A logo with a face and a moon

Description automatically generatedUNIVERSIDADE FEDERAL DO RECONCAVO BAIANO

CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICA

GUILHERME EUGÊNIO BISPO NASCIMENTO DE OLIVEIRA

Análise Preditiva de Evasão no Curso de Ciências Exatas Utilizando Machine Learning:

Um Estudo de Caso com Dados do Bacharelado em Ciências Exatas e Tecnológicas.

Cruz das Almas- Bahia

2024

GUILHERME EUGÊNIO BISPO NASCIMENTO DE OLIVEIRA

Análise Preditiva de Evasão em Cursos de Ciências Exatas Utilizando Machine Learning:

Um Estudo de Caso com Dados do Bacharelado em Ciências Exatas e Tecnológicas.

Trabalho de Conclusão de Curso apresentado ao Curso Ciências Exatas e Tecnológica da Universidade Federal do Reconcavo Baiano, como requisito parcial para obtenção do título de bacharel.

Área de concentração: Machine Learning

Orientador: Camila Bezerra da Silva

Coorientador:

Cruz das Almas - Bahia

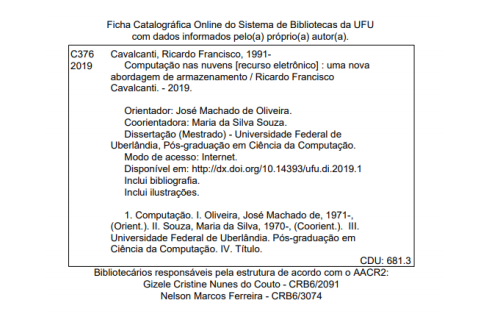
2024

A  **ficha** catalográfica contém as informações bibliográficas (autor, título, local de publicação, assuntos e outras) necessárias para identificação da obra.

A confecção da ficha catalográfica para teses e dissertações é feita de forma automatizada dentro do Portal do Aluno.

**Acesse**: [Tutorial para confecção da Ficha Catalográfica Online](http://www.bibliotecas.ufu.br/sites/bibliotecas.ufu.br/files/media/documento/ficha_catalografica_online_-_tutorial.pdf) **Acesse**: [www.portalestudante.ufu.br](http://www.portalestudante.ufu.br/)

Modelo:



NOME COMPLETO DO AUTOR

**Folha de aprovação** - **elemento pré-textual obrigatório**

Título do trabalho:

subtítulo (se houver)

Trabalho de Conclusão de Curso apresentado ao Curso Ciências Exatas e Tecnológica da Universidade Federal do Reconcavo Baiano, como requisito parcial para obtenção do título de bacharel.

Área de concentração: Machine Learning

Orientador: Camila Bezerra da Silva

Coorientador:

Cruz das Almas, Julho de 2024

Banca Examinadora:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome – Titulação (sigla da instituição)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome – Titulação (sigla da instituição)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome – Titulação (sigla da instituição)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome – Titulação (sigla da instituição)

**Dedicatória** - **elemento pré-textual opcional**

Dedico este trabalho aos meus pais, pelo estímulo, carinho e compreensão.

**AGRADECIMENTOS**

Agradeço ao professor e amigo... o incentivo, motivação e orientação nesta caminhada acadêmica.

Aos colegas... e...,

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa concedida durante os anos do curso.

Agradeço também, ao coordenador do setor..., da Universidade..., por permitir o meu afastamento....

**Agradecimentos** - **elemento pré-textual opcional**

**Epígrafe** - **elemento pré-textual opcional**

“Ninguém ignora tudo. Ninguém sabe tudo. Todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa.”

(FREIRE, 2002, p. 69)

**RESUMO**

Elemento obrigatório elaborado conforme NBR6028 da Associação Brasileira de Normas Técnicas (2003), apresentado em um só bloco de texto sem recuo de parágrafo consistindo na apresentação concisa das ideias do texto completo. Deve descrever de forma clara e sintética a natureza do trabalho, o objetivo, o método, os resultados e as conclusões, visando fornecer elementos para o leitor decidir sobre a consulta ao texto completo. Deve ser redigido em linguagem clara e objetiva, ser inteligível por si mesmo, empregar verbos na voz ativa e na terceira pessoa do singular e conter de 150 a 500 palavras. Deve-se evitar o uso de símbolos abreviaturas, fórmulas, quadros, equações. Após o texto do resumo, seguem as palavras-chave representativas do conteúdo do trabalho, que devem aparecer após um espaço em branco de 1,5, à margem esquerda, separadas entre si por ponto final.

**Palavras-chave**: Trabalhos acadêmicos. Formatação de documentos. Normalização bibliográfica

**Resumo** - **elemento pré-textual obrigatório**

**ABSTRACT**

Elemento obrigatório para trabalhos de conclusão de curso (graduação ou especialização), mestrado e doutorado. É a versão do resumo em português para outro idioma, neste exemplo em inglês. Deve aparecer em página distinta e seguindo a mesma formatação do resumo em português. A versão das palavras-chave em outro idioma deve seguir a mesma formatação utilizada no resumo em português.

**Keywords:** Academic Works. Abstracts. Documents formatting.

**LISTA DE ILUSTRAÇÕES**

**Abstract** - **elemento pré-textual obrigatório**

|  |  |  |
| --- | --- | --- |
| Gráfico 1 - | Xxxxxxxxxxxxxx............................................................................ | x |
| Fotografia 1 - | Xxxxxxxxxxxxxx............................................................................. | xx |
| Quadro 1 - | Xxxxxxxxxxxxxx ............................................................................ | xx |
| Gráfico 2 - | Xxxxxxxxxxxxxx ............................................................................ | xx |

**LISTA DE TABELAS**

**Lista de ilustrações** -  **elemento pré-textual opcional**

|  |  |  |
| --- | --- | --- |
| Tabela 1 - | Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx............... | x |
| Tabela 2 - | Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx............... | xx |
| Tabela 3 - | Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx............... | xx |

**LISTA DE ABREVIATURAS E SIGLAS**

**Lista de tabelas - elemento pré-textual opcional**

|  |  |
| --- | --- |
| ABNT | Associação Brasileira de Normas Técnicas |
| CNPq | Conselho Nacional de Desenvolvimento Científico e Tecnológico |
| IBGE | Instituto Brasileira de Geografia e Estatistica |
| UFU | Universidade Federal de Uberlândia |
|  |  |

**SUMÁRIO**

**Lista de abreviaturas - elemento pré-textual opcional**

[1 INTRODUÇÃO 12](#_Toc17219506)

[2 REVISÃO BIBLIOGRÁFICA 1](#_Toc17219508)3

[3 METODOLOGIA 15](#_Toc17219509)

[4 RESULTADOS 16](#_Toc17219510)

[5 DISCUSSÃO 17](#_Toc17219511)

[6 CONCLUSÃO 18](#_Toc17219512)

[REFERÊNCIAS 18](#_Toc17219513)

[APÊNDICE A – TÍTULO DO APÊNDICE 20](#_Toc17219514)

[APÊNDICE B – ORIENTAÇÕES SOBRE APRESENTAÇÃO DO TRABALHO 21](#_Toc17219515)

[ANEXO A – TÍTULO DO ANEXO 27](#_Toc17219516)

[ANEXO B – MODELO DE SUMÁRIO 28](#_Toc17219517)

[ANEXO C – MODELO DE LISTA DE REFERÊNCIAS 29](#_Toc17219518)

**Sumário - elemento pré-textual obrigatório elaborado de acordo com a NBR 6024 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2012b). Pode ser automático ou não. O sumário acima está no modo automático, podendo ser removido em “Referências”/ Sumário / Remover sumário.**

## 

# 1 INTRODUÇÃO

O ensino superior enfrenta um dilema complexo, onde a missão de educar e formar profissionais qualificados se confronta com desafios estruturais que muitas vezes resultam em altas taxas de evasão (Amare; Šimonová, 2021). Esse cenário torna-se ainda mais desafiador nos cursos de Ciências Exatas onde a demanda por habilidades especializadas é alta, mas a evasão representa um obstáculo significativo para a formação desses profissionais (Belli, 2006; Matos *et al*., 2023).

Nestes cursos, os estudantes frequentemente se deparam com uma série de desafios, que vão desde a complexidade dos conteúdos até a falta de preparo prévio em disciplinas fundamentais como matemática e física (Talbert, 2019; Moreira, 2021). Esse descompasso entre as habilidades exigidas e as habilidades adquiridas muitas vezes resulta em dificuldades de aprendizagem e, consequentemente, em taxas elevadas de evasão (Lock, 2021).

Para agravar ainda mais a situação, a evasão no ensino superior é um fenômeno multifacetado, influenciado por uma série de fatores individuais, acadêmicos, socioeconômicos e institucionais (Goldschmidt; Wang, 1999; Silva; Díaz, 2023). Identificar e abordar eficazmente esses fatores requer uma compreensão profunda dos padrões de comportamento dos estudantes e a capacidade de prever com precisão quais alunos estão em maior risco de desistência (Amare; Šimonová, 2021).

Nesse contexto, a utilização de métodos e técnicas avançadas, como análise de dados e aprendizagem de máquina, emerge como uma abordagem promissora para enfrentar esse desafio. Esses modelos podem examinar uma variedade de dados, incluindo desempenho acadêmico, participação em atividades extracurriculares, uso de recursos da instituição e até mesmo dados demográficos e socioeconômicos dos estudantes (Oqaidi; Aouhassi; Mansouri, 2022). Com base nesses dados, os modelos de Machine Learning podem identificar padrões sutis e correlações que escapam à percepção humana, permitindo uma intervenção mais proativa e direcionada (Pal, 2012).

Diante desse cenário, o objetivo deste trabalho é identificar padrões nos alunos que evadem do curso de Bacharelado Em Ciências Exatas e Tecnológicas da Universidade Federal do Recôncavo da Bahia (UFRB), utilizando técnicas de análise de dados e modelos de machine learning. A proposta visa fornecer insights que possam auxiliar a Pró-Reitoria de Assuntos Estudantis (PROPAAE) na criação de estratégias mais eficazes para intervir e prevenir a evasão dos alunos assistidos, promovendo assim a retenção e o sucesso acadêmico no Bacharelado em Ciências Exatas e Tecnológicas.

# 2 REVISÃO BIBLIOGRÁFICA

* 1. **EVASÃO ESCOLAR NO NÍVEL SUPERIOR**

A evasão escolar no ensino superior é um fenômeno complexo e multifacetado, que suscita o interesse de diversas áreas acadêmicas. Teorias como a Teoria da Integração de Tinto (1975), que destaca a importância da integração social e acadêmica dos estudantes para a permanência na universidade, e a Teoria do Capital Humano de Mincer (1958), que enfatiza o investimento em educação como um meio de aumentar as oportunidades no mercado de trabalho, oferecem diferentes perspectivas sobre os fatores que influenciam a decisão dos estudantes de abandonar os estudos.

Fatores individuais, como habilidades acadêmicas, motivação e saúde mental, juntamente com fatores institucionais, como qualidade do ensino, suporte acadêmico e políticas de retenção, desempenham papéis significativos nesse processo (Long; Noor,2023). Aspectos socioeconômicos e culturais, como renda familiar, etnia, gênero e cultura, também exercem influência na evasão, afetando o acesso, permanência e sucesso dos estudantes na educação superior (Santos; Ortiz; Pilatti, 2022).

Para lidar com esse desafio, têm sido implementadas intervenções e estratégias de retenção, como tutoria acadêmica, orientação vocacional e apoio financeiro. Estudos destacam que a evasão no ensino superior pode ser motivada por diversos fatores, como falta de perspectiva na carreira, baixo comprometimento com o curso e baixa participação em atividades acadêmicas (Polomino; Ortega, 2023; Silva; Díaz, 2023).

Além disso, a evasão pode ser influenciada por aspectos como acolhimento ruim por parte de colegas e professores, falta de identidade relacionada ao curso e falta de estratégias metodológicas alternativas. A complexidade desse fenômeno requer uma compreensão aprofundada, especialmente considerando as diferentes realidades e cenários em que se manifesta (Daitx; Loguercio; Strack, 2016).

Portanto, enfrentar esse desafio demanda abordagens multidisciplinares e políticas educacionais eficazes, considerando a complexidade dos fatores que influenciam a decisão dos estudantes de abandonar os cursos. A implementação de estratégias de identificação precoce e suporte personalizado, baseadas em uma compreensão aprofundada dos padrões de comportamento dos estudantes, é essencial para mitigar as altas taxas de evasão e promover a conclusão bem-sucedida dos cursos no ensino superior (Lima; Zago, 2018).

* 1. **MODELOS DE MACHINE LEARNING**

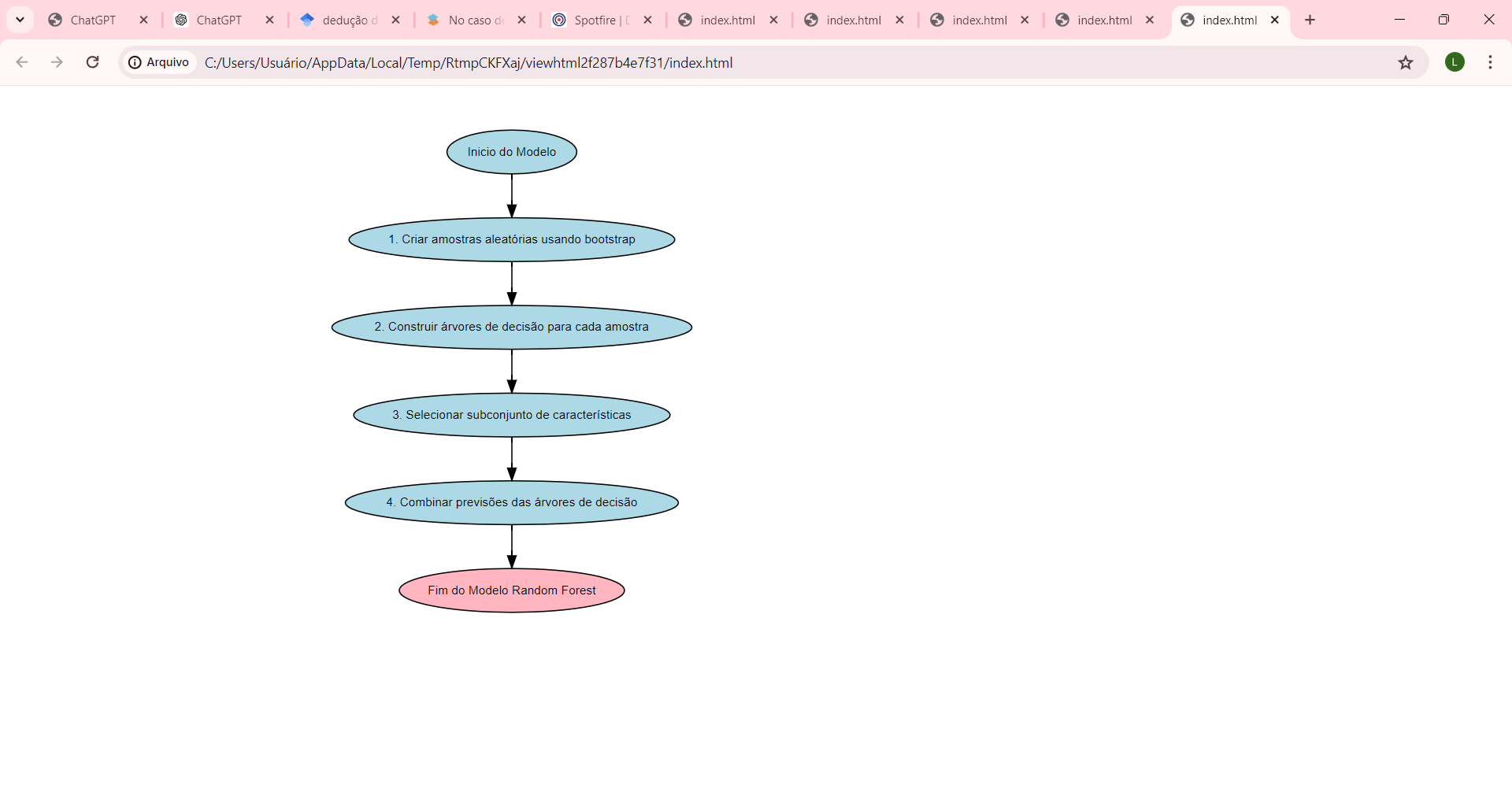
Os modelos de machine learning desempenham um papel crucial na análise de grandes volumes de dados, oferecendo insights valiosos e previsões precisas (Kaneko, 2023). Esses modelos são eficazes na identificação de preditores significativos, no tratamento de multicolinearidade e na descoberta de padrões complexos que os métodos estatísticos tradicionais frequentemente não conseguem captar (Grebovic *et al*., 2023). Entre os muitos modelos disponíveis, o Random Forest e a Regressão Logística destacam-se pela sua popularidade e eficiência em diversas aplicações (Yang *et al*., 2006; Mesquita, 2014; Gonzalez, 2018). Além destes, as redes neurais têm emergido como ferramentas poderosas para lidar com dados complexos e não lineares, ampliando significativamente as capacidades analíticas no contexto educacional (Hilbert *et al*., 2021).

* + 1. **Random Forest**

O Random Forest é um algoritmo de aprendizado supervisionado que constrói múltiplas árvores de decisão durante o treinamento e utiliza a média das predições dessas árvores para melhorar a precisão e controlar o overfitting (Svetnik *et al*., 2003). Ele é eficaz para conjuntos de dados grandes e com alta dimensionalidade (Ho, 1995).

Este modelo pode ser descrito em quatro etapas principais: primeiro, várias amostras aleatórias do conjunto de dados original são criadas utilizando a técnica de bootstrap (Figura 1). Em seguida, uma árvore de decisão é construída para cada amostra, permitindo que a árvore cresça até sua máxima profundidade sem podas. Na terceira etapa, um subconjunto de características é selecionado aleatoriamente em cada nó, e a melhor divisão é determinada com base nesse subconjunto. Por fim, as predições de todas as árvores são combinadas, por votação no caso de classificação ou média no caso de regressão, para produzir a predição final (Breiman, 2001; Yang *et al*., 2006).

**Figura 1**. Fluxograma do Modelo de Random Forest



Fonte: Próprio autor.

A equação do modelo Random Forest pode ser descrita de forma geral, considerando o processo de combinação das previsões individuais das árvores de decisão que compõem a floresta (Amit; Geman, 1997). Para simplificação, apresenta-se a formulação para problemas de regressão e classificação, descritos a seguir segundo o apresentado por Cutler, Cutler e Stevens (2012).

No caso de regressão, a previsão final ( é a média das previsões de todas as árvores:

Aqui, representa a previsão da-ésima árvore.

No caso de classificação, a previsão final é determinada pela votação majoritária das previsões das árvores. Formalmente, é a previsão da -ésima árvore para a classe :

Alternativamente, pode-se usar uma abordagem probabilística, onde a probabilidade de cada classe é a média das probabilidades preditas pelas árvores:

E a previsão final é a classe com a maior probabilidade:

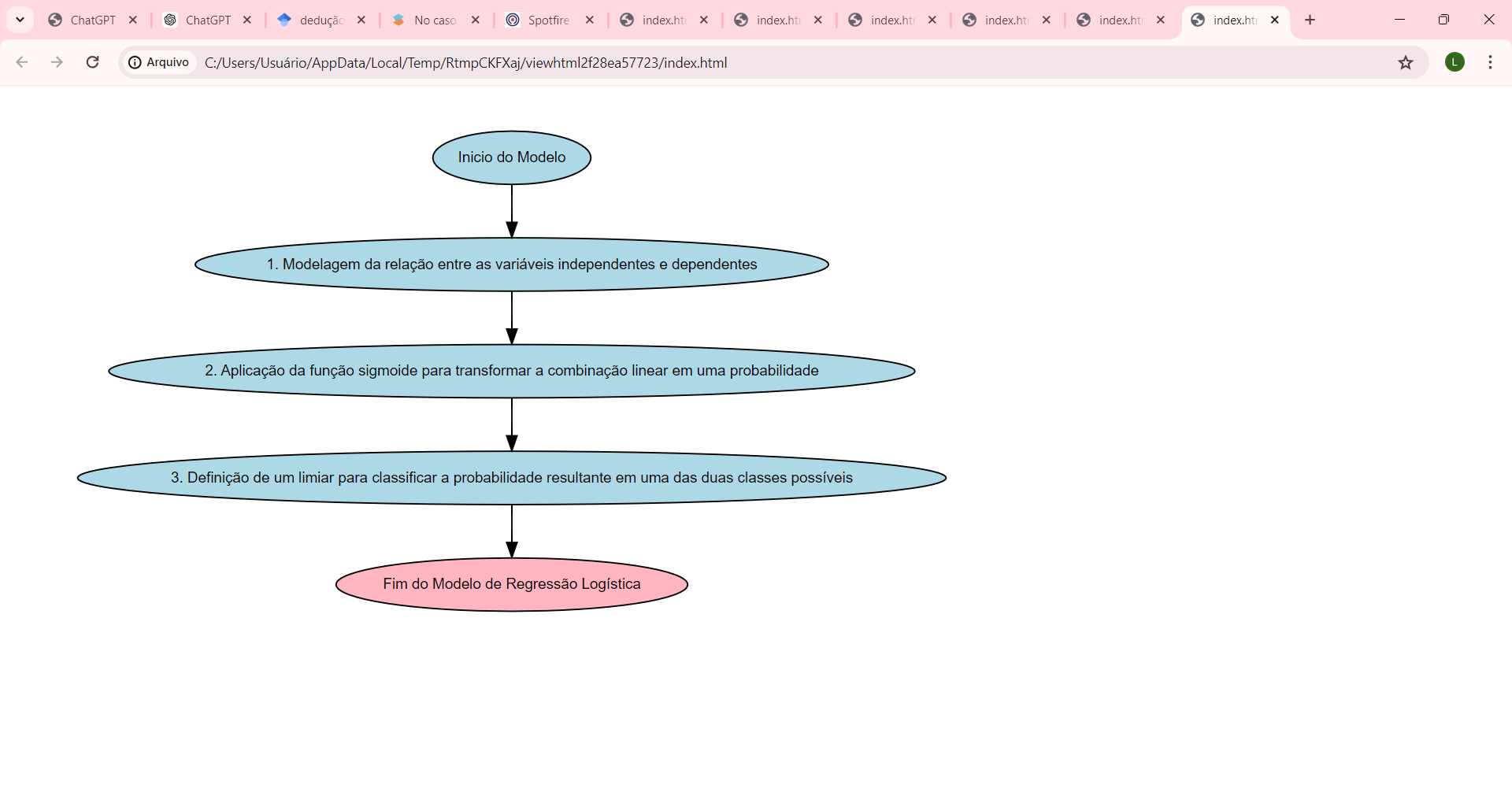
Essas equações capturam a essência do processo de agregação no Random Forest, onde a robustez do modelo decorre da combinação das previsões de múltiplas árvores, cada uma treinada em diferentes amostras dos dados e com diferentes subconjuntos de características.

* + 1. **Regressão Logística**

A Regressão Logística é um modelo estatístico amplamente utilizado em diversas áreas, como marketing, medicina e ciências, para prever a probabilidade de ocorrência de eventos binários, como a compra de um produto por um cliente ou o desenvolvimento de uma doença por um paciente (Mesquita, 2014).

Ao contrário da regressão linear, que prevê valores contínuos, a regressão logística mapeia a saída para um valor entre 0 e 1 por meio da função sigmoide, o que a torna ideal para problemas de classificação binária (Gonzalez, 2018). Esse modelo pode ser descrito em três etapas principais: modelagem da relação entre as variáveis independentes e dependentes, aplicação da função sigmoide para transformar a combinação linear em uma probabilidade, e definição de um limiar para classificar a probabilidade resultante em uma das duas classes possíveis (figura 2) (Sayed, 2022).

**Figura 2**. Fluxograma do Modelo de Regressão Logística



Fonte: Próprio autor.

Para deduzir o modelo de regressão logística, é necessário seguir uma série de etapas que envolvem a relação entre as variáveis independentes e dependentes, a aplicação da função sigmoide e a definição de um limiar. Incialmente, suponha-se que há um conjunto de dados com observações e variáveis independentes (Hosmer; Lemeshow, 1989). A variável dependente binária é denotada como , onde ∈ {0,1}. As variáveis independentes são representadas pelo vetor . O objetivo é modelar a probabilidade de dado x.

Em seguida, aplica-se a função sigmoide, sendo-a utilizada para transformar a combinação linear das variáveis independentes em uma probabilidade. A combinação linear é dada por:

Onde é o intercepto e e são os coeficientes das variáveis independentes. A função sigmoide, também conhecida como função logística, é definida como:

Assim, a probabilidade de dado x é:

Para classificar a probabilidade resultante em uma das duas classes possíveis, define-se um limiar 𝑡 (geralmente 𝑡 = 0.5). Se a probabilidade for maior ou igual a 𝑡, a observação é classificada como y =1; caso contrário, é classificada como y=0.

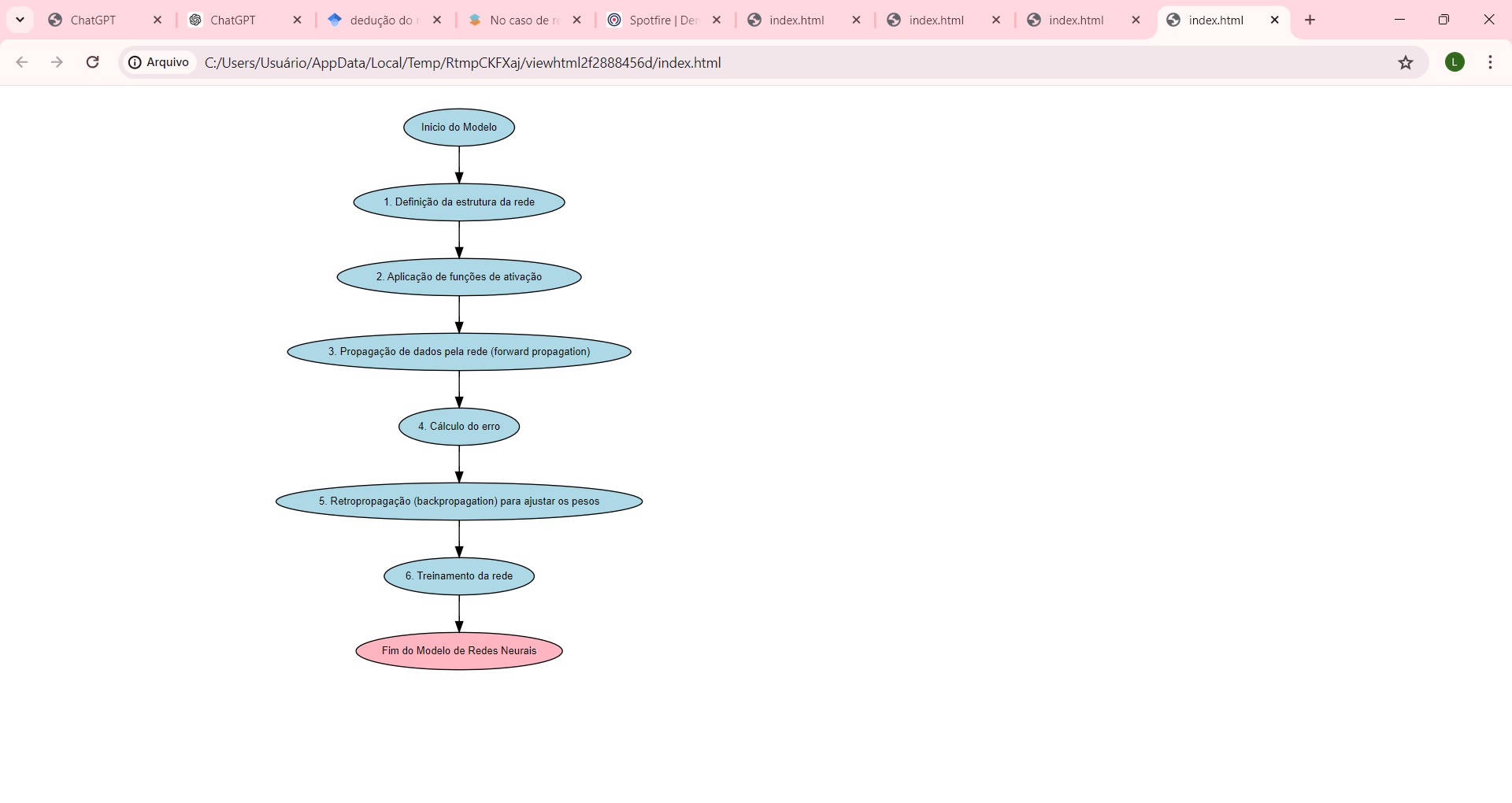
* + 1. **Redes Neurais**

Além dos modelos tradicionais, as redes neurais têm mostrado grande potencial na análise de dados complexos. Redes neurais artificiais (ANNs), inspiradas na estrutura e funcionamento do cérebro humano, são particularmente úteis para modelar relações não lineares e complexas entre variáveis, os quais consistem em camadas de nós (neurônios) que processam os dados e aprendem padrões por meio de um processo de treinamento (Lek et al, 1996). Redes neurais são especialmente eficazes em tarefas de classificação e previsão, onde a relação entre as variáveis de entrada e a saída não é óbvia ou linear. Sendo inspiradas no funcionamento do cérebro humano, são eficazes em identificar combinações complexas de fatores que levam a resultados como a evasão escolar, considerando variáveis demográficas, acadêmicas, socioeconômicas e comportamentais (Rigo *et al.*, 2014).

As redes neurais profundas (deep learning), que consistem em várias camadas de neurônios, são ainda mais poderosas, as quais capturam hierarquias de características complexas nos dados, permitindo uma análise mais detalhada e precisa (Wang *et* al., 2018). Isso é particularmente útil em grandes conjuntos de dados educacionais, onde as interações entre diferentes fatores podem ser altamente complexas e multifacetadas.

Para deduzir o modelo de redes neurais, seguem-se uma série de etapas que envolvem a definição da estrutura da rede, a aplicação de funções de ativação, a propagação de dados pela rede (forward propagation), o cálculo do erro, a retropropagação (backpropagation) para ajustar os pesos e o treinamento da rede (Figura 3) (Silva; Spotti; Flauzino, 2016).

**Figura 3**. Fluxograma do Modelo de Redes Neurais



Fonte: Próprio autor.

Primeiro, uma rede neural é composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada camada é formada por unidades de processamento chamadas neurônios. As camadas são conectadas por pesos que ajustam a força do sinal de uma camada para a próxima. A camada de entrada recebe os dados de entrada, as camadas ocultas processam os dados extraindo características relevantes e a camada de saída produz a saída final da rede.

Cada neurônio realiza uma combinação linear dos sinais de entrada e aplica uma função de ativação para introduzir não-linearidade no modelo. A combinação linear em uma camada é dada por:

Onde: é a combinação linear na camada para o neurônio ; é o peso da conexão entre o neurônio na camada e o neurônio na camadaé a ativação do neurônio na camada; é o viés do neurônio na camada A função de ativação pode ser, por exemplo, a função sigmoide, ReLU, ou a tangente hiperbólica, e a ativação do neurônio é dada por:

Na propagação para frente (forward propagation), os dados são propagados da camada de entrada para a camada de saída através das camadas ocultas. O vetor de dados de entrada é denotado por x. Nas camadas ocultas, a ativação é calculada como:

O erro é calculado na camada de saída utilizando uma função de custo , onde o *y* é o valor real e o é o valor predito. Dado pelo erro quadrático médio de:

Na retropropagação (backpropagation), os gradientes do erro em relação aos pesos e vieses são calculados e usados para atualizar os pesos da rede. O gradiente do erro na camada 𝑙 é dado por:

Em que o o é gradiente do erro na camada 𝑙 e é a derivada da fubção de ativação. Os pesos e vieses são atualizados usando um algoritmo de otimização, como o gradiente descendente. Sendo o processo de forward propagation, cálculo do erro, backpropagation e atualização dos pesos é repetido iterativamente para cada exemplo do conjunto de treinamento até que a rede neural aprenda a mapear os dados de entrada para a saída desejada com precisão.

# 4 METODOLOGIA

A descrição das técnicas usadas deve ser precisa e clara permitindo ao leitor a compreensão do trabalho, e tornar possível que outros pesquisadores repitam na íntegra o mesmo método.

# 5 RESULTADOS

**Parte** **do** **desenvolvimento**

A apresentação dos resultados deve ser clara, objetiva, lógica e acompanhada de comentários. A apresentação de tabelas e ilustrações facilita de maneira extraordinária esta parte do texto. Os dados obtidos, mesmo quando em grande quantidade, devem fazer parte do trabalho, mesmo que em forma de anexo.

# 6 DISCUSSÃO

**Parte** **do** **desenvolvimento**

Nessa etapa são comparados, avaliados e criticados os resultados. Discute-se o valor absoluto e relativo dos resultados. Da apresentação dos fatos pode-se passar para deduções paralelas, generalização cautelosa e enumeração das questões que ocorrem do autor para as quais não encontrou resposta e que requerem estudos e pesquisas além do limite do trabalho.

Pode ser elaborada uma única seção para “Resultados e discussão” em caso de textos não muito extensos e a discussão não muito detalhada. Em caso de resultados complexos é pertinente elaborar um texto para apresentar os resultados e outro para discussão a partir do que foi constatado.

# 7 CONCLUSÃO

**Ou ‘Considerações finais’**

Elemento obrigatório. Consta de deduções lógicas baseadas e fundamentadas no texto. Devem ser empregadas frases curtas e sucintas, apresentando sugestões para trabalhos futuros.

REFERÊNCIAS

AMARE, M.Y.; ŠIMONOVÁ, S. Global challenges of students dropout: A prediction model development using machine learning algorithms on higher education datasets. *In*: SHS WEB OF CONFERENCES, 129., 2021, [online]. **Anais do** ... [online]: EDP Sciences-Web of Conferences, 2021.

AMIT, Y.; GEMAN, D. Shape quantization and recognition with randomized trees. **Neural computation**, Cambridge, v. 9, n. 7, p.1545-1588, 1997.

BELLI, J.I.R. Avaliação do ensino: questões metodológicas. *In*: CONGRESSO BRASILEIRO DO ENSINO DE ENGENHARIA-COBENGE, 34., 2006, Passo Fundo. **Anais do** ... Passo Fundo: COBENGE, 2006.

BREIMAN, L. Random forests.**Machine learning**, Standford, v. 45, p. 5-32, 2001.

CUTLER, A.; CUTLER, D.R.; STEVENS, J.R. Random forests. *In:* ZHANG, C.; MA, Y. (Ed.). **Ensemble machine learning**: methods and applications. Berlim: Springer Science & Business. p. 157-175, 2012.

DAITX, A.C.; LOGUERCIO, R. de Q.; STRACK, R. Evasão e retenção escolar no curso de Licenciatura em Química do Instituto de Química da UFRGS. **Investigações em Ensino de Ciências**, Porto Alegre, v. 21, n. 2, p. 153-178, 2016.

GOLDSCHMIDT, P.; WANG, J. When can schools affect dropout behavior? A longitudinal multilevel analysis. **American Educational Research Journal**, Washington, D.C., v. 36, n. 4, p. 715-738, 1999.

GONZALEZ, L. de A. **Regressão Logística e suas Aplicações**. 2018. 46p. Monografia (Graduação em Ciências da Computação) - Universidade Federal do Maranhão, São Luís, 2018.

GREBOVIC, M.; FILIPOVIC, L.; KATNIC, I.; VUKOTIC, M.; POPOVIC, T. Machine learning models for statistical analysis. **International Arab Journal of Information Technology,** Amman, v. 20, n. 3A, p. 505-514, 2023.

HOSMER, D.W.; LEMESHOW, S.; **Applied logistic regression.** New York: John Wiley, 1989, 307 p.

KANEKO, Hiromasa. Interpretation of Machine Learning Models for Data Sets with Many Features Using Feature Importance. **ACS omega**, Washington, D.C., v. 8, n. 25, p. 23218-23225, 2023.

LEK, S.; DELACOSTE, M.; BARAN, P.; DIMOPOULOS, I.; LAUGA, J.; AULAGNIER, S.

Application of neural networks to modelling nonlinear relationships in ecology. **Ecological**

**Modelling**, Amsterdam, v.90, n.1, p.39-52, 1996.

LIMA, F. S. de; ZAGO, N. Desafios conceituais e tendências da evasão no ensino superior: a realidade de uma universidade comunitária. **Revista Internacional de Educação Superior**, Brasília, v. 4, n. 2, p. 366-386, 2018.

LOCK. F. Challenges Faced by Students. *In:* LOCK. F. **Cliff's Nodes with Appreciations**. Nova York: AIP Publishing LLC. 2021.

LONG, Z.A.; NOOR, M.F.M. Factors influencing dropout students in higher education. **Education Research International**, Caíro, v. 2023, P. 1-13, 2023.

MATOS, D.A.; FARIA, I.; OLIVEIRA, L.P.J.; IRIART, M.F.S. (Orgs.). **Juventudes universitárias**: ingresso e permanência em perspectiva. Cruz das Almas, BA: EDUFRB, 2023. 289p.

MESQUITA, P.S.B. **Um Modelo de Regressão Logística para Avaliação dos Programas de Pós–Graduação no Brasil.** 2014. 107p. Dissertação (Mestrado em Engenharia da Produção) — Universidade Estadual do Norte Fluminense, Campo dos Goytacazes, 2014.

MINCER, J. Investment in human capital and personal income distribution. **Journal of Political Economy**, Illinois, v. 66, n. 4, p. 281-302, 1958.

MOREIRA, M.A. Desafios no ensino da física. **Revista Brasileira de Ensino de Física**, Florianópolis,v. 43, p. 1-8, 2021.

OQAIDI, K.; AOUHASSI, S.; MANSOURI, K. Towards a students’ dropout prediction model in higher education institutions using machine learning algorithms. **International Journal of Emerging Technologies in Learning,** Graz, v. 17, n. 18, p. 103-117, 202

PAL, S. Mining educational data to reduce dropout rates of engineering students. **International Journal of Information Engineering and Electronic Business**, [online], v. 4, n. 2, p. 1, 2012.

PALOMINO, J.C.V.; ORTEGA, A.M. Dropout Intentions in Higher Education: Systematic Literature Review. **Journal on Efficiency and Responsibility in Education and Science**, Bratislava, v. 16, n. 2, p. 149-158, 2023.

RIGO, S.J.; CAMBRUZZI, W.; BARBOSA, J.L.; CAZELLA, S.C. Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. **Revista Brasileira de Informática na Educação**, Porto Alegre, v. 22, n. 01, p. 132, 2014.

SANTOS, C.O.; PILATTI, L.A.; BONDARIK, R. Evasão e políticas públicas para o ensino superior: entre o conceito e a realidade. **Educere-Revista da Educação da UNIPAR**, Umuarama, v. 21, n. 1, 2021.

SAYED, A. H. Chapter 59: Logistic Regression. *In*: SAYED, A. H. **Inference and learning from data learning**. v. 3. Lausann:École Polytechnique Fédérale de Lausanne. 2022. p.2457-2498.

SILVA, C.A.G.; DÍAZ, J.P. Deserción en estudiantes de educación superior: un estudio de caso. **Universidad, Ciencia y Tecnología**, [online], v. 27, n. 119, p. 18-28, 2023.

SILVA, I.N. da; SPATTI, D.H.; FLAUZINO, R.A. R**edes Neurais Artificiais para engenharia e ciências aplicadas.** 2ed. – São Paulo: Artliber, 2016.

TALBERT, R. **Guia para utilização da aprendizagem invertida no ensino superior**. Porto Alegre: Penso Editora, 2019.

TINTO, V. Dropout from higher education. A theoretical synthesis of recent research. **Review**

**of Educational Research**, Washington, v. 45, p. 89-125, 1975.

YANG, B.S.; DI, X.; HAN, E.T. Random forests classifier for machine fault Diagnosis. **Journal of Mechanical Science and Techonology**, Daejeon, v. 22, p.1716–1725, 2008.

WANG, J.; MA, Y.; ZHANG, L.; GAO, R.X.; WU, D. Deep learning for smart manufacturing: Methods and applications. **Journal of manufacturing systems**, Amsterdam, v. 48, p. 144-156, 2018.

Listagem de referências das obras citadas no trabalho, padronizadas conforme NBR 6023 da Associação Brasileira de Normas Técnicas (2018).

APÊNDICE A – TÍTULO DO APÊNDICE

O apêndice é um texto ou documento elaborado pelo próprio autor, a fim de complementar sua argumentação, sem prejuízo da unidade nuclear de trabalho.

Elemento opcional. Deve ser precedido da palavra APÊNDICE, identificado com letras maiúsculas do alfabeto, travessão e respectivo título (como exemplificado acima). O título ‘Apêndice’ deve estar centralizado e com a mesma tipologia utilizada para as seções primárias, sugestão: maiúsculo e negrito.

**Apêndice** – **elemento pós-textual opcional**

APÊNDICE B – ORIENTAÇÕES SOBRE APRESENTAÇÃO DO TRABALHO

As regras gerais de apresentação de trabalhos acadêmicos de acordo com NBR 14724 da Associação Brasileira de Normas Técnicas (2011) estão descritas a seguir:

1. tamanho das margens: superior e esquerda: 3cm; margens inferior e direita: 2cm. Para o caso de trabalhos digitados dos dois lados da folha, as margens do verso devem ser: direita e superior de 3cm e esquerda e inferior de 2cm. Fonte: tamanho 12 (todo o texto, inclusive capa), exceção para citações com mais de três linhas, notas de rodapé, legendas, fontes de ilustrações e tabelas e ficha catalográfica, que devem ser em tamanho menor e uniforme;
2. espaço entre linhas: 1,5 cm em todo o texto:

- exceção para citações de mais de três linhas, notas de rodapé, referências, legendas das ilustrações e tabelas, natureza (tipo do trabalho, objetivo, nome da instituição a que é submetido e área de concentração) que devem se digitadas em espaço simples;

- as referências ao final do trabalho devem ser separadas entre si por um espaço simples em branco;

c) ilustrações: para qualquer tipo de ilustração, a indicação deve aparecer na parte superior precedida da designação: desenho, esquema, fluxograma, fotografia, gráfico, mapa, organograma, planta, quadro, figura, imagem, entre outros, seguida de seu número de ordem de ocorrência no texto, em algarismos arábicos, travessão e do respectivo título. Na parte inferior, indicar a fonte, mesmo que seja o próprio autor. A fonte deve ser informada no formato de citação, com a referência completa mencionada na listagem ao final do trabalho.

d) tabelas: apresentação de informações nas quais os dados numéricos se destacam (quantificação de determinado fato específico observado). Devem ser padronizadas confome as normas de apresentação tabular (IBGE, 1993).

Títulos das Seções

Os títulos das seções devem seguir o indicativo numérico, separado deste por um espaço em branco sem hifen, ponto final ou qualquer outro sinal, destacados tipograficamente (recursos gráficos de maiúscula, negrito, itálico ou sublinhado e outros), de forma hierárquica, da seção primária à quinária. Devem ser separados do texto que os precede ou os sucede por um espaço de 1,5 cm em branco.

Títulos sem indicativos numéricos

Errata, agradecimentos, lista de ilustrações, lista de abreviaturas e siglas, lista de símbolos, resumos, sumário, referências, apêndice, anexo e índice devem estar centralizados, com o mesmo destaque tipogr´fico ds seções primárias.

Elementos sem título e sem indicativo numérico

Não deve constar títulos para os elementos: folha de rosto, folha de aprovação, dedicatória e epígrafe.

Numeração Progressiva

Deve ser elaborada de acordo com a norma de numeração progressiva das seções dos documentos conforme NBR 6024 da Associação Brasileira de Normas Técnicas (2012a), conforme segue:

1. todas as seções devem conter um texto com início na linha seguinte à do título e não na frente deste. O texto deve iniciar a um espaço em branco de 1,5cm do título.
2. deve ser utilizado algarismo arábico na numeração e limitar-se à seção quinaria.

Alíneas

Quando for necessário enumerar assuntos de uma seção que não possua título, a mesma deve ser subdividida em alíneas, indicadas com letras minúsculas do alfabeto e ordenadas adequadamente, seguir como exemplo o Apêndice B deste documento.

1.2.1.1 Apresentação das alíneas

As letras indicativas das alíneas são reentradas em relação à margem esquerda do trabalho.

O texto da alínea deve ser iniciado com letras minúsculas e finalizado por ponto e vírgula, exceto a última que termina por ponto final. As linhas seguintes à primeira linha da alínea começam sob a primeira letra do texto da própria alínea. Havendo subalínea, a respectiva alínea termina em dois pontos (ver Apêndice B).

Subalíneas

São subdivisões das alíneas, apresentando um recuo em relação à alínea. O texto das mesmas deve começar por um hífen seguido de espaço. As linhas seguintes de uma subalínea iniciam-se sob a primeira letra do texto da própria subalínea, seguir como exemplo a seção 5 deste documento.

Paginação

Todas as folhas devem ser contadas sequencialmente a partir da folha de rosto, sendo que a numeração começa a aparecer na primeira folha da parte textual (introdução), em algarismos arábicos, no canto superior direito da folha com tamanho de fonte menor do que a utilizada para digitação do texto.

Inserção de número de páginas

Para que seja possível configurar corretamente a numeração de página em um trabalho acadêmico é preciso organizá-lo em seções. Para isso utiliza-se o recurso de ‘Quebra de Páginas e de Seção’*[[1]](#footnote-1)* do Microsoft Word.

Um trabalho acadêmico formatado conforme a NBR 14724 da Associação Brasileira de Normas Técnicas (2011) precisa ser organizado basicamente em 3 seções:

* 1. a ‘primeira Seção’ corresponde a capa;
  2. a ‘segunda Seção’ conterá os demais elementos pré-textuais do trabalho;
  3. a ‘terceira Seção’ corresponde ao restante do trabalho, (elementos textuais e pós-textuais).

Para organizar o trabalho desta forma, deve-se inserir duas ‘Quebras de Seção’ ao documento.

As ‘quebras de seção’ permitem controlar a numeração de páginas, informando ao Word quais páginas devem ser numeradas ou não, quais devem ser contabilizadas e qual o formato de numeração a ser utilizado.

Podem ser acrescentadas outras seções ao trabalho conforme a necessidade como, por exemplo, quando for preciso incluir anexos ou mesmo alterar a orientação de uma página.

Além das ‘Quebras de Seção*,’* existem também as ‘Quebras de Página’ que são utilizadas para separar o conteúdo dentro de uma seção. Geralmente utiliza-se este recurso para separar as seções primárias do trabalho, que começam sempre em uma nova página.

O botão “Mostrar tudo” ou “Mostrar/Ocultar” do Microsoft Office, permite exibir e ocultar certos marcadores de formatação utilizados pelo Word. Com este botão ativado é possível ver onde foram inseridas linhas em branco, quantos espaços foram inseridos entre as palavras e também onde estão as ‘Quebras de Página’ e ‘de Seção.’ Para ativá-lo basta clicar no ícone  no grupo ‘Parágrafo’ dentro da aba 'Página Inicial,’ para ocultar as marcações basta clicar no mesmo ícone novamente.

1.3.1.1 Configurando as quebras de seção

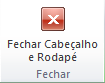
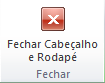
É necessário remover a numeração de página e as ‘Quebras de Seção’ caso elas já tenham sido inseridas anteriormente, para evitar conflitos na execução das configurações a seguir:

1. para inserir a seção que irá separar a capa dos elementos pré-textuais:
   * posicione o cursor na última linha da primeira página, que deve ser a página da capa;
   * clique na aba ‘Layout de Página’, depois em ‘Quebras’ e em seguida clique em ‘Próxima Página’;
2. para inserir a seção que separará os elementospré-textuais do restante do trabalho:
   * posicione o cursor na última linha da página do ‘Sumário;’
   * clique na aba ‘Layout de Página’, depois em ‘Quebras’ e em seguida clique em ‘Próxima Página’.

Ao inserir a segunda ‘Quebra de Seção*’* o trabalho estará divido nas três seções básicas tratadas anteriormente.

1.3.1.2 Configurando a numeração de página

Antes de inserir a numeração de página é preciso primeiro realizar algumas configurações:

1. clique duas vezes sobre o cabeçalho da segunda página (a página da folha de rosto) para editá-lo.
2. em seguida clique em . Isso desvinculará o cabeçalho das páginas da segunda seção da página da primeira seção (capa), permitindo ignorar esta última da contagem de páginas.
3. depois, na aba “Inserir” clique em “Número de página” > “Formatar Números de Página...”. Na janela que abrir marque “Iniciar em:” e coloque 1 na caixa ao lado. Depois clique em OK.
4. volte para a aba “Design” e clique em  para finalizar a edição do cabeçalho.
5. agora clique duas vezes sobre o cabeçalho da primeira página da terceira seção (provavelmente é a página da Introdução do seu trabalho).
6. clique em  para desvincular o cabeçalho e depois clique em  para encerrar a edição.
7. Para inserir a numeração, com o cursor na mesma página clique na aba “Inserir” depois vá em “Número de página” > “Início da página” e selecione a terceira opção: “Número sem formatação 3”.

A numeração foi inserida somente a partir da terceira seção, correspondente à parte textual do trabalho, considerando a quantidade de páginas da primeira seção correspondente aos elementos pré-textuais.

Sumário Automático

Neste template, automatizamos o sumário. Para que ele atenda corretamente à sua pesquisa, é preciso formatar os títulos das seções de seu trabalho com os devidos estilos de parágrafo. Para tanto, deixe o cursor na linha do título a ser formatado e, no conjunto “Estilo” da faixa de opções “Página Inicial,” escolha o estilo adequado:

para seções primárias, escolha Título 1;

para seções secundárias, escolha Título 2;

para seções terciárias, escolha Título 3;

para seções quaternárias, escolha Título 4;

para seções quinquenárias, escolha Título 5;

para as seções pós-textuais não numeradas, escolha Seção não numerada.

Depois de aplicados os estilos aos títulos do trabalho, basta clicar com o botão direito do mouse sobre a área do sumário, escolher Atualizar Campo, em seguida marcar Atualizar o índice inteiro e confirmar (botão OK).

Em caso de dúvida, consulte: [Inserir um sumário](https://support.office.com/pt-br/article/inserir-um-sum%C3%A1rio-882e8564-0edb-435e-84b5-1d8552ccf0c0) e [Atualizar um sumário](https://support.office.com/pt-br/article/atualizar-um-sum%C3%A1rio-6c727329-d8fd-44fe-83b7-fa7fe3d8ac7a).

**1.5 Links recomendados para auxiliar na elaboração de trabalhos acadêmicos**

[Mecanismo online para referências](http://www.more.ufsc.br/);

[Normalização - Citações;](https://www.bibliotecas.ufu.br/sites/bibliotecas.ufu.br/files/media/documento/citacoes_-_abr_2019_apresentacao_novo.pps)

[Normalização - Elaboração de trabalhos acadêmicos](https://prezi.com/m2nus57ydzeg/trabalhos-academicos-treinamento-virtual-normalizacao/);

[Normalização - Projeto de pesquisa;](https://prezi.com/shxghfrxhzvx/projeto-de-pesquisa-treinamento-virtual-normalizacao/)

[Normalização - Referências](https://www.bibliotecas.ufu.br/sites/bibliotecas.ufu.br/files/media/documento/referencias_6023_2018_apresentacao_novo.pps).

ANEXO A – TÍTULO DO ANEXO

O anexo constitui material não elaborado pelo autor, que serve de fundamentação, comprovação e ilustração do seu trabalho.

Elemento opcional. Deve ser precedido da palavra ANEXO, identificado com letras maiúsculas do alfabeto, travessão e respectivo título (como exemplificado acima). O título “**ANEXO**” deve estar centralizado e com a mesma tipologia utilizada para as seções primárias, sugestão: maiúsculo e negrito.

**Anexo – elemento pós-textual opcional**

ANEXO B – MODELO DE SUMÁRIO

**SUMÁRIO**

|  |  |  |
| --- | --- | --- |
| **1** | **INTRODUÇÃO** | **13** |
| **2** | **REFERENCIAL TEÓRICO** | **15** |
| **2.1** | **Prevalência** | **15** |
| **2.2** | **Classificação (ADA, 2011)** | **15** |
| **3** | **OBJETIVOS** | **23** |
| **3.1** | **Objetivo geral** | **23** |
| **3.2** | **Objetivos específicos** | **23** |
| **4** | **CASUÍSTICA E MÉTODOS** | **24** |
| **4.1** | **Considerações éticas** | **24** |
| **4.2** | **Casuística** | **25** |
| **4.3** | **Métodos** | **26** |
| ***4.3.1*** | ***Desenhos*** | **26** |
| ***4.3.2*** | ***Metodologia*** | **26** |
| *4.3.2.1* | *Variáveis estudadas* | 27 |
| *4.3.2.2* | *Definição das variáveis* | 27 |
| 4.3.2.2.1 | Equipamento utilizado | 28 |
| 4.3.2.2.2 | Técnica de exame | 28 |
| **4.4** | **Diagnóstico de Diabetes** | **29** |
| **4.5** | **Análise estatística** | **29** |
| **5** | **RESULTADOS** | **40** |
| **5.1** | **Caracterização clínica dos doentes** | **40** |
| **6** | **DISCUSSÃO** | **41** |
| **7** | **CONCLUSÃO** | **42** |
|  | **REFERÊNCIAS** | **53** |
|  | **APÊNDICE A – PROTOCOLO DE ESTADO – GRÁVIDAS NORMAIS E COM DMG** | **62** |
|  | **ANEXO A – PARECER DO CEP** | **63** |

**Sumário elaborado de forma não automática**

ANEXO C – MODELO DE LISTA DE REFERÊNCIAS

**REFERÊNCIAS**

AMARE, M.Y.; ŠIMONOVÁ, S. Global challenges of students dropout: A prediction model development using machine learning algorithms on higher education datasets. *In*: SHS WEB OF CONFERENCES, 129., 2021, [online]. **Anais do** ... [online]: EDP Sciences-Web of Conferences, 2021.

BELLI, J.I.R. Avaliação do ensino: questões metodológicas. *In*: CONGRESSO BRASILEIRO DO ENSINO DE ENGENHARIA-COBENGE, 34., 2006, Passo Fundo. **Anais do** ... Passo Fundo: COBENGE, 2006.

BREIMAN, L. Random forests.**Machine learning**, Standford, v. 45, p. 5-32, 2001.

DAITX, A.C.; LOGUERCIO, R. de Q.; STRACK, R. Evasão e retenção escolar no curso de Licenciatura em Química do Instituto de Química da UFRGS. **Investigações em Ensino de Ciências**, Porto Alegre, v. 21, n. 2, p. 153-178, 2016.

GOLDSCHMIDT, P.; WANG, J. When can schools affect dropout behavior? A longitudinal multilevel analysis. **American Educational Research Journal**, Washington, D.C., v. 36, n. 4, p. 715-738, 1999.

GONZALEZ, L. de A. **Regressão Logística e suas Aplicações**. 2018. 46p. Monografia (Graduação em Ciências da Computação) - Universidade Federal do Maranhão, São Luís, 2018.

GREBOVIC, M.; FILIPOVIC, L.; KATNIC, I.; VUKOTIC, M.; POPOVIC, T. Machine learning models for statistical analysis. **International Arab Journal of Information Technology,** Amman, v. 20, n. 3A, p. 505-514, 2023.

LIMA, F. S. de; ZAGO, N. Desafios conceituais e tendências da evasão no ensino superior: a realidade de uma universidade comunitária. **Revista Internacional de Educação Superior**, Brasília, v. 4, n. 2, p. 366-386, 2018.

LOCK. F. Challenges Faced by Students. *In:* LOCK. F. **Cliff's Nodes with Appreciations**. Nova York: AIP Publishing LLC. 2021.

LONG, Z.A.; NOOR, M.F.M. Factors influencing dropout students in higher education. **Education Research International**, Caíro, v. 2023, P. 1-13, 2023.

MATOS, D.A.; FARIA, I.; OLIVEIRA, L.P.J.; IRIART, M.F.S. (Orgs.). **Juventudes universitárias**: ingresso e permanência em perspectiva. Cruz das Almas, BA: EDUFRB, 2023. 289p.

MESQUITA, P.S.B. **Um Modelo de Regressão Logística para Avaliação dos Programas de Pós–Graduação no Brasil.** 2014. 107p. Dissertação (Mestrado em Engenharia da Produção) — Universidade Estadual do Norte Fluminense, Campo dos Goytacazes, 2014.

MINCER, J. Investment in human capital and personal income distribution. **Journal of Political Economy**, Illinois, v. 66, n. 4, p. 281-302, 1958.

MOREIRA, M.A. Desafios no ensino da física. **Revista Brasileira de Ensino de Física**, Florianópolis,v. 43, p. 1-8, 2021.

OQAIDI, K.; AOUHASSI, S.; MANSOURI, K. Towards a students’ dropout prediction model in higher education institutions using machine learning algorithms. **International Journal of Emerging Technologies in Learning,** Graz, v. 17, n. 18, p. 103-117, 202

PAL, S. Mining educational data to reduce dropout rates of engineering students. **International Journal of Information Engineering and Electronic Business**, [online], v. 4, n. 2, p. 1, 2012.

PALOMINO, J.C.V.; ORTEGA, A.M. Dropout Intentions in Higher Education: Systematic Literature Review. **Journal on Efficiency and Responsibility in Education and Science**, Bratislava, v. 16, n. 2, p. 149-158, 2023.

SANTOS, C.O.; PILATTI, L.A.; BONDARIK, R. Evasão e políticas públicas para o ensino superior: entre o conceito e a realidade. **Educere-Revista da Educação da UNIPAR**, Umuarama, v. 21, n. 1, 2021.

SAYED, A. H. Chapter 59: Logistic Regression. *In*: SAYED, A. H. **Inference and learning from data learning**. v. 3. Lausann:École Polytechnique Fédérale de Lausanne. 2022. p.2457-2498.

SILVA, C.A.G.; DÍAZ, J.P. Deserción en estudiantes de educación superior: un estudio de caso. **Universidad, Ciencia y Tecnología**, [online], v. 27, n. 119, p. 18-28, 2023.

TALBERT, R. **Guia para utilização da aprendizagem invertida no ensino superior**. Porto Alegre: Penso Editora, 2019.

TINTO, V. Dropout from higher education. A theoretical synthesis of recent research. **Review**

**of Educational Research**, Washington, v. 45, p. 89-125, 1975.

YANG, B.S.; DI, X.; HAN, E.T. Random forests classifier for machine fault Diagnosis. **Journal of Mechanical Science and Techonology**, Daejeon, v. 22, p.1716–1725, 2008.

1. Neste documento utiliza-se terminologia Seção com sentidos distintos. O termo Seção empregado pela Microsoft Word é um recurso do MSWord utilizado para agrupar as páginas de um documento em conjuntos, que permite aplicação de formatações específicas para cada conjunto. De acordo com a NBR 6024 da Associação Brasileira de Normas Técnicas (2012, p. 1) seção é a “parte em que se divide um texto de um documento que contém as matérias consideradas afins na exposição ordenada do assunto.” [↑](#footnote-ref-1)