.</i>
Viana, Santana e Rabêlo (2022)	Algoritmos RF, ET, MLP e SVM.
Rafiq, Rabbi e Ahammad (2021)	Software Python e Random Forest.
Santos <i>et al.</i> (2019)	<i>Random Forest, K-Nearest Neighbor (KNN), Support Vector Machines, Naive Bayes.</i>
Aguirre e Pérez (2020)	Software ERP Ellucian.
Deb <i>et al.</i> (2024)	<i>Random Forest, K-Nearest Neighbor (KNN), Support Vector Machines, Naive Bayes, Stochastic Gradient Descent.</i>
Serrano <i>et al.</i> (2024)	Random Forest e algoritmo de balanceamento <i>Synthetic Minority Oversampling Technique</i> (SMOTE).
Tariq <i>et al.</i> (2023)	<i>K-Nearest Neighbor (KNN).</i>
García <i>et al.</i> (2020)	<i>Random Forest, K-Nearest Neighbor (KNN), Support Vector Machines e Gradient Boosting.</i>
Hossain <i>et al.</i> (2022)	<i>Support Vector Machines, Random Forest, K-Nearest Neighbor (KNN) e Naive Bayes.</i>
Santos <i>et al.</i> (2020)	<i>Support Vector Machines e Naive Bayes.</i>
Gopalakrishnan <i>et al.</i> (2017)	<i>AdaBoost, Naive Bayes e Support Vector Machines.</i>

Fonte: Autoria própria (2024).

QP1 – Quais os tipos de dados (sociais, financeiros, acadêmicos) encontrados para essa temática?

Quadro 8: Os tipos de dados de evasão encontrados nos trabalhos selecionados.

TRABALHO	RESULTADO
Murillo <i>et al.</i> (2023)	Sociais e acadêmicos.
Salas e Caldas (2024)	Acadêmicos e socioeconômicos.
Osorio e Santacoloma (2023)	Acadêmicos.
Teodoro e Kappel (2020)	Acadêmicos.
Assis e Marcolino (2024)	Acadêmicos.
Hortêncio Filho <i>et al.</i> (2020)	Registros institucionais.
Souza <i>et al.</i> (2022)	Sociais, demográficos e acadêmicos.
Martins <i>et al.</i> (2023)	Acadêmicos.
Ramos e Araújo (2022)	Socioeconômicos e acadêmicos.
Santos e Goya (2020)	Acadêmicos e sociodemográficos.
Santos <i>et al.</i> (2021)	Dados acadêmicos, por base de exames e registros de atividades educacionais.
Barbosa <i>et al.</i> (2023)	Base de dados da plataforma Q-ACADÊMICO.
Viana, Santana e Rabêlo (2022)	Sociais e acadêmicos.
Rafiq, Rabbi e Ahammad (2021)	Sociopolíticos e acadêmicos.
Santos <i>et al.</i> (2019)	Acadêmicos.
Aguirre e Pérez (2020)	Socioeconômicos e acadêmicos.
Deb <i>et al.</i> (2024)	Acadêmicos, demográficos e socioeconômicos.

Serrano <i>et al.</i> (2024)	Acadêmicos.
Tariq <i>et al.</i> (2023)	Demográficos e acadêmicos.
García <i>et al.</i> (2020)	Acadêmicos e sociais.
Hossain <i>et al.</i> (2022)	Acadêmicos e demográficos.
Santos <i>et al.</i> (2020)	Sociais e acadêmicos.
Gopalakrishnan <i>et al.</i> (2017)	Acadêmicos e socioeconômicos.

Fonte: Autoria própria (2024).

QP2 – Como os resultados são apresentados aos *stakeholders*?

Quadro 9: Resultados informados aos *stakeholders* pelos trabalhos selecionados.

TRABALHO	RESULTADO
Murillo <i>et al.</i> (2023)	Tabelas, gráficos, figuras e representações gráficas do desempenho do algoritmo.
Salas e Caldas (2024)	Tabelas, gráficos e figuras esquemáticas.
Osorio e Santacoloma (2023)	Figuras esquemáticas e tabelas.
Teodoro e Kappel (2020)	Análise e interpretação dos dados por meio gráficos, fórmulas matemáticas, tabelas e fluxogramas.
Assis e Marcolino (2024)	Dados estáticos por meio de gráficos, tabelas, figuras e representações gráficas do desempenho do algoritmo.
Hortêncio Filho <i>et al.</i> (2020)	Os resultados foram apresentados por meio de gráficos e fórmulas matemáticas.
Souza <i>et al.</i> (2022)	Gráficos e tabelas.
Martins <i>et al.</i> (2023)	Tabelas.
Santos e Goya (2020)	Tabelas.

Santos <i>et al.</i> (2021)	Figuras esquemáticas e tabelas.
Barbosa <i>et al.</i> (2023)	Tabelas, representação em gráficos do desempenho do algoritmo e figuras.
Viana, Santana e Rabêlo (2022)	Dados estáticos por meio de gráficos, tabelas e figuras esquemáticas.
Rafiq, Rabbi e Ahammad (2021)	Fluxogramas, tabelas e figuras com representações gráficas de desempenho.
Santos <i>et al.</i> (2019)	Figuras.
Aguirre e Pérez (2020)	Análise e interpretação dos dados por meio gráficos, quadros, figuras esquemáticas e fluxogramas.
Deb <i>et al.</i> (2024)	Figuras e tabelas.
Serrano <i>et al.</i> (2024)	Figuras e tabelas.
Tariq <i>et al.</i> (2023)	Os resultados foram apresentados por meio de gráficos, figuras esquemáticas e tabelas.
García <i>et al.</i> (2020)	Tabelas, quadros e figuras.
Hossain <i>et al.</i> (2022)	Tabelas, gráficos, figuras e representações gráficas do desempenho do algoritmo.
Santos <i>et al.</i> (2020)	Dados estáticos por meio de quadros, tabelas, figuras esquemáticas e representações gráficas do desempenho do algoritmo.
Gopalakrishnan <i>et al.</i> (2017)	Tabelas e figuras esquemáticas.

Fonte: Autoria própria (2024).

A partir das respostas das questões de pesquisa, percebe-se alguns aspectos:

- As causas da evasão podem vir de aspectos bastante diversos, desde aspectos puramente acadêmicos até problemas sociais, financeiros e familiares. É preciso conhecer a realidade do dia a dia dos acadêmicos para compreender melhor o que os desmotivam em seus estudos;

- A maioria dos estudos usam dados institucionais ou de uso de alguma plataforma em que os alunos têm cadastro. O problema dessas abordagens, especialmente para dados que não são de desempenho acadêmico, é que elas podem estar com informações desatualizadas ou que não foram preenchidas com compromisso pelos discentes;
- Utiliza-se a Mineração de Dados Educacionais com focos em análises dos dados e a aplicação de algum algoritmo de ML para auxiliar na predição dos possíveis evadidos;
- Não existe um método único de ML que seja melhor ou mais rápido para realizar as predições. Em cada caso uma técnica pode ser melhor que outras, sendo preciso a aplicação e comparação de várias, utilizando métricas de comparação dos modelos;
- O conhecimento é apresentado aos *stakeholders* através de relatórios com gráficos, tabelas e estatísticas, bem como com modelos que fazem predições de acordo com dados fornecidos como entrada.

Os resultados extraídos da RSL serviram de base para a determinação da metodologia de desenvolvimento deste trabalho, cuja descrição e resultados obtidos são expostos e trabalhados nas próximas seções, que irão abordar detalhadamente as etapas seguidas na construção da pesquisa, incluindo a escolha das ferramentas e a forma como os dados foram interpretados. Além disso, serão discutidas as implicações dos resultados para o contexto estudado, destacando suas contribuições e possíveis limitações.

4 METODOLOGIA

Neste capítulo, são apresentados os aspectos que direcionaram essa investigação, bem como a metodologia adotada para a coleta de dados. A pesquisa concentrou-se em compreender os desafios enfrentados pelos estudantes por meio de coleta de dados e análise documental, proporcionando uma visão abrangente dos fatores que influenciam tanto a permanência quanto a evasão dos alunos.

4.1 CONTEXTUALIZAÇÃO

Angicos, no Rio Grande do Norte, é uma cidade pequena com uma população de 11.632 habitantes, de acordo com o Censo de 2022 divulgado pelo IBGE⁴. A cidade é marcada por uma densidade demográfica baixa, com cerca de 15,69 habitantes/km². Em termos de crescimento populacional, houve um aumento de apenas 83 pessoas em relação ao censo de 2010, o que indica um crescimento muito modesto ao longo da última década.

Grande parte da população de Angicos está ligada à zona rural, o que é comum em municípios menores do interior do Nordeste. A economia local é predominantemente baseada na agricultura e pecuária, refletindo a forte presença de áreas rurais. De acordo com o G1 (2023), a cidade ocupa a 49ª posição em termos de população no estado do Rio Grande do Norte.

Por esta razão, Angicos sempre enfrentou desafios típicos de regiões pouco desenvolvidas, como o crescimento comercial e industrial limitado e altas taxas de desemprego. A economia local se manteve tradicionalmente concentrada em atividades de subsistência e em um setor de serviços pouco diversificado, o que limita suas oportunidades de crescimento.

A implementação da Universidade Federal Rural do Semi-Árido (UFERSA) em Angicos, por meio da Resolução CONSUNI/UFERSA nº 006/2008, foi um marco transformador para a cidade. Suas atividades acadêmicas, iniciadas em março de 2009, representaram uma resposta às limitações econômicas locais. A decisão de implantar o campus da UFERSA em Angicos foi baseada na forte tradição educacional do município, que, em 1963, foi palco de um experimento inovador conduzido pelo educador Paulo Freire. A

⁴ <https://cidades.ibge.gov.br/brasil/rn/angicos/panorama>

experiência resultou na alfabetização de cerca de 300 adultos em apenas 40 horas, o que levou a cidade ao reconhecimento nacional e internacional.

Com a chegada da UFERSA, Angicos passou a experimentar uma revitalização econômica significativa, especialmente no setor de serviços, que começou a se expandir para atender às demandas da comunidade acadêmica e dos novos moradores. Áreas como construção civil, alimentação e hospedagem ganharam força, proporcionando novas oportunidades de negócios e geração de empregos, como pôde ser observado entre 2005 e 2007, onde a economia local apresentou um crescimento médio de quase 8% ao ano (Cunha *et al.*, 2023). Embora o desenvolvimento comercial e industrial ainda seja limitado, a presença da universidade trouxe novas perspectivas para o crescimento econômico da cidade.

Além disso, a UFERSA tem desempenhado um papel essencial na promoção do desenvolvimento regional, oferecendo formação acadêmica de qualidade e incentivando a pesquisa aplicada. Este processo não só tem aumentado a qualificação profissional da população local, mas também impulsionado a criação de empresas e iniciativas que contribuem para uma economia mais diversificada e sustentável a longo prazo. Mesmo assim, Angicos ainda enfrenta desafios estruturais em relação ao seu desenvolvimento econômico, a instalação da universidade representa um ponto de virada na trajetória da cidade, trazendo novas possibilidades de progresso.

Contudo, a infraestrutura de transporte em Angicos é limitada, o que impõe sérios desafios à mobilidade dos estudantes da UFERSA. O único meio de transporte público disponível é o ônibus fornecido pela própria universidade, conhecido como "Circular", que serve como a principal opção de deslocamento para muitos alunos. No entanto, a capacidade e a frequência desse ônibus são insuficientes para atender a toda a demanda, especialmente durante os horários de pico, o que causa limitações significativas para os estudantes que dependem exclusivamente desse serviço. Além do ônibus da UFERSA, alguns estudantes que residem em Angicos utilizam transporte pessoal, enquanto outros recorrem a caronas de conhecidos ou familiares.

Há também aqueles que, devido à falta de outras opções, fazem o percurso até o campus a pé, enfrentando situações perigosas, como assaltos, arriscando sua própria vida, o que pode afetar negativamente o rendimento acadêmico. Para os estudantes que residem em outros municípios, a situação é ainda mais complexa. Muitos dependem de ônibus

intermunicipais, que enfrentam problemas, o que pode representar uma barreira. Entre as dificuldades comuns estão as quebras de ônibus, a falta de motoristas, a limitação de horários, com apenas uma saída disponível por dia, e o tamanho reduzido dos veículos. Para alguns, a dependência de carros particulares agrava esses problemas. A falta de uma rede de transporte público eficaz entre as cidades da região e Angicos não apenas dificulta o acesso diário ao campus, podendo levar ao atraso, à ausência constante e, eventualmente, à desistência do curso, assim como também pode desencorajar potenciais alunos de se matricularem.

Ressalta-se, contudo, que o campus da UFERSA é um importante centro de Educação Superior que representa uma significativa oportunidade para a população local e das áreas circunvizinhas. Desempenha um papel crucial na formação acadêmica e profissional de estudantes provenientes de diversas regiões do estado e até mesmo de outros estados, refletindo a rica diversidade cultural e social da sua comunidade acadêmica.

A UFERSA atrai um amplo espectro de estudantes, cada um com suas próprias histórias e circunstâncias. Esses alunos vêm de diferentes contextos socioeconômicos e culturais, incluindo aqueles que se deslocam de outras cidades. A universidade busca atender a essa diversidade por meio de uma oferta curricular abrangente e de programas de apoio estudantil. Esses esforços visam garantir a equidade e a inclusão, abordando as dificuldades emocionais e estruturais que alguns estudantes podem enfrentar e proporcionando um ambiente que apoie seu sucesso acadêmico e pessoal.

Além de sua função educacional, o campus da UFERSA tem um impacto na região do Semiárido. Ao oferecer cursos nas áreas de Ciências Humanas e Exatas, a universidade contribui para o desenvolvimento regional, preparando profissionais que atendem às necessidades locais e promovendo a inovação e o avanço tecnológico.

Os alunos da UFERSA que vêm de outros municípios também enfrentam desafios relacionados à moradia. A escassez de moradia a preço acessível e adequada pode forçar os alunos a enfrentar longos deslocamentos ou a viver em condições precárias. A conciliação entre emprego, estudos e deslocamento pode gerar estresse e afetar o desempenho acadêmico, tornando a permanência no curso mais desafiadora. Para muitos, a incerteza habitacional e o custo associado podem ser fatores decisivos na escolha de abandonar os estudos.

Apesar dos auxílios oferecidos pela universidade, como o Programa Institucional de Assistência Estudantil (PIAE), que inclui bolsas, auxílios e moradia estudantil, ainda existem desafios que afetam a permanência dos alunos. Embora esses auxílios ajudem a aliviar algumas das dificuldades enfrentadas, a diversidade de problemas e a persistência de alguns deles indicam a necessidade de soluções contínuas para melhorar o ambiente de estudo e apoio aos alunos.

Além disso, fatores financeiros têm um impacto considerável. Muitos alunos enfrentam problemas econômicos que podem limitar sua capacidade de arcar com despesas relacionadas aos estudos, como materiais didáticos, alimentação e outras necessidades básicas. Embora programas de assistência como o PIAE ofereçam suporte financeiro, esses auxílios muitas vezes não cobrem todas as necessidades.

A percepção da comunidade acadêmica sobre a evasão também é relevante. A maneira como os problemas são vistos e discutidos dentro da universidade pode influenciar a abordagem para lidar com a evasão. A falta de um diálogo aberto sobre os desafios enfrentados pelos alunos pode resultar em soluções inadequadas ou em uma resposta institucional que não aborda de forma eficaz as reais necessidades da comunidade estudantil.

Na intenção de calcular a média de evasão utilizou-se como exemplo o número de matrículas e o número de abandonos (Figura 6.). No ano de 2023 é constatado que em Licenciatura em Computação e Informática tem-se 55 abandonos, Engenharia Civil 6 abandonos, Engenharia de Produção 6 abandonos, Ciência e Tecnologia Diurno e Noturno 335 abandonos, Pedagogia 27 abandonos e Sistema da Informação 61 abandonos.

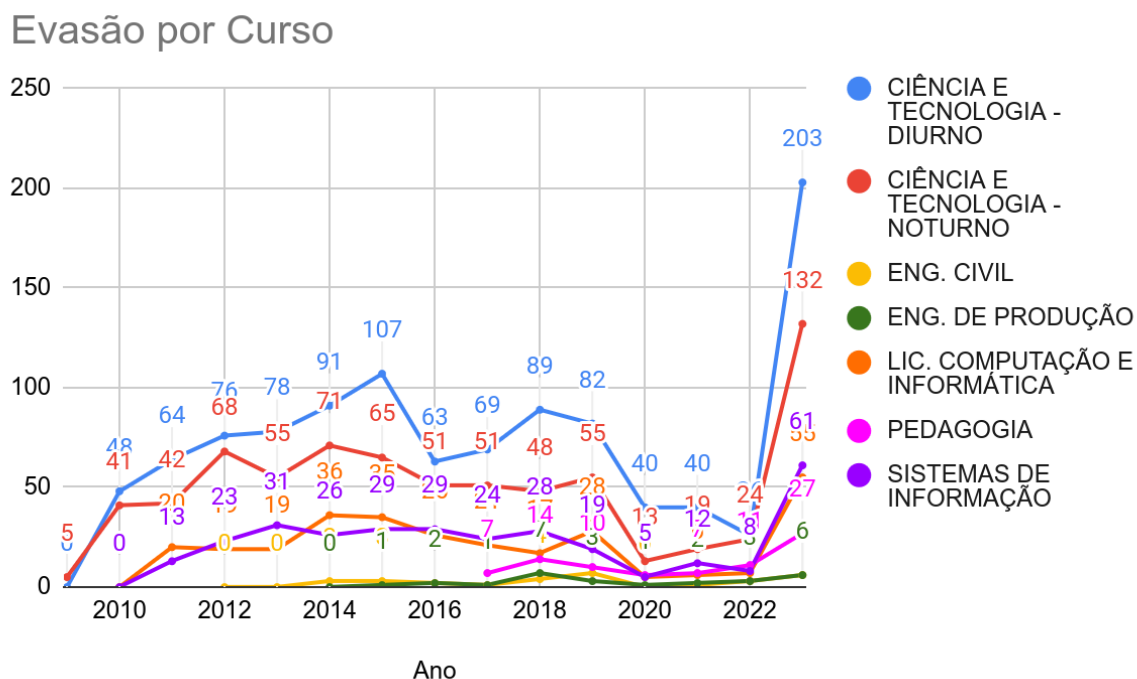
Portanto, é importante que a universidade adote uma abordagem integrada para enfrentar a evasão acadêmica. Isso inclui melhorar a infraestrutura, oferecer apoio contínuo, e promover um ambiente de suporte onde os alunos possam expressar seus problemas e receber ajuda adequada. Medidas proativas e uma análise constante dos fatores que contribuem para a evasão podem ajudar a criar um ambiente acadêmico mais inclusivo e sustentável.

4.2 ELABORAÇÃO DO QUESTIONÁRIO

Um questionário foi desenvolvido (Apêndice A) para identificar as principais limitações que contribuem para a evasão dos estudantes nas instituições públicas com ênfase no campus Angicos, como ilustrado a Figura 6. Compreender os fatores que levam à evasão é

essencial para criar estratégias eficazes de prevenção e retenção. O estudo investigou fatores socioeconômicos, emocionais, dificuldades de adaptação à vida acadêmica e desafios de deslocamento. A pesquisa ofereceu uma visão abrangente das causas da evasão, possibilitando a reflexão sobre soluções para minimizar seu impacto e promover maior inclusão e sucesso acadêmico.

Figura 6. Quantidade de abandonos de curso no Campus Angicos - UFRSA.



Fonte: Divisão de Registros Acadêmicos da UFRSA/Angicos (2023)

A criação de uma base de dados com as respostas foi uma etapa fundamental para entender realmente os fatores que contribuem para a evasão. Para a construção dessa base, a atenção foi dada com foco nos discentes, buscando explorar tanto questões relacionadas às dificuldades acadêmicas quanto aos desafios sociais e financeiros que eles enfrentam.

O questionário foi elaborado com 16 perguntas, sendo 13 objetivas, que facilitaram a quantificação de dados como a frequência de determinados problemas, e 3 subjetivas, que permitiram uma análise qualitativa mais profunda das percepções dos alunos sobre os fatores que os motivam a abandonar o curso. As questões objetivas incluíam temas como transporte, tempo de deslocamento, e infraestrutura disponível no campus. Já as perguntas subjetivas exploravam o impacto emocional e psicológico, como o estresse e a pressão para conciliar estudo e trabalho.

Após a coleta de dados no campo, foi feito um trabalho de organização das informações. Os dados foram inseridos em uma planilha no Google Docs, facilitando o agrupamento e categorização dos fatores identificados. Isso permitiu a criação de gráficos e tabelas que não só quantificaram as respostas, mas também evidenciaram tendências importantes sobre a evasão. Dessa forma, a análise dos resultados pôde ser feita de maneira visual e acessível, auxiliando na identificação de padrões e na construção de inferências sobre os motivos mais críticos para a evasão dos estudantes.

O questionário, apresentado no Apêndice A, foi organizado com o objetivo de captar uma visão abrangente sobre os desafios enfrentados pelos estudantes e as possíveis causas de evasão acadêmica no campus Angicos. A seguir, serão apresentadas as perguntas utilizadas e suas respectivas justificativas.

A primeira pergunta solicitava que o estudante indicasse o curso no qual estava matriculado. Compreender a distribuição dos problemas entre os cursos ajudaria a correlacionar o nível de satisfação com a área de estudo, contribuindo para a identificação de padrões específicos.

Em seguida, foi perguntado o ano de ingresso na universidade, com o intuito de entender há quanto tempo o aluno estava matriculado. Esse dado é crucial para analisar a relação entre o tempo de permanência na instituição e o risco de evasão. Estudantes com maior tempo de matrícula, mas que ainda não concluíram o curso, podem enfrentar desafios relacionados ao ritmo acadêmico ou à progressão, o que pode impactar negativamente sua motivação para continuar os estudos.

Outro aspecto investigado foi o meio de transporte utilizado pelos alunos para chegar à universidade. Foram oferecidas quatro opções objetivas e uma subjetiva para que o aluno pudesse descrever seu tipo de transporte, caso não se encaixasse nas alternativas apresentadas. A finalidade dessa pergunta era identificar se as dificuldades de locomoção estavam entre os fatores que levam à evasão, visto que a localização da instituição em uma área interiorana pode aumentar a dependência de transporte público ou privado.

Além disso, os estudantes foram questionados sobre o acesso ao transporte, se ele era próprio, público, particular, cedido, ou se as opções não eram aplicáveis. Esse questionamento permitiu explorar se o transporte era uma barreira significativa para a frequência às aulas. Em regiões mais afastadas, a falta de um transporte adequado pode ser um fator decisivo para a

permanência ou não dos alunos, especialmente se eles enfrentarem longas distâncias ou se o transporte for irregular.

A questão sobre a localização de moradia dos alunos foi direcionada para entender se eles residiam no município de Angicos ou em outras regiões, e qual o tempo necessário para ir e voltar do campus. Essa pergunta permitiu avaliar se a moradia distante influenciava negativamente a frequência às aulas. Longas distâncias, somadas a problemas com transporte, podem desencorajar a continuidade nos estudos e afetar diretamente a motivação do discente.

O questionário também incluiu uma questão abordando se o estudante havia sofrido algum tipo de preconceito ou violência ao longo do curso, relacionada a características como cor de pele, condição financeira, aparência, deficiência, dificuldade de aprendizado, gênero, curso, ou outros fatores. Essa pergunta se justifica pela importância de identificar possíveis formas de discriminação e violência que podem desestimular os alunos a continuar seus estudos. A existência de um ambiente acadêmico hostil pode ser um forte motivo para a evasão.

Foi perguntado aos alunos se eles se identificavam com o curso que estavam cursando. Essa questão visava medir o nível de satisfação dos estudantes com sua escolha acadêmica. Os alunos podiam responder se estavam satisfeitos com o curso, se queriam concluí-lo mesmo sem identificação, ou se estavam em dúvida sobre a continuidade.

Outra questão abordava a acessibilidade no campus, pedindo que os estudantes indicassem se a consideravam adequada ou inadequada, com um espaço aberto para especificações. O objetivo era identificar se as condições de acessibilidade influenciavam na permanência dos alunos, especialmente para aqueles com deficiência ou mobilidade reduzida.

Em relação ao tempo dedicado aos estudos, os alunos foram questionados se consideravam o tempo disponível suficiente para realizar as atividades acadêmicas. Essa pergunta permitia identificar se a carga horária exigida pelo curso estava adequada à rotina dos estudantes ou se contribuía para a sobrecarga.

Além disso, o questionário investigou se os estudantes trabalhavam, especificando se possuíam emprego formal, informal, temporário, ou se eram bolsistas ou estagiários. Compreender a situação de trabalho dos alunos é essencial para analisar como conciliar trabalho e estudo.

Para os que trabalhavam, foi questionado o horário de trabalho, buscando entender a sobrecarga que poderia interferir na vida acadêmica. Estudantes que trabalham em tempo integral, em regime de plantão ou escala, podem ter problemas em conciliar os horários, o que pode levar ao abandono das atividades acadêmicas.

Questões adicionais abordaram se os alunos eram casados ou estavam em união estável, se tinham filhos, e se contribuíam para o sustento financeiro da família. Esses aspectos são importantes para compreender a carga de responsabilidades familiares e econômicas que poderiam sobrecarregar ainda mais os estudantes.

Por fim, os alunos foram solicitados a descrever as principais dificuldades para sua permanência no curso e se já haviam pensado em trancar disciplinas ou abandonar o curso. Essa pergunta aberta permitiu uma visão mais qualitativa das experiências dos estudantes, fornecendo um espaço para que expressassem seus pensamentos de forma mais detalhada e pessoal.

4.3 APLICAÇÃO DOS QUESTIONÁRIOS

Durante a fase de pesquisa de campo, a coleta de dados foi realizada por meio da aplicação presencial dos questionários em papel. Essa abordagem foi escolhida para garantir que os alunos participassem. Os questionários foram distribuídos fisicamente aos alunos durante as aulas, com a permissão dos professores que estavam administrando as turmas no momento.

Isso permitiu que os alunos os preenchessem com tranquilidade e sem pressa. Para assegurar a privacidade e a confidencialidade das respostas, os questionários não continham campos de identificação pessoal, permitindo que os participantes respondessem com segurança e sem intimidação.

A participação no estudo foi inteiramente voluntária. Os alunos foram informados de que poderiam optar por não responder as ou a qualquer pergunta se se sentissem desconfortáveis ou se a pergunta não se aplicasse à sua situação. Essa flexibilidade foi importante para respeitar as diferentes experiências e a vontade dos participantes.

Além disso, a aplicação presencial dos questionários possibilitou uma interação direta com os alunos, o que facilitou a clarificação de dúvidas e garantiu que todos compreendessem

corretamente as perguntas e instruções. Essa metodologia assegurou a coleta de um número maior de respostas, contribuindo significativamente para os resultados da pesquisa.

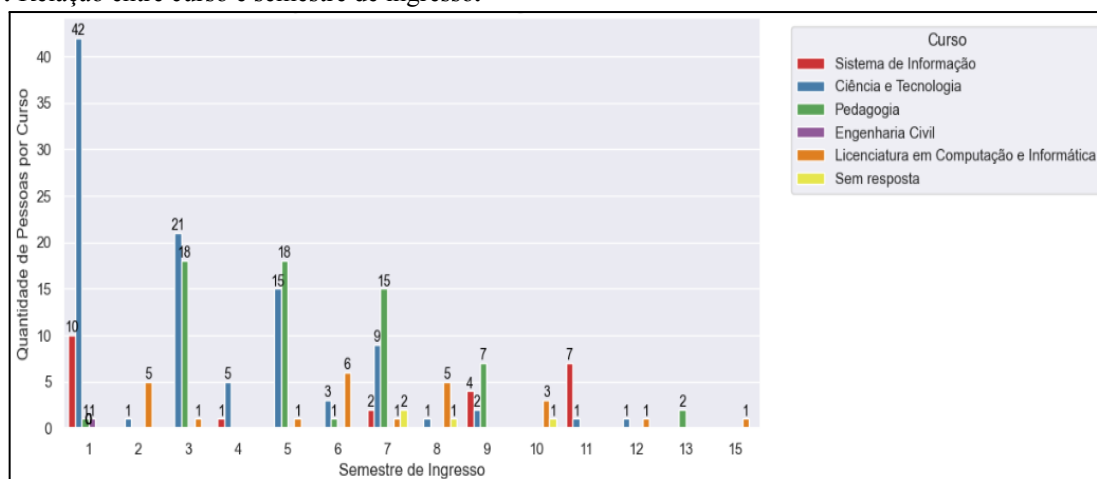
5 RESULTADOS

Neste capítulo, são apresentados os resultados coletados por meio dos questionários aplicados durante a coleta de dados desta pesquisa, que teve como objetivo compreender melhor as condições enfrentadas pelos estudantes da UFERSA e identificar os principais fatores que influenciam a evasão acadêmica. Os questionários foram aplicados em setembro de 2023, contando com dados até o semestre 2023.1.

5.1 DADOS COLETADOS

Na pesquisa de mestrado, foram distribuídos e respondidos um total de 266 questionários. Os resultados descritos a seguir apresentam as respostas de cada pergunta de forma separada por temática. A primeira pergunta, mais geral, foi sobre o curso em que estavam matriculados e o período em relação ao início dos estudos (Figura 7). Durante a pesquisa, constatou-se que os questionários mais respondidos foram do curso de Ciência e Tecnologia (CeT) e dos primeiros períodos. A análise dos dados do gráfico revelou que 11% dos alunos optaram por não responder a essa pergunta, o que pode sugerir que alguns estudantes podem se sentir desconfortáveis em divulgar informações sobre seus cursos. A ausência de respostas pode ser interpretada como dados relacionados à privacidade e confidencialidade dos alunos. A investigação futura deve aprofundar a compreensão desses dados ausentes, explorando suas causas e implicações.

Figura 7: Relação entre curso e semestre de ingresso.



Fonte: Autoria própria (2024).

Vale ressaltar que, a partir de dados disponibilizados pela direção do Campus, verificou-se que, no período de 2018 a 2023, o curso de CeT teve uma média de 30 a 50

alunos novos por semestre, tendo os demais cursos entrada regular anual, além de outras semestrais excepcionais. Em Engenharia Civil, houve uma redução no número de ingressantes, variando de 15 a 30, ao longo desses anos. Em Engenharia de Produção, o ingresso manteve-se entre 23 a 30 alunos. Para Pedagogia, o número de ingressantes variou de 45 a 51. Em Sistemas da Informação, a variação foi de 1 a 49, enquanto em Licenciatura da Computação e Informática foi de 34 a 42, diminuindo para 16 a 29 ingressantes nos anos de 2021 a 2023.

Esses dados revelam padrões interessantes, que podem indicar mudanças significativas nas tendências de ingresso ao longo dos anos. Enquanto o curso de CeT tem entrada regular semestral, os demais têm entradas anuais, sendo alguns cursos de segundo ciclo, isto é, só poderiam ser cursados após a conclusão do curso de CeT. Outro dado interessante é que nos anos de 2020 a 2023, que sofreram a influência negativa da pandemia de COVID-19, a entrada anual se reduziu.

Durante a pesquisa, observou-se que, no período de 2018 a 2023.1, os cursos da UFERSA-Angicos apresentaram diferentes cenários em relação à conclusão dos estudos. No curso de Sistema da Informação, dos 267 ingressantes, 196 estão ativos, 13 concluíram e 46 evadiram, dos quais 12 trancaram. Da mesma forma, no curso de Computação e Informática, dos 187 ingressantes, 149 estão ativos, 4 concluíram e 31 evadiram, sendo que 3 trancaram. Na Pedagogia, dos 294 ingressantes, 216 ainda estão ativos, 24 concluíram e 46 evadiram, com 8 trancamentos. Em Ciência e Tecnologia, dos 749 ingressantes, 434 permanecem ativos, 113 concluíram e 181 evadiram, dos quais 21 trancaram. O curso de Engenharia de Produção, com 302 ingressantes, apresentou 164 ativos, 69 concluíram e 32 evadiram, incluindo 37 trancamentos. Por fim, o curso de Engenharia Civil, com 230 ingressantes, tem 85 ativos, 131 concluíram e 5 evadiram, dos quais 9 trancaram. Esses dados destacam a complexidade das trajetórias acadêmicas dos estudantes e evidenciam a necessidade de compreender os motivos e de adotar medidas para promover a conclusão dos cursos e reduzir a evasão.

5.2 ANÁLISE SOBRE MORADIA E TRANSPORTE

Inicialmente, escolheu-se a temática de moradia e transporte pelas características da cidade de Angicos. Situada no interior do estado e tendo pouco mais de 11.000 habitantes, muitos estudantes do Campus da UFERSA são de outras cidades, algumas ainda menores e menos desenvolvidas economicamente. Isso faz com que a infraestrutura relacionada à

moradia e ao transporte público seja precária. Muitos estudantes se deslocam diariamente para assistir às aulas e assim vivenciar o sonho de fazer um curso superior sem precisar mudar para grandes centros urbanos.

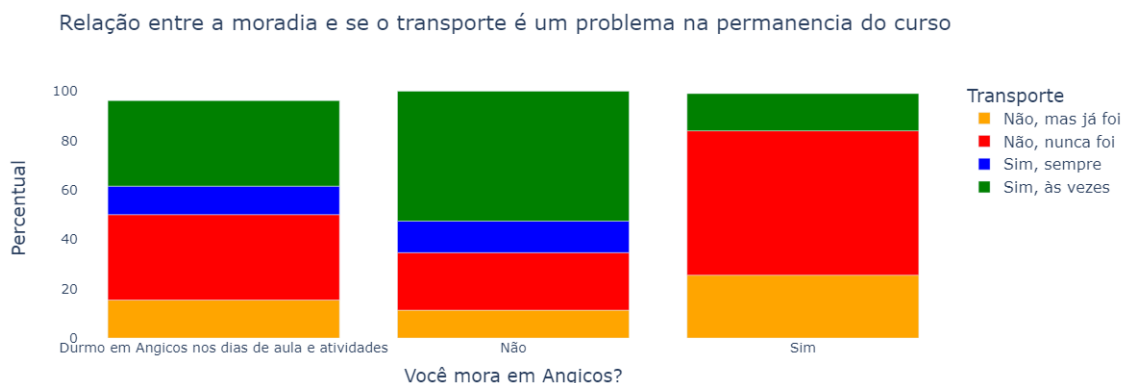
É importante, contudo, investigar se o deslocamento desses estudantes pode influenciar negativamente para a conclusão do curso. Para explorar essa dinâmica, questionou-se quantos alunos residem na cidade onde o campus está localizado e quantos vêm de áreas circunvizinhas. Os dados coletados revelaram o seguinte: 50,2% dos alunos não residem em Angicos; 40% dos estudantes habitam na cidade em que o campus está situado; e 9,8% relataram que pernoitam lá nos dias de aula. Isto comprova que, apesar do campus estar situado em uma cidade pequena, ele atrai alunos de áreas vizinhas. É interessante observar também que o número de alunos que residem na cidade é menor comparado àqueles que moram em municípios vizinhos.

Em seguida, investigou-se como era o deslocamento dos alunos para a Universidade, analisando se isto poderia ser considerado como uma forma de contribuir para a evasão. Os dados destacam que os alunos utilizam a moto (10,2%) e o ônibus (87,2%) como seu principal meio de deslocamento para o campus. Embora haja outras opções disponíveis, elas não representam números significativos.

Também foi analisado se o transporte utilizado para frequentar as atividades no Campus representava uma barreira. Observou-se aí os diferentes níveis de dificuldades enfrentadas pelos alunos em relação ao transporte. Cerca de 39% dos entrevistados afirmaram que isso nunca representou uma barreira em sua trajetória acadêmica. Contudo, os demais participantes da pesquisa declararam que o transporte já foi ou ainda é um obstáculo que afeta seus estudos, sendo que: para 17% o transporte até o campus já foi, mas atualmente não é mais um problema; 36% indicam que é, mas só às vezes; e 8% dizem que sim, o transporte sempre representa um problema para sua permanência no curso.

Além disso, foi feita uma comparação com alunos que residem ou não na cidade, identificando se o transporte é ou não considerado como uma dificuldade (Figura 8). Os dados destacam que a maioria dos estudantes que ainda têm problemas de transporte para realizar seus estudos em Angicos, sempre (em azul) e às vezes (em verde), são aqueles que não moram na cidade ou que só dormem lá nos dias de aula.

Figura 8: Relação entre a moradia e o transporte se é um problema na permanência do curso.



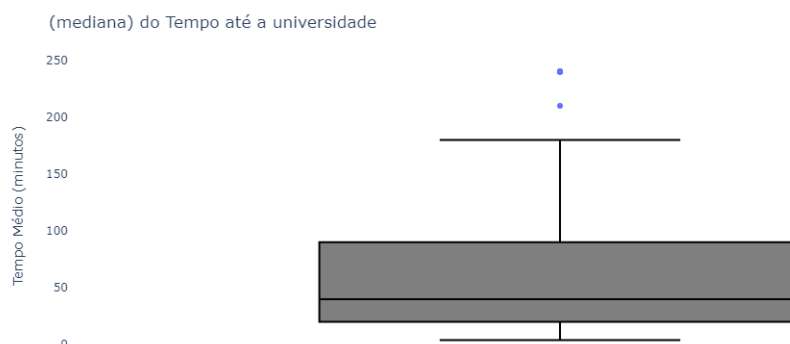
Fonte: Autoria própria (2024).

Complementando estes dados, perguntou-se quanto tempo, em média, os alunos levavam para se deslocar de casa até o Campus (Figura 9). Com base na análise do gráfico boxplot, observa-se que a mediana do tempo de deslocamento dos estudantes até o campus é de 40 minutos. Além disso, os quartis destacam que 25% dos dados estão abaixo de 20 minutos (primeiro quartil), 25% estão acima de 90 minutos (quarto quartil). O valor máximo registrado foi de 240 minutos, enquanto o mínimo foi 0. Percebe-se que alguns dos estudantes residem nas redondezas do campus, mas uma parcela considerável leva mais de uma hora e meia para fazê-lo em uma cidade sem problemas com trânsito pesado e congestionamentos, como nos grandes centros.

Essas análises evidenciam a diversidade de desafios enfrentados pelos discentes em relação à moradia e ao transporte. Alguns afirmaram que esses fatores nunca representaram uma barreira, enquanto outros, relataram que já passaram ou ainda passam por obstáculos. Percebe-se que para alguns essas questões foram e/ou podem ser motivo para contribuir com o pensamento de desistência.

Esse levantamento é essencial para avaliar se os estudantes das regiões próximas a Angicos precisam pernoitar na cidade para cumprir suas atividades acadêmicas, considerando as limitações na infraestrutura de transporte. As informações obtidas são valiosas para analisar a logística de deslocamento dos alunos, influenciando diretamente as estratégias de suporte e intervenção, especialmente em relação à moradia. Compreender essa dinâmica permitirá a implementação de soluções mais eficazes que atendam às necessidades dos estudantes, contribuindo significativamente para sua permanência e sucesso acadêmico.

Figura 9: Boxplot do tempo médio no deslocamento até a universidade.



Fonte: Autoria própria (2024).

5.3 ANÁLISE SOBRE INCLUSÃO, PRECONCEITOS E VOCAÇÃO

Os estudantes também apresentaram outros aspectos que podem interferir na continuação do curso. Um dos investigados foi em razão aos preconceitos que muitas vezes podem atrapalhar a vida acadêmica. Cerca de 50% dos entrevistados afirmaram que não foram vítimas de preconceitos. Outros participantes optaram por não responder a pergunta, o que sugere que muitos podem não perceber ou se sentir confortáveis ao discutir questões sensíveis referentes ao contexto. Essa constatação é relevante, pois sugere que, para muitos, a ausência de experiências relacionadas a preconceitos pode contribuir para uma perspectiva mais positiva em relação à permanência nos estudos, minimizando a possibilidade de pensamentos de evasão.

Contudo, alguns estudantes afirmaram que sofreram ou sofrem com preconceitos na Universidade. Diante disso, foi construída uma nuvem de palavras com as respostas que foram dadas (Figura 10). A nuvem de palavras proporciona uma abordagem visual eficaz para compreender as percepções e experiências dos alunos em relação às situações mencionadas. Nesse contexto, as palavras em maior destaque denotam aquelas que surgiram com maior frequência.

Dentre os preconceitos destacados, observa-se a maior discriminação com relação ao curso. Há a questão cultural de considerar certas profissões ou cursos mais conceituados na sociedade. Outros fatores também foram apontados, com pessoas sendo alvo de preconceito devido a: dificuldades de aprendizagem em relação à absorção nos conteúdos; a aparência relacionada à forma de vestir; o gênero pelo qual a pessoa se identifica; a cor da pele; a presença de alguma deficiência; e a idade, especialmente o etarismo. Além disso, houve uma

citação contra gestantes, frequentemente criticadas ao se encontrarem nesse ambiente. Torna-se evidente que diversos pilares impactam negativamente o desenvolvimento acadêmico.

Figura 10 : Análise de experiências: preconceito e violência no contexto educacional.



Fonte: Autoria própria (2024).

A observação de que a maioria dos preconceitos está relacionada aos cursos e às dificuldades de aprendizagem destaca a necessidade de preparar os futuros professores para estarem prontos a oferecer didáticas e ações inclusivas, garantindo à todos os alunos que se sintam integrados e apoiados no contexto acadêmico. O desenvolvimento de estratégias pedagógicas que levem em consideração a diversidade contribui não apenas para um ambiente mais igualitário, mas também para um aprendizado mais efetivo e enriquecedor.

Diante dos elementos coletados também investigou-se quantos discentes se identificavam com o curso em que estavam matriculados. Os resultados mostraram que a maioria, correspondente a 80,75%, afirmou que se identifica com o curso. No entanto, 14,72% destacaram que não se identificavam, mas expressaram o desejo de concluir o curso. E, 4,53% mencionaram que não se identificavam com o curso e também não tinham certeza se iria concluir.

A maioria dos estudantes estão matriculados no curso desejado, contudo, nem sempre essas pessoas se identificam com a área que optam. Alguns optam por concluir o curso apesar dessa falta de identificação, buscando as oportunidades que um diploma de nível superior pode proporcionar, como a participação em certos concursos públicos, por exemplo. Enquanto

isso, outros decidem não continuar, buscando mudar para a área com a qual se identificam ou desistindo de fato.

Durante a pesquisa também questionou-se sobre a acessibilidade do campus, onde 50% relataram que o ambiente é adequado, 26% consideraram inadequado, e 24% optaram por não responder. Essa avaliação auxilia a compreender como a infraestrutura do campus impacta o bem estar dos alunos. A acessibilidade inadequada pode representar uma barreira significativa, prejudicando a autonomia dos estudantes e, eventualmente, influenciando a decisão de permanecer ou evadir.

A inclusão da diversidade continua a ser um objetivo em busca de concretização, e isso não é diferente na vida universitária na Universidade Federal Rural do Semi-Árido em Angicos. Dentre os desafios enfrentados, destacam-se diversos tipos de preconceitos que podem prejudicar a jornada acadêmica. Essa realidade é evidenciada pelas respostas do questionário, nas quais pessoas afirmaram terem sido vítimas de algum tipo de preconceito.

5.4 ANÁLISE SOBRE EMPREGO, RENDA E TEMPO DE DEDICAÇÃO

No contexto da pesquisa, houve a inclusão de uma questão específica para avaliar se os estudantes que trabalham tinham mais desafios comparados àqueles que não trabalham. O Gráfico da Figura 11 aprofunda a análise, mostrando as respostas.

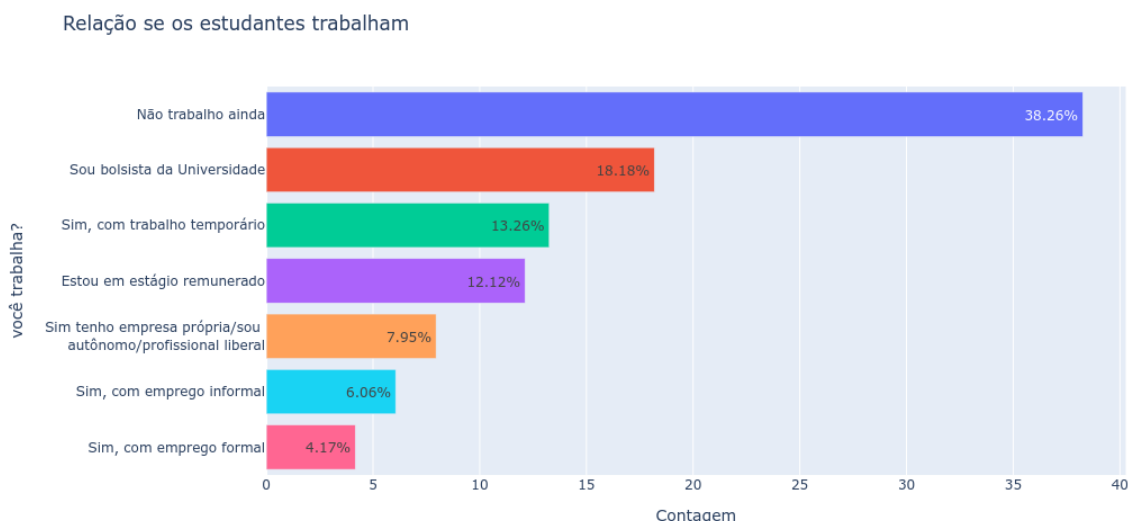
Ao analisar os dados, percebe-se que 37% dos alunos ainda não estão empregados. Essa constatação sugere que a falta de trabalho pode ser um fator contribuinte para a evasão escolar, pois como a maioria dos alunos não reside em Angicos necessitam de recursos para se manter. No entanto, é importante considerar que a falta de tempo livre devido ao trabalho pode impactar negativamente o desempenho.

Na sequência, foi gerada a Figura 12, que apresenta uma análise dos discentes que trabalham, investigando o horário de suas jornadas de trabalho. A construção dessa observação tem como objetivo visualizar como os diferentes regimes de trabalho impactam a vida acadêmica dos alunos, uma vez que tanto o horário quanto a carga horária de trabalho são fatores diretamente relacionados ao tempo que o estudante pode dedicar aos estudos e à frequência nas aulas.

Com base nos resultados apresentados, observa-se que, relativo aos alunos que afirmaram que trabalham, 39,61% dos participantes não possuem um horário fixo de trabalho.

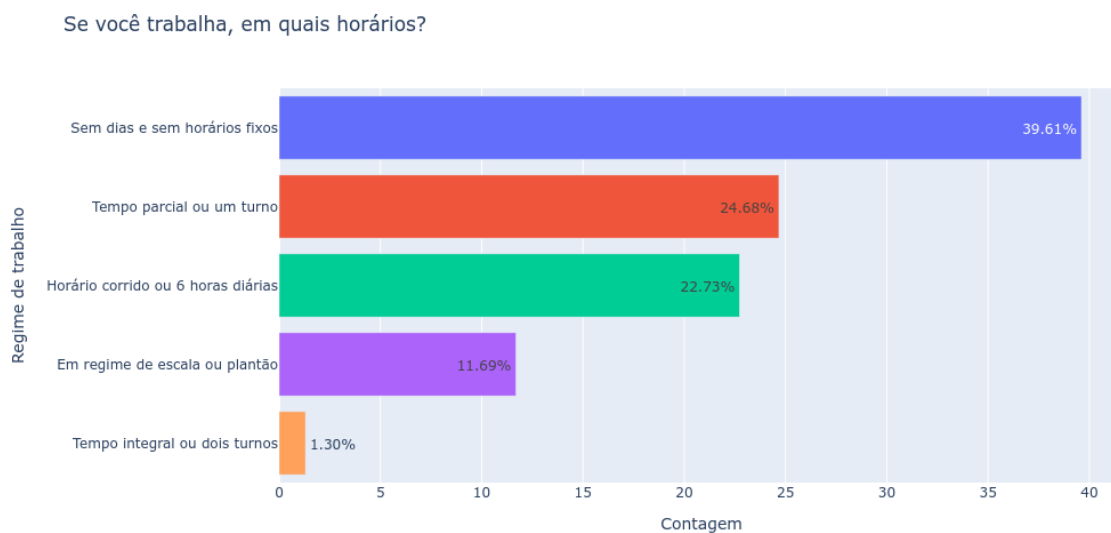
Enquanto isso, 24,68% estão empregados em regime de trabalho parcial ou de um turno. Cerca de 22,58% estão em regime de horário corrido ou trabalham seis horas diárias, enquanto 11,61% estão sob regime de escala ou plantão. Apenas 1,29% trabalham em período integral ou em dois turnos. Isto mostra que a maior parte dos discentes não ocupa uma grande parte do tempo no trabalho, deixando tempo para o estudo.

Figura 11: Percentuais dos estudantes que trabalham.



Fonte: Autoria própria (2024).

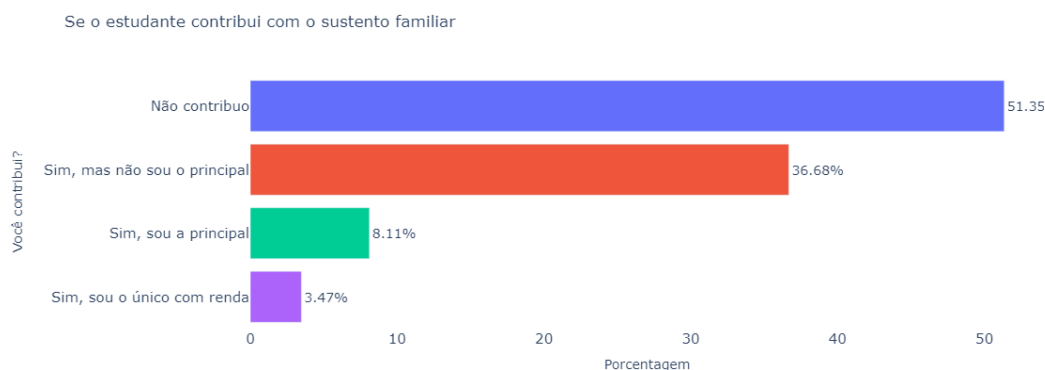
Figura 12: Jornada de trabalho dos estudantes.



Fonte: Autoria própria (2024).

Posteriormente foi verificado se as pessoas contribuem com a renda familiar, na perspectiva de saber se os que trabalham colaboram com os gastos de suas residências ou só com seus estudos. O gráfico da Figura 13 apresenta os dados que foram coletados.

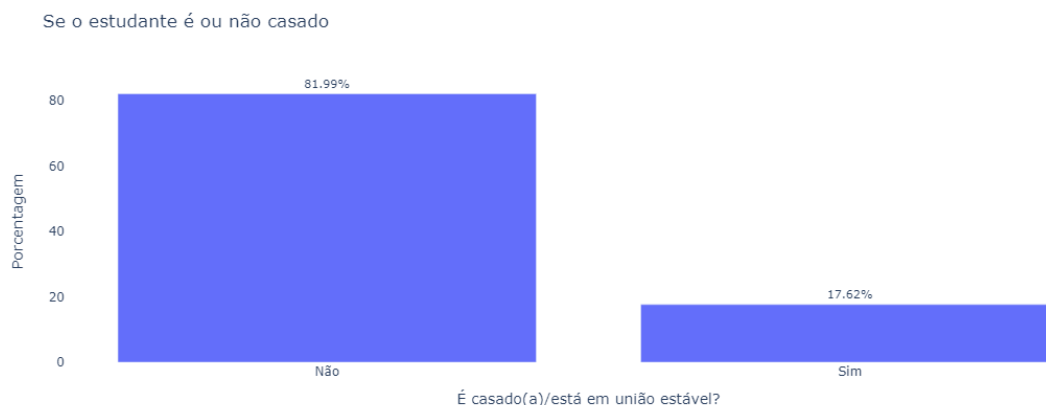
Figura 13 : Contribuição na renda familiar pelos estudantes.



Fonte: Autoria própria (2024).

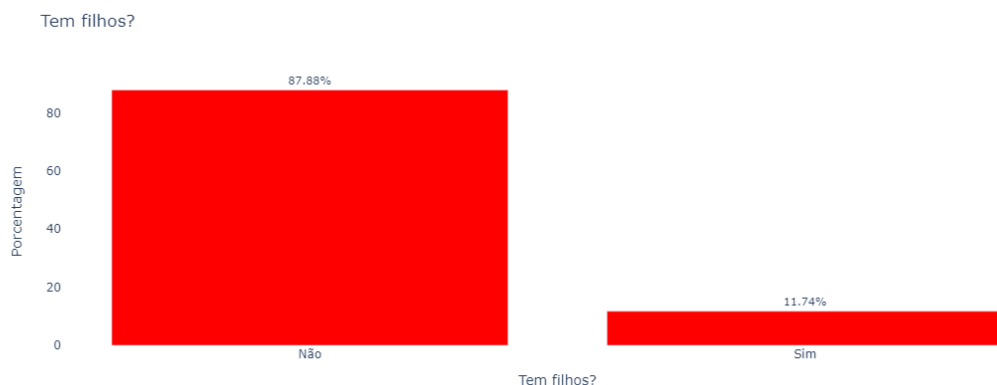
Com base nas respostas fornecidas, vê-se que não há uma grande pressão sob os estudantes entrevistados já que a maioria não contribui (51%) ou contribui apenas parcialmente para o sustento da família (36%). Já para os outros quase 12%, o fato de serem a principal ou única fonte de renda lhes impõe uma grande responsabilidade.

Também foi analisada a quantidade de estudantes casados ou que vivem em união estável, em comparação com aqueles que não possuem compromisso, com o objetivo de compreender se aqueles com responsabilidades familiares apresentam uma maior probabilidade de evasão. A relação entre estado civil e permanência acadêmica é relevante, pois estudantes com compromissos familiares frequentemente enfrentam desafios adicionais, como equilibrar estudos, trabalho e vida familiar, o que pode aumentar as chances de abandono do curso. A Figura 14 apresenta os dados dos discentes casados ou não, fornecendo uma visão clara dessa relação.

Figura 14: Discentes casados e não casados.

Fonte: Autoria própria (2024).

Os resultados mostram que a maioria, cerca de 82%, são de pessoas solteiras, indicando que se trata de um público mais jovem, ainda sem relacionamentos com compromissos sérios. Foi averiguado também se esses discentes têm filhos ou não. Na Figura 15, são apresentados os dados coletados sobre essa questão.

Figura 15: Estudantes com e sem filhos.

Fonte: Autoria própria (2024).

De forma semelhante, quase 88% dos entrevistados ainda não tinham filhos, o que sugere que essa condição pode não ser um fator significativo para a evasão dos alunos, visto que a ausência de responsabilidades familiares diretas pode facilitar a dedicação aos estudos e à permanência na vida acadêmica. Com esses resultados destaca-se que a maioria dos entrevistados não carrega o fardo da responsabilidade de prover o sustento familiar. Contudo, para os que têm filhos, os seus entes dependem da sua renda e trabalham em tempo integral, a universidade pode representar tanto uma oportunidade quanto um desafio.

modalidade são regulamentadas na Resolução CONSUNI/UFERSA N° 003/2020 (UFERSA, 2020) e compreendem como:

- Bolsa Acadêmica: Apoio financeiro mensal para estudantes envolvidos em atividades de ensino, pesquisa, extensão e cultura, com orientação de um docente ou técnico-administrativo (10 horas semanais).
- Bolsa Esporte: Suporte financeiro mensal para alunos com habilidades esportivas, incentivando a participação em atividades esportivas da UFERSA, também com orientação de profissionais da instituição.
- Moradia Estudantil: Vagas para estudantes que não possuem renda familiar na cidade do campus, garantindo permanência durante o período do curso.
- Auxílio Moradia: Suporte financeiro mensal para alunos que não obtiveram vaga na Moradia Estudantil, priorizando suplentes até que surjam novas vagas.
- Auxílio Alimentação: Refeições gratuitas nos Restaurantes Universitários da UFERSA durante o semestre letivo; nos campi sem restaurante, o auxílio é concedido em forma de assistência financeira mensal.
- Auxílio Transporte: Assistência financeira mensal para ajudar estudantes com as despesas de deslocamento até o campus.
- Auxílio Didático: Suporte financeiro para aquisição de materiais e recursos complementares necessários aos cursos.
- Auxílio Inclusão Digital: Fornece apoio para aquisição de equipamentos tecnológicos e planos de internet para participação em atividades remotas, por meio de editais.
- Auxílio Creche: Suporte financeiro para estudantes com filhos em idade pré-escolar (até 6 anos incompletos), cobrindo despesas relacionadas à creche ou cuidados com a criança.
- Auxílio Acessibilidade: Apoio mensal a estudantes com deficiência, transtornos globais de desenvolvimento, altas habilidades e superdotação.

- Auxílio Saúde: Assistência para estudantes que enfrentam problemas de saúde física ou mental, com suporte para tratamentos, medicamentos ou exames não cobertos pelo SUS, mediante avaliação da equipe multiprofissional.
- Auxílio Emergencial: Subvenção financeira para situações excepcionais e momentâneas, como mudanças na situação familiar ou financeira, avaliadas por estudo social.

As modalidades de apoio estudantil são fundamentais para que os alunos permaneçam em seus cursos, contribuindo para a redução da taxa de evasão. No entanto, esses programas, embora eficazes, não conseguem eliminar completamente os desafios que os discentes enfrentam, como problemas sociais, financeiros, preconceitos e de desmotivação. Portanto, apesar de sua importância, os apoios oferecidos não garantem que a evasão não ocorra.

Os resultados da pesquisa sobre a trajetória acadêmica dos estudantes da UFERSA em Angicos revelam que os desafios diários impactam na permanência acadêmica. Diante disso, é essencial pensar-se em estratégias institucionais que abordam esses desafios, como melhorias na infraestrutura e apoio para os estudantes. Essas ações visam criar um ambiente mais acolhedor e propício ao aprendizado, o que pode aumentar a permanência e o sucesso acadêmico dos alunos.

Outra estratégia, de cunho preventivo, é tentar prever que estudantes correm mais risco de abandono do curso. Isso pode ser feito através da Mineração de Dados com a base construída com as respostas dos questionários, utilizando técnicas de *Machine Learning* para fazer uma classificação se um discente tem potencial ou não para desistência.

5.6 MINERAÇÃO DOS DADOS DOS QUESTIONÁRIOS

Utilizando o processo do KDD, a base de dados com as respostas dos questionários foi analisada para a extração de conhecimentos. Na primeira etapa, de seleção dos dados, decidiu-se que todas as respostas do questionário seriam utilizadas, exceto a questão “Cite as principais dificuldades para sua permanência no Curso?” por trazer uma experiência pessoal e muito diversificada. Contudo, a maioria das respostas a essa pergunta estavam cobertas indiretamente pelas demais do questionário. Com isso, a base de dados selecionada continha 266 observações com 16 variáveis cada.

Na segunda etapa, de pré-processamento, o tratamento envolveu a organização e a limpeza, seguido da terceira etapa com a transformação dos dados brutos em um formato adequado para a Mineração dos Dados, com o objetivo de garantir que as informações estivessem preparadas para responder às questões de pesquisa propostas. Essas duas fases tiveram um papel importante no *pipeline* de *Machine Learning*, garantindo a qualidade e a consistência dos dados para torná-los utilizáveis em modelos preditivos. Durante o projeto, foram aplicadas diversas técnicas para lidar com dados ausentes, reduzir ruídos, ajustar formatos e preparar o conjunto de dados para os algoritmos. As principais etapas desse processo estão previstas a seguir.

5.6.1. Tratamento e Limpeza de Dados

Inicialmente, foi realizada a limpeza do conjunto de dados, corrigindo inconsistências e eliminando duplicatas resultantes da inserção dos dados oriundos do questionário em papel no arquivo da planilha. Além disso, foi feita a padronização de formatos de datas, textos e números, garantindo uniformidade nas informações. Merecem destaque:

- *Correção de valores inválidos*: dados fora dos intervalos esperados foram ajustados ou removidos;
- *Revisão de formatos*: foram padronizados formatos de datas, textos e números para evitar problemas de inconsistência;
- Técnicas de filtragem de valores extremos, como análise do Z-score e IQR (Intervalo Interquartil), foram aplicadas para tratar *outliers* que poderiam distorcer as análises;
- Tratamento de dados ausentes: diferentes estratégias foram utilizadas para lidar com valores ausentes como:
 - Imputação por algoritmos preditivos: valores ausentes foram preenchidos com base em modelos treinados;
 - Remoção de linhas ou colunas: em casos de dados excessivamente incompletos, essas entradas foram descartadas.

- Codificação de variáveis categóricas: *Label Encoding* e *OneHot Encoding* foram utilizados para transformar variáveis categóricas em representações numéricas ordinais e binárias.
- Conversão de tipos: colunas originalmente armazenadas como texto (*strings*) foram convertidas para formatos numéricos apropriados.

Neste conjunto de dados, valores de respostas categóricas nominais tiveram um processo de conversão do tipo Codificação em categoria binária (*OneHot Encoding*), em que cada resposta possível era transformada em uma coluna. Por exemplo, na questão “*Você já sofreu algum tipo de preconceito?*”, cada resposta (curso, aparência, etc.) se transformou em uma *feature* com o valor 0, caso não tenha sofrido, ou 1 se já sofreu aquele tipo de preconceito.

Os atributos categóricos ordinais das variáveis foram convertidos em números, onde o valor reflete o grau de dificuldade ou relevância direcionado na resposta dos alunos em relação à evasão. O valor 0 indica ausência ou menor grau de dificuldade, enquanto o valor 1 representa a presença ou maior relevância da dificuldade relacionada. Por exemplo, na questão “*O transporte representa uma barreira/dificuldade para frequentar a universidade?*” as respostas que relataram que isso nunca foi um problema eram atribuídas ao valor 0, “*Não, mas já foi*” ao valor 1, “*Sim, às vezes*” ao valor 2 e “*Sim, sempre*” ao valor 3. Essa conversão permite uma análise quantitativa mais objetiva, alinhando-se ao propósito do estudo. O Quadro 10 traz um dicionário com todas as transformações realizadas, indicando a pergunta, o tipo de transformação e os valores mapeados na base de dados transformada, incluindo os valores substituídos quando estava sem resposta.

Quadro 10: Perguntas e Respostas sobre a Análise dos Dados Encontrados.

Pergunta	Tipo de transformação	Mapeamento na base de dados
Curso	Codificação categoria binária	Feita com a técnica <i>One-Hot Encoding</i>
Semestre	Quantidades de semestres cursados em número inteiro	1º semestre - 1 2º semestre - 2 ...
Mora em Angicos?	Codificação ordinal	Sim - 0 Dorme nos dias de aula - 1 Não - 2
Como é o seu deslocamento até a universidade?	Codificação ordinal	Carro - 0 Moto - 1 Onibus, van, micro-onibus = 2

Com relação ao transporte do item anterior	Codificação ordinal	Próprio - 0 Não se aplica, público, sem resposta - 1 Cedido - 2 Particular - 3
O transporte representa uma barreira/dificuldade para frequentar a universidade?	Codificação ordinal	Não, nunca foi - 0 Não, mas já foi - 1 Sim, às vezes - 2 Sim, sempre - 3
Quanto tempo, em média, você demora para se deslocar todos os dias até o campus (ida e volta)?	Conversão do tempo em minutos	Número de minutos
Você sofreu algum tipo de preconceito ou violência durante o Curso	Codificação categoria binária	Feita com a técnica <i>One-Hot Encoding</i>
Você se identifica com o curso que está fazendo?	Codificação ordinal	Sim - 0 Não, mas quero concluir - 1 Não, não sei se concluirei - 2
Em relação ao tempo necessário como discente para dedicar no estudo?	Codificação ordinal	É suficiente - 0 É insuficiente, mas desempenho a maioria das atividades - 1 É insuficiente, mas só realizo as atividades obrigatórias - 2 É insuficiente e não consigo realizar as atividades obrigatórias - 3
Você trabalha?	Codificação ordinal	Não trabalho ainda - 0 Sou bolsista da universidade - 1 Sim, com trabalho temporário - 2 Estou em estágio remunerado - 3 Sim tenho empresa própria/sou autonomo/profissional liberal - 4 Sim, com emprego informal - 5 Sim, com emprego formal - 6
Se você trabalha, em quais horários?	Codificação ordinal	Sem dias e sem horários fixos, sem resposta - 0 Tempo parcial ou um turno - 1 Horário corrido ou 6 horas diárias - 2 Em regime de escala ou plantão - 3 Tempo integral ou dois turnos - 4
É casado(a)/está em união estável	Codificação ordinal	Não - 0 Sim - 1
Tem filho?	Codificação ordinal	Não - 0 Sim - 1
Você contribui para o sustento financeiro da família?	Codificação ordinal	Não contribuo - 0 Sim, mas não sou o principal - 1 Sim, sou o principal - 2 Sim, sou o único com renda - 3
Já pensou em trancar	Codificação ordinal	Não - 0

disciplinas, período ou abandonar o curso? Se sim, porquê?		Sim - 1
Cite as principais dificuldades para sua permanência no Curso?	Nenhuma Codificação Aplicada	X

Fonte: Autoria própria (2024)

5.6.2. Aplicação de Modelos de *Machine Learning*

Na última etapa da Mineração dos Dados, foram aplicadas técnicas específicas para preparar os dados para a aplicação dos algoritmos de ML (Mckinney, 2017):

- Divisão em conjuntos de treino e teste (*Holdout*): O conjunto de dados foi separado para avaliar o desempenho do modelo, mantendo proporções consistentes entre as classes (estratificação) daqueles com e sem a intenção de evadir.
- *Feature Scaling*: A normalização garantiu que a escala das variáveis não influenciasse indevidamente os modelos.

Existem outras metodologias para o teste de algoritmos de ML, como a validação cruzada para separação de conjuntos de treinamento e teste, e testes de significância estatística, além das métricas baseadas na matriz de confusão. Contudo, o foco deste trabalho não é realizar testes de desempenho de técnicas e modelos, apenas mostrar uma aplicação da aprendizagem de máquinas no problema em questão.

Para a aplicação dos modelos de ML utilizou-se o PyCaret⁵, uma biblioteca de Python que oferece uma abordagem simplificada e eficiente para construir, treinar e avaliar modelos. Ele é uma ferramenta que automatiza muitas etapas do ciclo de vida de aprendizado de máquina, desde o pré-processamento dos dados até a seleção do melhor modelo. Esta abordagem é tratada como *AutoML* ou ML automatizado. A biblioteca PyCaret foi escolhida por utilizar a mesma linguagem de programação que já estava sendo utilizada para o desenvolvimento das demais etapas do estudo, além de ser fácil de utilizar, possuir um bom desempenho e implementação eficiente de técnicas e ML. Para este trabalho, utilizou-se as configurações padrões para a execução dos algoritmos. O uso do PyCaret permitiu:

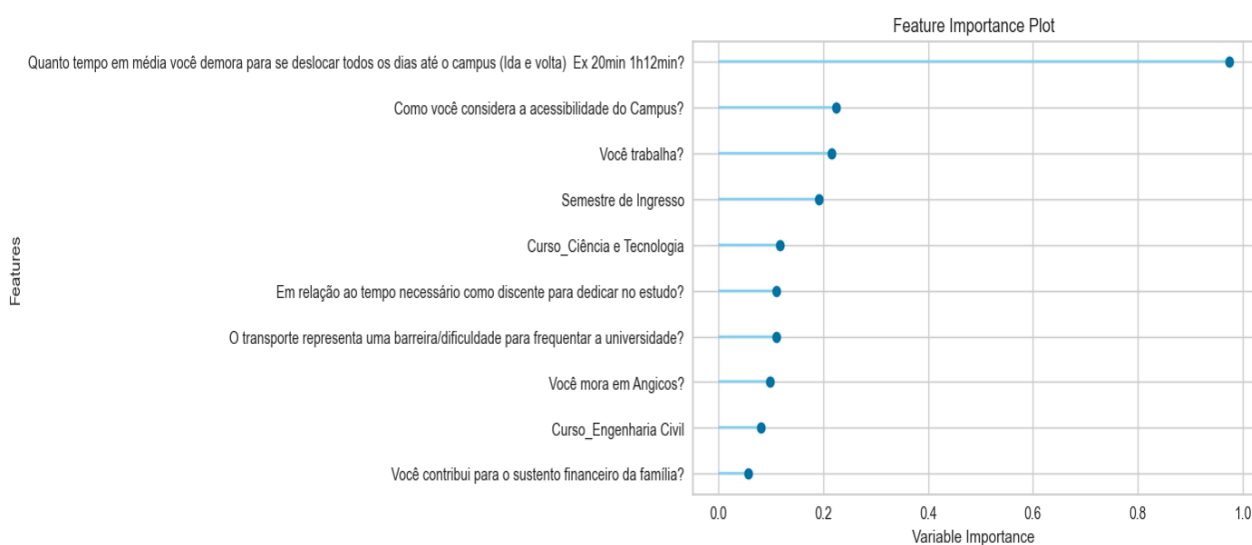
- Comparação de diversos algoritmos;

⁵ <https://pycaret.org/>

- Determinação dos valores dos hiperparâmetros dos algoritmos de maneira automatizada;
- Análises detalhadas de métricas e visualizações interpretativas.

Durante a execução, o PyCaret também realizou uma análise automática de atributos importantes. A Figura 17 aponta as 10 principais características que influenciam na decisão de se evadir ou não segundo as respostas dos questionários.

Figura 17: Gráfico sobre a importância de cada fator de evasão.



Fonte: Autoria própria (2024).

Os resultados mostraram que a questão do transporte, com o tempo de deslocamento para o campus, é, na prática, um fator determinante. Isto é muito preocupante pois, como já visto, a maioria dos entrevistados mora em outras cidades e o fato de morar em Angicos ou não também aparece no *Top 10*. Além disso, em Angicos não há um sistema de transporte público satisfatório. Cabe também destacar que a acessibilidade vem como segundo fator. Já a sobrecarga com o trabalho vem em terceiro lugar, influenciando negativamente a motivação dos discentes. O semestre de ingresso, na sequência, também mostra que o acúmulo de atividades vai minando a resiliência dos discentes com o tempo. Com relação aos cursos, vê-se que os cursos de CeT e Engenharia Civil aparecem também como fator de influência, indicando que é preciso uma atenção especial a ser dada pelas coordenações dos cursos com relação a este tema.

Em relação aos resultados obtidos com o PyCaret, para facilitar a compreensão, foram elaboradas tabelas comparativas utilizando dois conjuntos de dados: um com todas as colunas

(incluindo o tempo de deslocamento) e outro sem a coluna referente a essa variável. O Quadro 11, contendo todas as colunas, conseguiu atingir métricas de acurácia, área sob a curva (Area Under the Curve - AUC), precisão, *recall*, *f1-score*, coeficiente Kappa e coeficiente de correlação de Matthews (Matthews Correlation Coefficient - MCC) de 100% com dois modelos, *Logistic Regression* e KNN, sendo o último com o melhor tempo de execução. Uma das possíveis causas para isto pode ser a quantidade pequena de instâncias da base, o que não possibilita uma grande diversidade de instâncias.

Quadro 11: Resultado do PyCaret com variável tempo de deslocamento.

Modelo	Acur.	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Logistic Regression	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	64.00
K Neighbors Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.05
Ridge Classifier	0.9947	0.9977	1.0000	0.9917	0.9957	0.9890	0.9896	54.00
Extra Trees Classifier	0.9842	0.9966	0.9909	0.9833	0.9865	0.9674	0.9690	223.00
SVM - Linear Kernel	0.9725	1.0000	0.9809	0.9742	0.9761	0.9438	0.9476	52.00
Random Forest Classifier	0.9509	1.0000	1.0000	0.9263	0.9600	0.8974	0.9064	0.19
Naive Bayes	0.8798	0.8625	1.0000	0.8305	0.9057	0.7462	0.7750	0.05
Quadratic Discriminant Analysis	0.8798	0.9977	1.0000	0.8305	0.9057	0.7462	0.7750	51.00
CatBoost Classifier	0.8523	1.0000	1.0000	0.8002	0.8868	0.6859	0.7261	816.00
Decision Tree Classifier	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	49.00
Ada Boost Classifier	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	62.00
Gradient Boosting Classifier	0.5649	0.9955	1.0000	0.5649	0.7219	0.0000	0.0000	146.00
Linear Discriminant Analysis	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	76.00
Light Gradient Boosting Machine	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	68.00
Dummy Classifier	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	73.00

Fonte: Autoria própria (2024).

Já o Quadro 12, gerada sem a variável do tempo de deslocamento apresentou algumas diferenças nos resultados, indicando que outros fatores ganharam maior destaque quando essa informação foi excluída, mas mantendo o algoritmo *Logistic Regression* como o que obteve a maior acurácia, com mais de 97%. Isso evidencia que, embora o tempo de deslocamento seja

uma variável de impacto considerável, outros fatores também desempenham um papel expressivo na análise e predição da evasão acadêmica. Esses resultados reforçam a complexidade do fenômeno da evasão, que não pode ser explicado por uma única variável, mas sim por uma interação de múltiplos fatores que influenciam a experiência e a permanência dos discentes no ensino superior.

O tempo de deslocamento destacou-se como uma variável crítica, mas a exclusão dessa dimensão revelou a relevância de outros elementos, como moradia, acessibilidade e sobrecarga de trabalho. Esses resultados mostram a importância de modelos preditivos para antecipar melhor as motivações dos estudantes e orientar estratégias eficazes de combate à evasão. Assim, as informações analisadas oferecem subsídios valiosos para intervenções direcionadas, promovendo um ambiente acadêmico mais inclusivo e favorável à permanência dos alunos.

Quadro 12: Resultado do PyCaret sem a variável tempo de deslocamento.

Modelo	Acur.	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Logistic Regression	0.9784	0.9964	1.0000	0.9667	0.9822	0.9550	0.9582	1.45
K Neighbors Classifier	0.9545	0.9755	0.9709	0.9226	0.9444	0.8649	0.8713	143.00
Ridge Classifier	0.9509	0.9976	0.9318	0.9809	0.9539	0.9014	0.9058	0.06
Extra Trees Classifier	0.8909	0.9789	0.9609	0.8712	0.9112	0.7721	0.7856	299.00
SVM - Linear Kernel	0.9567	0.9964	0.9800	0.9536	0.9635	0.9099	0.9187	56.00
Random Forest Classifier	0.6348	0.9840	1.0000	0.6112	0.7573	0.1773	0.2603	142.00
Naive Bayes	0.8263	0.8000	1.0000	0.7707	0.8689	0.6265	0.6779	58.00
Quadratic Discriminant Analysis	0.8263	0.9988	1.0000	0.7707	0.8689	0.6265	0.6779	59.00
CatBoost Classifier	0.5649	0.9524	1.0000	0.5649	0.7219	0.0000	0.0000	1.80
Decision Tree Classifier	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	98.00
Ada Boost Classifier	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	136.00
Gradient Boosting Classifier	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	213.00
Linear Discriminant Analysis	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	113.00
Light Gradient Boosting Machine	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	0.20

Dummy Classifier	0.5649	0.5000	1.0000	0.5649	0.7219	0.0000	0.0000	44.00
------------------	--------	--------	--------	--------	--------	--------	--------	-------

Fonte: Autoria própria (2024).

5.7 DISCUSSÃO ACERCA DA QUESTÃO GERAL DE PESQUISA E HIPÓTESES

Sobre a Questão Geral de Pesquisa deste trabalho “Como identificar os fatores que causam a evasão de forma precoce dos discentes presentes nos cursos de graduação do Campus de Angicos da UFRSA e utilizá-los para definir estratégias de prevenção para esse fenômeno”, é possível afirmar que a identificação dos fatores que levam à evasão deve considerar aspectos além do desempenho acadêmico. Como mostra o Quadro 12, fatores como a demora para o deslocamento até o Campus, a falta de acessibilidade e a falta de tempo para estudar devido ao trabalho podem influenciar muito na decisão de abandonar os estudos. Para prevenir que isso ocorra, é preciso identificar previamente os estudantes que têm maior tendência para que isso ocorra, utilizando métodos de aprendizagem de máquina como o exposto neste trabalho. Além disso, é essencial a criação e fortalecimento de políticas de permanência estudantil que levem em consideração estes fatores, como residências universitárias, auxílio moradia, melhoria na infraestrutura do campus e a implementação de atividades não-presenciais.

Sobre as hipóteses investigadas neste trabalho:

- O maior índice de evasão ocorre nos primeiros períodos: estudos realizados em instituições de ensino superior geralmente apontam que os primeiros períodos têm, de fato, os maiores índices de evasão (CRUB, 2016). Isso ocorre devido à dificuldade de adaptação dos discentes ao ritmo acadêmico, às expectativas do curso, e à incompatibilidade entre o que esperavam e o que encontram na graduação. No caso deste trabalho, o período cursado não teve tanta relevância como outros fatores;
- Os discentes com dificuldade em se locomover até o/ou no campus têm maior tendência a desistir dos seus estudos: A precariedade do transporte público, apresenta como um fator determinante na evasão. Estudantes que depender do transporte fornecido pela UFRSA (o Circular) ou de outros meios limitados tende a levar ao desejo de desistência;
- Discentes que são responsáveis por famílias estão mais propensos a se evadirem: discentes com responsabilidades familiares enfrentam frequentemente desafios

adicionais, como a necessidade de trabalhar para sustentar a família e cuidar dos dependentes. Esses fatores podem levar à escolha de priorizar a família em vez dos estudos;

- Discentes de cursos das áreas de exatas e tecnológicas têm maior risco de cometer evasão: cursos das áreas de exatas e tecnológicas tendem a ter maiores taxas de evasão devido ao alto nível de dificuldade, exigência em disciplinas como cálculo e programação, e desgaste mental relacionado a essas áreas. Nos resultados vê-se isso se confirmar para dois casos, que foram os Bacharelados em Ciência e Tecnologia e Engenharia Civil;
- Discentes que apresentam algum problemas de conciliar os estudos com o trabalho se evadem com maior frequência: a necessidade de conciliar trabalho e estudo é uma das principais causas de evasão, pois reduz o tempo disponível para estudo e descanso, levando ao baixo desempenho acadêmico e ao desejo de desistir. Isto deveria ser o oposto, já que um curso superior abre um leque de oportunidades na carreira, mas até que isso ocorra o discente já pode ter abandonado o seu curso.

Sendo assim, observa-se que quase todas as questões levantadas foram confirmadas através dos resultados da pesquisa, apesar de que o mais indicado é que seja feito um estudo mais aprofundado para confirmar o que apontam os indícios obtidos nesta pesquisa.

6. CONCLUSÃO

Este estudo teve como objetivo investigar as causas da evasão no ensino superior, de forma específica, no campus da UFERSA em Angicos/RN. A investigação mostrou que a evasão é um problema multifacetado, influenciado por uma série de fatores, incluindo transporte até o campus, desafios financeiros, falta de tempo, questões pessoais e percepções de acessibilidade e inclusão.

A metodologia adotada incluiu uma abordagem qualitativa, utilizando estudos em documentos oficiais da instituição, questionários aplicados aos alunos e análise de literatura. Adicionalmente, foi explorado o desenvolvimento e a aplicação de técnicas de *Machine Learning* (ML) e Mineração de Dados (MD) para prever as causas da evasão e fornecer uma previsão mais precisa, permitindo que os gestores educacionais adotem medidas proativas para reduzir esse índice.

Os resultados indicaram que os estudantes enfrentam desafios relacionados a deslocamento, tempo de estudo insuficiente e necessidade de conciliar trabalho e estudo. Questões de preconceito e discriminação, embora menos reportadas, também foram identificadas como fatores de evasão, destacando a importância de um ambiente mais inclusivo.

As ações propostas, como a aplicação de questionários detalhados, o desenvolvimento de modelos preditivos de evasão e a análise descritiva dos dados, visam fornecer uma base sólida para o desenvolvimento de políticas educacionais que incentivem a permanência dos estudantes. Os resultados desta pesquisa serão apresentados à direção e aos departamentos do campus, com o intuito de implementar estratégias eficazes para reduzir a evasão.

Com a pesquisa de literatura, constatou-se os fatores que influenciam a evasão no ensino superior no campus estudado. Com esses dados, foi possível avaliar as necessidades dos discentes, bem como as ações de permanência realizadas pelo campus. Isso proporcionou a oportunidade de utilizar a Mineração de Dados como ferramenta para avaliar os registros acadêmicos existentes, focando na evasão universitária.

A partir dessa avaliação, destaca-se que as questões enfrentadas pelos discentes são relativas a fatores externos, refletindo as responsabilidades adicionais de uma população mais madura. Entre os principais desafios encontrados, estão a dificuldade de moradia, a

conciliação entre trabalho e estudo e as complicações relacionadas ao deslocamento. Com a implementação de ações direcionadas, espera-se transformar essa realidade e assegurar que mais alunos completem com êxito seu percurso acadêmico.

Para pesquisas futuras, recomenda-se a realização de estudos longitudinais que analisem a trajetória dos alunos ao longo do tempo buscando verificar como os fatores de evasão mudam ao longo do tempo, além da exploração de variáveis adicionais, como o contexto familiar e a percepção de apoio institucional. A inclusão de amostras mais amplas, considerando outros campi e cursos, e até mesmo outras universidades do Nordeste, pode proporcionar uma visão mais abrangente sobre a evasão no ensino superior em diferentes escalas, do âmbito local à esfera regional. Outras condições relatadas na literatura também podem ser verificadas, como as condições de moradia e desempenho acadêmico no início do curso.

A quantidade de discentes participantes nesta pesquisa pode ser considerada sua maior limitação, já que ela foi focada em um campus da UFERSA e contém um percentual de discentes. Por esta razão, pretende-se realizar a validação do modelo com dados não vistos antes, e será feita uma discussão mais aprofundada dos resultados em relação às métricas calculadas. Também será criado um cenário de aplicação e uso do modelo escolhido, com a disponibilização do modelo por meio de uma aplicação web. Sobre as questões subjetivas, também podem ser exploradas outras dimensões, além das palavras chave, como alguma técnica de análise de sentimento.

Outras possibilidades é avaliar a inclusão de outras variáveis de entrada, como a situação de moradia, medidas de desempenho escolar e distância até o campus, ao invés do tempo de deslocamento. Por fim, para apresentar os resultados de melhor forma, também é aventada a possibilidade de criação de um *dashboard* que permita aos gestores visualizar a propensão à evasão dos discentes.

7. REFERÊNCIAS BIBLIOGRÁFICAS

AGUIRRE, Cindy Espinoza; PÉREZ, Jesus Carretero. Técnicas de análise preditiva de dados aplicadas ao abandono de estudos universitários. In: **XLVI CONFERÊNCIA LATINO-AMERICANA DE INFORMÁTICA (CLEI)**, 46., 2020, Loja, Equador. Proceedings. Piscataway, NJ: IEEE, 2020. p. 1–10. DOI: 10.1109/CLEI50758.2020.9280247. Disponível em: <https://ieeexplore.ieee.org/document/9280247>. Acesso em: 8 set. 2024.

ASSIS, Marcos Vinicius Oliveira de; MARCOLINO, Anderson Santos. A Predictive Model for Dropout Risk in a Computer Science Education Program. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 35., 2024, Rio de Janeiro. Anais da 35ª edição do SBIE, 2024. Porto Alegre: Sociedade Brasileira de Computação, 2024. p. 1560–1573. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/31341>. Acesso em: 8 set. 2024

BARBOSA, Dyego; CABRAL, Luciano; DWAN, Filipe; FREITAS, Elyda; MELLO, Rafael Ferreira. Previsão da Evasão Escolar através da Análise de Dados e Aprendizagem de Máquina: Um estudo de caso. In: WORKSHOP DE APLICAÇÕES PRÁTICAS DE LEARNING ANALYTICS EM INSTITUIÇÕES DE ENSINO NO BRASIL (WAPLA), 2., 2023, Passo Fundo/RS. Anais do 2º WAPLA, 2023. Porto Alegre: Sociedade Brasileira de Computação, 2023. Disponível em: <https://sol.sbc.org.br/index.php/wapla/article/view/26127>. Acesso em: 2 nov. 2024.

BRASIL. **Lei nº 9.394/96 de 20 de Dezembro de 1996**. Estabelece as Diretrizes e Bases da Educação Nacional. Brasília: Ministério da Educação, 1996.

BRASIL. **Lei nº 12.711, de 29 de Agosto de 2012**. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. Brasília: Ministério da Educação, 2012.

BRASIL. **Plano de Reestruturação e Expansão das Universidades Federais (REUNI)**. Brasília, 2007. Disponível em: http://portal.mec.gov.br/index.php?option=com_content&view=article&id=864&Itemid=86. Acesso em: 5 mai. 2025.

BRASIL. **Lei nº 10.861, de 14 de abril de 2004**. Institui o Sistema Nacional de Avaliação da Educação Superior (SINAES). Brasília, 2004. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2004/110.861.htm. Acesso em: 6 abr. 2025.

BRASIL. **Plano Nacional de Assistência Estudantil - PNAES**. Brasília, 2010. Disponível em: <https://www.mec.gov.br/assistencia-estudantil/pnaes>. Acesso em: 8 mar. 2025.

BRASIL. Ministério da Educação. *Plano Nacional de Educação: Lei nº 13.005, de 25 de junho de 2014*. Diário Oficial da União, Brasília, 26 jun. 2014. Disponível em: <http://pne.mec.gov.br/>. Acesso em: 10 dez. 2024.

CASTRO, Daniel Gomes, FERRARI Leandro Nunes de. Introdução à Mineração de Dados: **Conceitos Básicos, Algoritmos e Aplicações**. Editora Saraiva, 2016. *E-book*. ISBN

978-85-472-0100-5. Disponível em:

<https://app.minhabiblioteca.com.br/#/books/978-85-472-0100-5/>. Acesso em: 13 jan. 2023.

CARRANO, Davi; ALBERGARIA, Elisa Tuler de; INFANTE, Carlos; ROCHA, Leonardo. Combinando técnicas de Mineração de Dados para melhorar a detecção de indicadores de evasão universitária. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2019. p. 1321. Disponível em: <http://ojs.sector3.com.br/index.php/sbie/article/view/8864>. Acesso em: 27 jan. 2023.

COLLAZOS, Kathya; BARRETO, Jorge Muniz; PELLEGRINI, Gisele F. Análise do Prontuário médico para a utilização com KDD. In: **CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE-CBIS**. 2000.

CRUB, CRUB sedia segunda edição do Seminário Evasão no Ensino Superior. Conselho de Reitores das Universidades Brasileiras, 2016. Disponível em: <http://www.crub.org.br/?p=4977>. Acesso em: 18 fev. 2025.

CUNHA, Graziela Souza; SILVA, Alderir; FERREIRA, Francisco da Silva; GOMES, Thiago Geovane Pereira. Mudança estrutural na economia de Angicos: **o impacto da Ufersa**. Revista de Economia Mackenzie, 20(2), 191–222, 2023. Disponível em: https://www.researchgate.net/publication/376294526_MUDANCA_ESTRUTURAL_NA_ECONOMIA_DE_ANGICOS_O_IMPACTO_DA_UFERSA. Acesso em: 18 fev. 2025.

DEB, Sreedham; SAMMY, Sakira Rezowana; TUSHER, Abdur Nur; SAKIB, Rabino Salehin; HASAN, Firoz; AUNIK, Ajharul Islam. Previsão da evasão escolar: uma abordagem de aprendizado de máquina. In: **15ª Conferência Internacional sobre Computação em Nuvem e Engenharia de Serviços (International Conference on Cloud Computing and Services Science – CLOSER 2024)**. [S.l.]: IEEE, 2024. Disponível em: <https://ieeexplore.ieee.org/document/10726161>. Acesso em: 28 nov. 2024.

DIAS SOBRINHO, José. Democratização, qualidade e crise da educação superior: faces da exclusão e limites da inclusão. **Educação & Sociedade**, v. 31, p. 1223-1245, 2010. Disponível em: <https://www.scielo.br/j/es/a/dFtMDqfdWm75WSc5vKXHCtq/?format=pdf&lang=pt>. Acesso em: 13 jan. 2023.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. Da mineração de dados à descoberta de conhecimento em bancos de dados. **Revista AI**, v. 17, n. 3, p. 37-37, 1996. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1609/aimag.v17i3.1230>. Acesso em: 18 mar. 2023.

FALQUETO, Leonan Entringer; SUTERIO, Ricardo; PAES, Rafael Lemos; PASSARO, Angelo. kNN e Rede Neural Convolucional para o Reconhecimento de Plataformas de Petróleo em Imagens SAR do Sentinel-1. **Simpósio de Aplicação Operacionais em Áreas de Defesa-SIGE**, 2019. São José dos Campos: ITA. Disponível em: https://www.sige.ita.br/edicoes-antiores/2019/st/ST_03_3.pdf. Acesso em: 13 mar. 2023.

FERNÁNDEZ-GARCÍA, Antonio Jesús; RODRÍGUEZ-ECHEVERRÍA, Roberto; PRECIADO, Juan Carlos; CONEJERO MANZANO, José Maria; SÁNCHEZ-FIGUEROA, Fernando. Criação de um Sistema de Recomendação para Apoiar Estudantes do Ensino

Superior na Decisão de Matrícula em Disciplinas. *IEEE Access*, v. 8, p. 20, 2020. Disponível em: <https://ieeexplore.ieee.org/document/9019855>. Acesso em: 8 abr. 2024.

FERREIRA, Caíque Augusto. **Indução de árvore de decisão utilizando meta-aprendizado**. 2022. Tese (Doutorado em Ciências e Letras). Universidade de São Paulo. Disponível em: https://www.teses.usp.br/teses/disponiveis/59/59143/tde-05122022-145141/publico/Dissertacao_Corrigida_Caique_Ferreira.pdf. Acesso em: 13 mar. 2023.

G1. População de Angicos (RN) é de 11.632 pessoas, aponta o Censo do IBGE. G1, 28 jun. 2023. Disponível em: <https://g1.globo.com/rn/rio-grande-do-norte/noticia/2023/06/28/populacao-de-angicos-rn-e-de-11-632-pessoas-aponta-o-censo-do-ibge.ghtml>. Acesso em: 06 fev. 2025.

GUALHANO, Mariana Abreu; VASCONCELOS, Aline Pires Vieira de. Análise dos cursos de licenciatura da Rede Federal utilizando Mineração de Dados. **Educação e Pesquisa**, v. 46, 2020. Disponível em: <http://educa.fcc.org.br/pdf/ep/v46/1517-9702-ep-46-e219576.pdf>. Acesso em: 26 jan. 2023.

GOMES, Alex Sandro; GOMES, Claudia Roberta Araújo. Classificação dos tipos de pesquisa em Informática na Educação. **Jaques, Patrícia Augustin**, 2019. Disponível em: https://metodologia.ceie-br.org/wp-content/uploads/2019/06/livro1_cap4.pdf. Acesso em: 27 fev. 2023.

GOPALAKRISHNAN, Aparna; KASED, Rama; YANG, Hui; AMOR, Maria Beth; GRATEROL, Célia; SHADA, Alycia. *A multifaceted data mining approach to understand what factors lead college students to persist and graduate*. In: **2017 IEEE MIT Undergraduate Research Technology Conference (URTC)**, 2017, Cambridge, MA, USA. Proceedings of the 2017 IEEE MIT Undergraduate Research Technology Conference. Piscataway: IEEE, 2017. p. 1–4. DOI: 10.1109/URTC.2017.8252128. Disponível em: <https://ieeexplore.ieee.org/document/8252128>. Acesso em: 8 abr. 2024.

GÜNTHER, Hartmut. Pesquisa qualitativa versus pesquisa quantitativa: esta é a questão?. *Psicologia: Teoria e Pesquisa*, v. 22, n. 2, p. 201-209, maio/ago. 2006. Disponível em: <https://www.scielo.br/j/ptp/a/HMpC4d5cbXsdt6RqbrmZk3J/?lang=pt#>. Acesso em: 06 fev. 2025.

HORTÊNCIO FILHO, Fernando Wagner Brito; VINUTO, Tiago Silva; LEAL, Bruno Carvalho. **Análise de Classificadores para Predição de Evasão dos Campi de uma Instituição de Ensino Federal**. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 31., 2020. Anais do 31º Simpósio Brasileiro de Informática na Educação. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 1132-1143. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/12869>. Acesso em: 5 set. 2024.

HOSSAIN, Mia; HOSEN, Shams Bin Md. Md Labib; KHAN, Shahidul Islam; MASUM, Abdul Kadar Mohammad. *Predictive Analysis of School Dropout Rate in Bangladesh during COVID-19*. In: **2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)**, 2022, Sakheer, Bahrain. Proceedings of the 2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies. Piscataway: IEEE, 2022. p. 1–6. DOI: 10.1109/3ICT56505.2022.9775898. Disponível em: <https://ieeexplore.ieee.org/document/9775898>. Acesso em: 8 abr. 2025.

IBGE (2019). Agência IBGE/notícias. **PNAD Contínua 2019**: Mais da metade das pessoas de 25 anos ou mais não completaram o ensino médio. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/28285-pnad-educacao-2019-mais-da-metade-das-pessoas-de-25-anos-ou-mais-nao-completaram-o-ensino-medio>. Acesso em 10 dez. 2024.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). Resumo Técnico do Censo da Educação Superior 2020. Brasília: Inep, 2022. Disponível em: https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/resumo_tecnico_censo_da_educacao_superior_2020.pdf. Acesso em 20 nov. 2024.

LOPES, Vagner José; CARVALHO, Deborah Ribeiro; LAHM, Janaina Verônica. KDD na avaliação da usabilidade do prontuário eletrônico do paciente por profissionais da enfermagem. **Revista Brasileira de Inovação Tecnológica em Saúde**, v.6, n.3, 2016. Disponível em: <https://periodicos.ufrn.br/reb/article/view/10414/7952>. Acesso em: 27 jan. 2023.

LORENZETT, Cassio Dal Castel. TELOKEN, Alex. Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão. **Simpósio de Pesquisa e Desenvolvimento em Computação**, v. 2, n. 1, 2016.

MARTINS, Caio Vinicius Monteiro; LACERDA, Filipe Carvalho; CARMO, Igor Procópio do; SILVA, Edmilson Vitorino Scovino da; ALVES, Tatiane Ornelas Martins; GOMES, Johnny Moreira; CAMPOS, Ricardo Silva. Modelos de previsão de evasão tardia na graduação de uma universidade pública. In: **CONGRESSO SOBRE TECNOLOGIAS NA EDUCAÇÃO (Ctrl+e)**, 8., 2023, Santarém/PA. Anais do Congresso sobre Tecnologias na Educação (Ctrl+e). Porto Alegre: Sociedade Brasileira de Computação, 2023. Disponível em: <https://sol.sbc.org.br/index.php/ctrl+e/article/view/25782>. Acesso em: 9 set. 2024.

MCKINNEY, Wes. Python for Data Analysis. 2. ed. Sebastopol: O'Reilly Media, 2017.

MORAES FILHO, Moacir José de Almeida. SOS–Stimulate Our Skills: **guia de estratégias didáticas criativas para o ensino e aprendizagem na licenciatura em letras-língua inglesa**. 2021. Dissertação (Programa de Pós-Graduação Criatividade e Inovação em Metodologias de Ensino Superior do Núcleo de Inovação e Tecnologias Aplicadas a Ensino e Extensão) - Universidade Federal do Pará, 2021. Disponível em: http://repositorio.ufpa.br/bitstream/2011/13684/2/Dissertacao_StimulateOurSkills.pdf. Acesso em: 05 fev. 2023.

NASCIMENTO, Fernanda Ferreira. **Implementação e Avaliação de Diferentes Modelos de Machine Learning Aplicados à Predição de Estudantes em Risco de Evasão Estudantil em Diferentes Cursos do Ensino Superior**. Orientador: Paulo Gabriel Gadelha Queiroz. 2022. p.140. Pós-Graduação (Ciência da Computação), 2022.

NAU, Jonathan; DAZZI, Rudimar Luís Scaranto; HAENDCHEN FILHO, Aluizio; FERNANDES, Anita. Processamento do Discurso em Textos Dissertativos-Argumentativos: Uma Abordagem Baseada em Mineração de Argumentos e Aprendizado Supervisionado de

Máquina. In: **Anais do XLVII Seminário Integrado de Software e Hardware**. SBC, 2020. p. 48-59. Disponível em: <https://sol.sbc.org.br/index.php/semish/article/view/11316/11179>. Acesso em: 01 fev. 2023.

NETTO, Amilcar; MACIEL, Francisco. **Python para Data Science e Machine Learning Descomplicado**. Rio de Janeiro: Editora Alta Books, 2021. *E-book*. ISBN 9786555203172. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9786555203172/>. Acesso em: 06 fev. 2023.

OLIVEIRA, Pamella Letícia Silva; RODRIGUES, Rodrigo Lins; RAMOS, Jorge Luis Cavalcanti; SILVA, João Carlos Sedraz. Identificação de Pesquisas e Análise de Algoritmos de Clusterização para a Descoberta de Perfis de Engajamento. **Revista Brasileira de Informática na Educação**, v. 30, p. 01-19, 2022.

OLIVEIRA, C. V. S. B. de; BEZERRA, D. H. D.; TORRES, G. V. de S. Revisão sistemática da literatura sobre as causas de evasão da educação a distância no Brasil. **Em Rede - Revista de Educação a Distância**, [S. l.], v. 8, n. 1, p. 1–15, 2021. DOI: 10.53628/emrede.v8i1.656. Disponível em: <https://www.aunirede.org.br/revista/index.php/emrede/article/view/656>. Acesso em: 17 mar. 2023.

PEREIRA, Wendell Soares; CYSNEIROS, Gilberto; AGUIAR, Yuska Paola Costa. Diretrizes para o Desenvolvimento de Serious Games: Um Mapeamento Sistemático da Literatura. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2019. p. 714. Disponível em: <http://ojs.sector3.com.br/index.php/sbie/article/view/8799/6357>. Acesso em: 17 mar. 2023.

PROIFES FEDERAÇÃO. [Infográfico Interativo] Impacto dos cortes de investimentos nas universidades federais. 2023.

PAREDES, Alberto Sanchez; DURHAM, Eunice Ribeiro. Evasão do terceiro grau em Curitiba. 1994.

RAFIQ, Md Abdur; RABBI, Abir Mahamud; AHAMMAD, Rasel. A data science approach to predict the university students at risk of semester dropout: Bangladeshi university perspective. *International Journal of Advanced Computer Science and Applications*, v. 12, n. 6, p. 341–348, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9453067>. Acesso em: 20 set. 2024.

RAMOS, Ana Isabela M.; ARAÚJO, Fabíola O. Questões de Gênero e a Evasão de Mulheres nos Cursos de Computação: Um Estudo de Caso na Região Metropolitana de Belém. In: **WOMEN IN INFORMATION TECHNOLOGY (WIT)**, 16., 2022, Niterói. Anais do XVI Women in Information Technology – WIT 2022. Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 239-244. Disponível em: <https://sol.sbc.org.br/index.php/wit/article/view/20880>. Acesso em: 10 set. 2024.

ROLIM, Maria José; ALMEIDA, Danusa Mendes. A evasão estudantil no curso de letras português da FECLESC. **Ensino em Perspectivas**, v. 2, n. 1, p. 1-12, 2021. Disponível em: <https://revistas.uece.br/index.php/ensinoemperspectivas/article/view/4532>. Acesso em: 17 jan. 2023.

SACCARO, Alice; FRANÇA, Marco Túlio Aniceto; JACINTO, Paulo de Andrade. Fatores Associados à Evasão no Ensino Superior Brasileiro: **um estudo de análise de sobrevivência para os cursos das áreas de Ciência, Matemática e Computação e de Engenharia, Produção e Construção em instituições públicas e privadas**. Estudos Econômicos (São Paulo), v. 49, p. 337-373, 2019. Disponível em: <https://www.scielo.br/j/ee/a/9YxHxWkk6Dzy35CpgmxXbPt/abstract/?lang=pt>. Acesso em: 16 jan. 2023.

SALAS, Fabio; CALDAS, Josué. Prevendo o desempenho acadêmico de estudantes de graduação em uma universidade destacada do Peru: **uma abordagem com ferramentas de Machine Learning**. *Educación*, v. 33, n. 64, p. 55-85, 2024. Disponível em: <https://www.redalyc.org/journal/7178/717877776004/>. Acesso em: 21 jul. 2024.

SANTOS, Jennifer Manfrin dos. A utilização da inteligência artificial no julgamento de casos simples para a promoção da processual nos juizados especiais cíveis. 2024. Disponível em: <https://repositorio.uninter.com/handle/1/1721>. Acesso em: 06 mai. 2025.

SANTOS, Gás; BELLOZE, K. T.; TARRATACA, L.; HADDAD, D. B.; BORDIGNON, A. L.; BRANDÃO, D. N. *EvolveDTree: analisando a evasão escolar nas universidades*. In: INTERNATIONAL CONFERENCE ON SYSTEMS, SIGNALS AND IMAGE PROCESSING (IWSSIP), 2020, Niterói, RJ. *Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. Piscataway: IEEE, 2020. p. 261–266. Disponível em: <https://ieeexplore.ieee.org/document/9157882>. Acesso em: 8 jul. 2024.

SANTOS, Patricia; GOYA, Denise. Aprendizado de Máquina Aplicado à Análise de Evasão em Cursos de Sistemas de Informação. In: FÓRUM DE EDUCAÇÃO EM SISTEMAS DE INFORMAÇÃO – SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI), 16., 2020, Evento Online. *Anais do Fórum de Educação em Sistemas de Informação - SBSI 2020*. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 210-213. DOI: <https://doi.org/10.5753/sbsi.2020.13145>. Disponível em: https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/13145. Acesso em: 11 set. 2024.

SANTOS, Carlos Henrique D. C.; MARTINS, Simone de Lima; PLASTINO, Alexandre. É possível prever evasão com base apenas no desempenho acadêmico? In: **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE)**, 31., 2020, Recife. *Anais do 31º SBIE*. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 686–695. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/18107>. Acesso em: 7 nov. 2024.

MACHADO SEGUNDO, Hugo de Brito (ed.). **Direito e Inteligência Artificial**: O que os Algoritmos têm a Ensinar sobre Interpretação, Valores e Justiça. Indaiatuba SP: Roberta Densa, 2023. *E-book* (348p.) P&B. ISBN: 13334-050. DOI: 10.2022. Disponível em: [https://books.google.com.br/books?hl=pt-BR&lr=&id=OFyYEAQAQBAJ&oi=fnd&pg=PT2&dq=A+evolu%C3%A7%C3%A3o+do+ML+se+deu+atrav%C3%A9s+da+Intelig%C3%AAncia+Artificial+\(IA\)+do+ingl%C3%AAs,+Artificial+Intelligence+\(AI\).+Desde+do+in%C3%ADcio+os+pesquisadores+venham+se+interessando+construir+m%C3%A1quinas+que+pude+sem+desenvolver+algoritmos+capazes+de+aprender+com+os+erros+&ots=jynoMuaPPD&sig=oKxphctXx8HBM2rjDw9hMWJTak#v=onepage&q&f=false](https://books.google.com.br/books?hl=pt-BR&lr=&id=OFyYEAQAQBAJ&oi=fnd&pg=PT2&dq=A+evolu%C3%A7%C3%A3o+do+ML+se+deu+atrav%C3%A9s+da+Intelig%C3%AAncia+Artificial+(IA)+do+ingl%C3%AAs,+Artificial+Intelligence+(AI).+Desde+do+in%C3%ADcio+os+pesquisadores+venham+se+interessando+construir+m%C3%A1quinas+que+pude+sem+desenvolver+algoritmos+capazes+de+aprender+com+os+erros+&ots=jynoMuaPPD&sig=oKxphctXx8HBM2rjDw9hMWJTak#v=onepage&q&f=false). Acesso em: 31 jan. 2023.

SERRANO, Catarina; SERRA, Yenny; RIOS-CRUZ, Sônia-Gisela; PARDO, Fredy. Modelos de aprendizado de máquina para prever a evasão acadêmica em estudantes universitários. In: **CONGRESSO INTERNACIONAL DE COMPUTAÇÃO, 2024**. Anais do Congresso Internacional de Computação. Piscataway: IEEE, 2024. Disponível em: <https://ieeexplore.ieee.org/document/10854829>. Acesso em: 8 abr. 2025.

SILVA FILHO, Roberto Leal Lobo e; MONTEJUNAS, Paulo Roberto; HIPÓLITO, Oscar; LOBO, Maria Beatriz Carvalho Melo. A Evasão no Ensino Superior Brasileiro. **Cadernos de Pesquisa**. São Paulo, v. 37, n. 132, set./dez. 2007. Disponível em: <https://publicacoes.fcc.org.br/cp/article/view/346/350>. Acesso em: 5 mai. 2025.

SILVA, Vislene Querino Fofano da. **Política de permanência no Ensino Superior no Brasil e formação de qualidade**. 2020. 142 f. Dissertação (Mestrado em Educação) – Universidade Federal de Juiz de Fora, Juiz de Fora, 2020. Disponível em: <https://repositorio.ufjf.br/jspui/handle/ufjf/11590>. Acesso em: 5 mai. 2025.

SILVA, Fernanda Cristina da. **Variáveis para modelos preditivos à evasão na educação superior**. 2021. Tese de Doutorado, Universidade Federal de Santa Catarina, Florianópolis. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/231181/PCAD1180-T.pdf?sequence=1&isAllowed=y>. Acesso em: 17 jan. 2023 .

SILVA, José Marcos Toscano Da; FREITAS, Manoel Cosme Pimenta de Freitas. **Evasão no curso superior em pedagogia da UAB: campus Laranjal do Jari**. 2022. Disponível em: <http://repositorio.ifap.edu.br/jspui/bitstream/prefix/729/3/SILVA%20%282022%29%20%e2%80%93%20Evas%20no%20curso%20superior.pdf>. Acesso em: 31 jan. 2023.

SILVA, Francisco Daniel Lima da. Implementando adaptação em ambientes inteligentes utilizando sistemas multiagentes e aprendizado por reforço. Orientador: Marcos Antonio de Oliveira. 2019. Monografia (Sistema de Informação) – Universidade Federal do Ceará, 2019. Disponível em: https://repositorio.ufc.br/bitstream/riufc/49760/1/2019_tcc_fldasilva.pdf. Acesso em: 14 mar. 2023.

SOUZA, Maria Caiane Lino; RAMOS, Maria Eduarda Lacerda; SERAFIM, Alexandre do Prado Caldas. Evasão Escolar e Participação no Ensino Remoto. In: **VIII Encontro Nacional das Licenciaturas**, 2021. Online. Disponível em: https://www.editorarealize.com.br/editora/anais/enalic/2021/TRABALHO_EV163_MD3_SA101_ID817_26102021161120.pdf. Acesso em: 17 jan. 2023.

SOUZA, Joubert Alexandrino de; KOMATI, Karin Satie; ANDRADE, Jefferson Oliveira. Análise de Sobrevivência: um estudo de caso em um Curso de Sistemas de Informação. In: **WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO (WEI)**, 2022. Anais do WEI 2022. Porto Alegre: Sociedade Brasileira de Computação, 2022. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/20847>. Acesso em: 9 set. 2024.

TARIQ, Arham; AMIN, Ahmad; MASOOD, Yasser; MUZAFFAR, Muhammad; IQBAL, Junaid. Predicting Early Dropout of University Students: A Comparative Study between KNN and Decision Tree. In: **2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)**, 2022, Prague, Czech Republic. Proceedings of the 2022 ICECET. Piscataway: IEEE, 2022. p. 1–6. DOI: 10.1109/ICECET55527.2022.10089706. Disponível em: <https://ieeexplore.ieee.org/document/10089706>. Acesso em: 8 jun. 2024

TEODORO, L. de A.; KAPPEL, M. A. A. Aplicação de Técnicas de Aprendizado de Máquina Para Predição de Risco de Evasão Escolar em Instituições Públicas de Ensino Superior no Brasil. *Revista Brasileira de Informática na Educação*, v. 28, p. 838–863, 2020. Disponível em: <https://journals-sol.sbc.org.br/index.php/rbie/article/view/3691>. Acesso em: 6 set. 2024.

TOMIM, Solange Rodrigues. Os desafios da escola pública paranaense na perspectiva do professor PDE. Orientador: Marco Antonio Batista Carvalho, Professora PDE, 2016/201. Graduação (Pedagogia). Disponível: http://www.diaadiaeducacao.pr.gov.br/portals/cadernospde/pdebusca/producoes_pde/2016/2016_artigo_ped_unioeste_solangerodrigues.pdf. Acesso em: 10 jan. 2023.

UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO. DEMONSTRATIVO DRA, 2022.1. **DIVISÃO DE REGISTRO ACADÊMICO**, [S. l.], 24 set. 2014. Disponível em: <https://dra.ufersa.edu.br/estatisticas/>. Acesso em: 7 fev. 2023.

VIANA, Filipe Soares; SANTANA, André Macêdo; RABÊLO, Ricardo de Andrade Lira. Avaliação de Classificadores para Predição de Evasão no Ensino Superior Utilizando Janela Semestral. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 33., 2022, Manaus. Anais do 33º SBIE. Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 908–919. DOI: <https://doi.org/10.5753/sbie.2022.224764>. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/22469>. Acesso em: 3 nov. 2024.

WEGNER, Rubia Cristina. Evasão no ensino superior: Digressões motivadas a partir da pandemia do novo coronavírus. **Revista Docência e Cibercultura**, v. 6, n. 1, p. 01-22, 2022. Disponível em: <https://www.e-publicacoes.uerj.br/index.php/re-doc/article/view/67043>. Acesso em: 26 dez. 2023.

WIERINGA, Roel. **Design science methodology for information systems and software engineering**. Springer-Verlag Berlin Heidelberg, 2014.

APÊNDICE A - QUESTIONÁRIO APLICADO COM OS PARTICIPANTES

Pesquisa de Mestrado

Esta pesquisa científica consiste em um estudo de caso planejado pela Mestranda Camyla Micaely Silveira Peixoto, orientada pelo professor Sebastião Emidio Alves Filho do Programa de Pós-Graduação em Ciência da Computação UERN/UFERSA. O tema do estudo é a "Evasão dos Discentes nos Cursos da UFERSA/Angicos". Esse formulário será aplicado na intenção de coleta de dados que contribua com a pesquisa da Defesa de Mestrado. O formulário não requer identificação pessoal e nenhum dado coletado será utilizado para identificar o respondente. As respostas fornecidas serão usadas somente na pesquisa da mestranda e em outras pesquisas relacionadas ao tema.

Agradeço antecipadamente pela colaboração,
Camyla Micaely Silveira Peixoto

Curso:

Semestre de ingresso:

1. Como é o seu deslocamento até a universidade (Tipo de transporte)?

() Carro. () Moto. () Ônibus. () A pé.

() Outro: _____

2. Com relação ao transporte do item anterior, ele é:

() Próprio. () Cedido. () Público(gratuito).

() Particular(táxi/moto-táxi). () Não se aplica.

3. O transporte representa uma barreira/dificuldade para frequentar a universidade?

() Sim, sempre. () Sim, às vezes. () Não, mas já foi.

() Não, nunca foi.

4. Você mora em Angicos ?

() Sim. () Não.

() Durmo em Angicos nos dias de aula e atividades.

5. Quanto tempo, em média, você demora para se deslocar todos os dias até o campus (ida e volta)? _____

6. Você sofreu algum tipo de preconceito ou violência durante o Curso relativo a (marque todas que forem aplicáveis):

() Cor de pele. () Condição financeira. () Aparência.

() Deficiência. () Dificuldade de aprendizado. () Gênero.

() Curso. () Outro: _____

7. Você se identifica com o curso que está fazendo?

() Sim. () Não, mas quero concluir. () Não, não sei se concluirei.

8. Como você considera a acessibilidade no campus ?

() Adequada. () Inadequada - Especificar:

9. Em relação ao tempo necessário como discente para dedicar no estudo?

() É suficiente.

() É insuficiente, mas desempenho a maioria das atividades.

() É insuficiente, mas só realizo as atividades obrigatórias.

() É insuficiente e não consigo realizar as atividades obrigatórias.

10. Você trabalha?

() Sim, tenho empresa própria / sou autônomo / profissional libera.

() Sim, com emprego formal. () Sim, com trabalho temporário.

() Sim, com emprego informal. () Sou bolsista da Universidade.

() Estou em estágio remunerado. () Não trabalho ainda.

11. Se você trabalha, em quais horários?

() Tempo integral ou dois turnos. () Tempo parcial ou um turno.

() Horário corrido ou 6 horas diárias. () Em regime de escala ou plantão. () Sem dias e horários fixos.

12. É casado(a) / está em união estável: () Sim. () Não.

13. Tem filhos? () Não. () Sim - Quantos: _____

14. Você contribui para o sustento financeiro da família:

() Sim, sou o único com renda. () Sim, sou a principal.

() Sim, mas não sou o principal. () Não contribuo.

15. Cite as principais dificuldades para sua permanência no Curso?

16. Já pensou em trancar disciplinas, período ou abandonar o curso? Se sim, porquê?

