



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE EDUCAÇÃO E SAÚDE
UNIDADE ACADÊMICA DE FÍSICA E MATEMÁTICA
CURSO DE LICENCIATURA EM MATEMÁTICA**

DAMIANA NATALIA ALVES DA SILVA

**MÉTODOS DE CLASSIFICAÇÃO POR TEORIA DA DECISÃO PARA
MENSURAÇÃO DE DADOS DO ENEM**

**CUITÉ - PB
2021**

DAMIANA NATALIA ALVES DA SILVA

**MÉTODOS DE CLASSIFICAÇÃO POR TEORIA DA DECISÃO PARA
MENSURAÇÃO DE DADOS DO ENEM**

Monografia apresentada à Banca Examinadora, como exigência parcial à conclusão do Curso de Licenciatura em Matemática, da Universidade Federal de Campina Grande - campus Cuité.

Orientador: Jorge Alves de Sousa

**CUITÉ – PB
2021**

S586m Silva, Damiana Natalia Alves da.

Métodos de classificação por teoria da decisão para mensuração de dados do ENEM. / Damiana Natalia Alves da Silva. - Cuité, 2021.
34 f.: il. color.

Trabalho de Conclusão de Curso (Licenciatura em Matemática) - Universidade Federal de Campina Grande, Centro de Educação e Saúde, 2021.

"Orientação: Profa. Jorge Alves de Sousa".

Referências.

1. Matemática. 2. Algoritmo de classificação. 3. Teoria de decisão. 4. Árvore de decisão. I. Sousa, Jorge Alves de. II. Título.

CDU 51(043)

DAMIANA NATALIA ALVES DA SILVA

Monografia apresentada à Banca Examinadora, como exigência parcial à conclusão do Curso de Licenciatura em Matemática, da Universidade Federal de Campina Grande Campus Cuité.

Aprovada em: 06/10/2021

BANCA EXAMINADORA



Prof. Dr. Jorge Alves de Sousa - UFCG

(Orientador)

gov.br Assinado de forma digital
por ALEXANDRO ALVES
VIEIRA:02940732485

Prof. Dr, Alexandro Alves Vieira -
UFCG

(Membro da Banca)



Prof. Dr. Ramilton Marinho Costa -

UFCG(Membro da Banca)

**CUITÉ - PB
2021**

AGRADECIMENTOS

Agradeço à Deus, primeiramente pelo dom da vida, por me permitir chegar aonde cheguei, por ser meu sustento ao longo da caminhada, por me dá saúde, coragem, força e muita sabedoria para continuar em busca dos meus sonhos. Gratidão por ser meu alicerce em todos os dias da minha vida e por me tornar uma grande vencedora. Toda honra e toda glória sejam dadas á Ele!

Agradeço a minha família pelo apoio que sempre me deram durante toda a minha vida. Por sempre terem acreditado em mim e pela dedicação que sempre tiveram durante essa trajetória, sem vocês esse sonho não seria possível. Em especial aos meus pais, Marilene Vicente da Silva Alves (in memoria) e José Alves Procópio, meus maiores exemplos de vida e amor, sou grata por me ensinarem sempre a importância de ter coragem e dedicação ao que se faz e persistir sempre em busca dos objetivos, mesmo em situações difíceis. Aos meus irmãos Marcos José, Nayara e Natan, aos meus primos Renato e Robervânia, agradeço por sempre me apoiar, me dar forças e sempre vibrar comigo todas as vitórias ao logo dessa caminhada.

Agradeço ao meu namorado, Mateus Araújo, por sempre ter acreditado em mim, por me encorajar, por sua paciência, compreensão, incentivo, por seu carinho e todo seu amor. Obrigada por segurar minha mão e me ajudar em absolutamente tudo.

Meu eterno agradecimento ao meu orientador, Jorge Alves de Sousa, pelos os ensinamentos, por ter acreditado em mim e por agregar tanto conhecimento em minha vida. Agradeço por toda cobrança e exigência, realizadas de maneiras sutis. Em seu nome, aproveito para agradecer a todos os mestres que já passaram ao longo dessa trajetória.

Agradeço aos meus grandes amigos que construí durante essa longa caminhada, em especial a Vinicius, Erica, Janaína, Ailton, Ivo, Fernando, Anailde, Maria, Isaac, Isabel, Alane, Carina, Caliane, Felipe, Vilânia e Eduarda. por cada conselho, por me ensinarem o valor da amizade e por me ajudarem em tudo quando eu mais precisei. Vocês estarão sempre presentes no meu coração. E a todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

Não se conhece completamente
uma ciência, a menos que se
saiba a sua história.”

(Auguste Comte)

RESUMO

Neste trabalho, se utilizou o algoritmo da árvore de decisão para analisar a relação da nota da redação com variáveis socioeconômicas de candidatos que prestaram o ENEM 2019 e declararam residência no município de Cuité-PB. Nosso objetivo foi gerar uma árvore de decisão usando o algoritmo de classificação, com base na qual poderíamos encontrar fatores que são importantes para o desempenho dos candidatos na redação. Atingimos especificamente esse objetivo construindo um modelo de decisão, cujas etapas incluíram a preparação de dados, pré-processamento de dados (limpeza de dados, conversão), construção de modelo (treinamento de algoritmo) e otimização de algoritmo para simulação do modelo. Neste cenário, dentre as 11 variáveis utilizadas no estudo, apenas três variáveis foram consideradas significativas na plotagem da árvore de decisão, que foram: estado civil, idade e computador em casa com uma precisão de acerto de 75%. Diante do exposto, apesar do modelo de previsão obtido neste estudo, se mostra adequado, entendemos que outros modelos devem ser validados em trabalhos futuros buscando melhorar esta precisão inserindo-se novas variáveis a análises.

Palavras-chave: Árvore de decisão. Algoritmo de classificação. Otimização

ABSTRACT

In this work, the decision tree algorithm was used to analyze the relation of the essay grade with socioeconomic variables of candidates who took the ENEM 2019 and declared residence in the city of Cuité - PB. Our objective was to generate a decision tree using the ranking algorithm, based on which we could find factors that are important for the candidates' performance in the essay. We specifically achieved this goal by building a decision model, whose steps included data collection and preparation, data pre-processing (data cleaning, conversion), model building (algorithm training) and algorithm optimization for model simulation. In this scenario, only three variables were considered significant in the decision tree plot, which were marital status, age and home computer with a 75% accuracy. Given the above, it is believed that other models should be validated in future work seeking to improve this accuracy by inserting new variables into analysis.

Keywords: Decision tree. Classification algorithm. Optimization.

LISTA DE FIGURAS

Figura 1: Exemplo de construção de uma árvore	19
Figura 2: Exemplo de construção de uma árvore decisão.....	20
Figura 3: árvore de decisão induzida a partir dos exemplos das figuras 3 e 4	20
Figura 4: Distribuição de percentuais das variáveis categóricas utilizadas na pesquisa, referente aos candidatos que prestaram ENEM 2019 e declararam residência no município de Cuité-PB.....	29
Figura 5: Índice de importância da variável para o modelo.....	30
Figura 6: Árvore de decisão para matriz de dados dos candidatos do ENEM/2019 com residência em Cuité/PB.....	30

Lista de Siglas

ENEM - Exame Nacional do Ensino Médio

LDB - Lei de Diretrizes e Bases da Educação

MEC - Ministério da Educação e Cultura

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais

PCNs - Parâmetros Curriculares Nacionais

PDA - Plano de Dados Abertos

MMFDH - Ministério da Mulher, da Família e dos Direitos Humanos

LAI - Lei de Acesso à Informação

LISTA DE QUADROS

Quadro 1: Nome e descrição das variáveis utilizadas no processo de análise.....	27
Quadro 2: Distribuição de medidas de posição e dispersão das variáveis quantitativas utilizadas na análise dos candidatos que prestaram ENEM 2019 e declararam residência no município de Cuité-PB.....	29
Quadro 3: Matriz de confusão para os dados de teste dos candidatos do ENEM/ 2019 com residência em Cuité/ PB.....	31

SUMÁRIO

INTRODUÇÃO.....	13
DEFINIÇÕES DE PROBABILIDADE	15
PROBABILIDADE CONDICIONAL E TEOREMA DE BAYES	16
PROBABILIDADE CONDICIONAL	17
TEOREMA DA DECISÃO.....	18
DIAGRAMA DE ÁRVORE DE DECISÃO	19
HISTÓRICO DO ENEM	23
DADOS ABERTOS.....	26
MATERIAIS E MÉTODOS	27
RESULTADOS	28
ANÁLISE DESCRITIVA	28
ÁRVORE DE DECISÃO.....	30
CONCLUSÃO	32
REFERÊNCIAS	33

INTRODUÇÃO

A probabilidade, pode ser apresentada de forma constante em nosso dia a dia, bem como ainda, os princípios probabilísticos aos quais vem se tornando ferramenta de trabalho para diversas áreas do conhecimento, tendo em vista que inúmeros fenômenos são controvertidos como experimentos aleatórios.

Além disso, a probabilidade pode ser definida ainda como sendo bem mais que uma vertente matemática, tendo em vista que esta ferramenta é utilizada em diversos modelos de Pesquisa Operacional, no qual, suas considerações e aplicabilidades são doutrinados de forma constante no ensino médio e superior objetivando assim, tornar mais compreensível um procedimento tão importante.

Contudo, a abordagem seguida controverte sobre a probabilidade como exemplo da informação disposta para quantificar incerteza, sendo, deste modo, relativa ao sujeito no qual possa ter posse desta informação, assim, dá-se o nome de interpretação subjetiva, bem como ainda, de sua utilização no auxílio de tomadas de decisão.

Em se tratando do Exame Nacional do Ensino Médio (ENEM), este surgiu no ano de 1998, ao qual foi aplicada sua primeira edição, este exame fazia parte de uma política de reestruturação do Ensino Médio, o qual passou a existir com a LDB, Lei 9.394.

Além disso, pode-se acrescentar ainda que o ENEM, no decorrer dos anos, desde sua implementação, passou por modificações, como por exemplo, aplicação da prova em dois domingos sucessivos, apesar disso, passou a ser conservada a sua estrutura referente ao número de questões por área e a aplicação da redação.

Assim, a presente pesquisa apresenta a seguinte problemática: quais os métodos de classificação são mais adequados para análise da relação da nota dos estudantes, residentes na cidade de Cuité-PB, na redação do ENEM, com algumas variáveis socioeconômicas.

Deste modo, pode-se dizer que, este estudo tem como objetivo geral apresentar os métodos de classificação por teoria da decisão e probabilidade subjetiva para encontrar alguns fatores que são importante e influenciam no desempenho dos candidatos, residentes na cidade de Cuité-PB, na redação do

ENEM/2019, bem como estimular a inserção da teoria da decisão ao currículo do ensino médio.

HISTÓRIA DA PROBABILIDADE

Segundo Borges (2011) na antiguidade, confiava-se bastante na superstição e atribuía-se aos deuses a vitória ou a derrota em meio a um jogo ou fato. Contudo, o cálculo de probabilidades poderia, quem sabe, ter sido criado bem antes, no entanto, especialistas, em particular os gregos, tendo em vista que por crerem que o futuro se encontrava nas mãos dos deuses e persistirem na veracidade do fato irrestrito demonstrado por meio da lógica, apresentavam dificuldade em acolher a aleatoriedade.

Entretanto, Dantas (2004) assinala que, em se tratando das primeiras civilizações a fazerem uso do processo de aleatoriedade pode-se destacar os romanos, aos quais passaram a iniciar sua criação. Com isso, pode-se acrescentar ainda que os romanos tinham mais preocupação com a medição e a contagem do que com a geometria.

Conforme James (2006) apenas no século XVI passou a criar a elaboração do cálculo de probabilidades. Contudo, tem-se como sendo ponderado por muitos como sendo o Girolamo Cardano (1501-1576) o iniciador do estudo matemático das probabilidades, tendo em vista que este passou a ser o primeiro a introduzir técnicas de combinatória para que se fosse assim possível calcular a quantidade de possibilidades aos quais pudessem ser favoráveis em um evento aleatório e, de tal modo, calcular a probabilidade do acontecimento de um evento tendo em vista a razão entre o número de possibilidades adeptas e o número total de possibilidades agregadas a experimentação.

Pode-se também destacar outros italianos que também se dedicaram ao estudo da aleatoriedade, como são os casos de Galileu Galilei (1564- 1642) e Tartaglia (1499-1557). Em se tratando de Galileu, pode-se dizer que o mesmo era físico, astrônomo, matemático e filósofo, este conjecturou que um dado apresenta a probabilidade igual de cair em qualquer um dos seis lados (MOREIRA 2015).

Nas palavras de Moreira (2015) o mesmo assinala que Jacob Bernoulli (1654-1705) foi outro matemático que também passou seus dias a se dedicar

aos estudos das probabilidades, o mesmo confiava que, para que se fosse possível tomar decisões racionais, seria necessário um procedimento matemático de confiança para produzir probabilidades.

Contudo, hoje em dia, o estudo pertinente ao cálculo de probabilidades se apresenta como sendo de enorme acuidade, além deste, haver sua aplicação em inúmeras áreas, como são os casos da “estatística, economia, engenharia, física, química, sociologia, psicologia, biologia, entre outros ramos do conhecimento” (MOREIRA 2015, p. 12).

DEFINIÇÕES DE PROBABILIDADE

Para Magalhães e Lima (2008), ao realizarmos um experimento sempre teremos um grau de incerteza associado a seus resultados, mesmo que este já tenha sido realizado várias vezes. A probabilidade com a qual podemos esperar que um evento ocorra, é representada como um número real entre 0 e 1. Ou seja, a chance de ocorrência de um evento está sempre entre a impossibilidade e a certeza. Assim Passamos então a definição formal de probabilidade. Seja S espaço amostral. Uma função P definida para todos os subconjuntos de S é chamada de probabilidade se:

1. $0 \leq P(A) \leq 1$, para todo evento $A \subset S$, isto é, a probabilidade de qualquer evento está entre 0 e 1;
2. $P(S) = 1$, onde S é o espaço amostral;
3. $P(\emptyset) = 0$, isto é, o evento impossível tem probabilidade zero;
4. Se A_1 e A_2 são disjuntos (também chamados de mutuamente exclusivos) temos que: $[P(A_1 \cup A_2) = P(A_1) + P(A_2)]$.

Em se tratando das definições de probabilidade, pode-se dizer segundo Montgomery e Runger (2012), que a probabilidade se apresenta como sendo de grande relevância e importância em nosso cotidiano, visto que, continuamente, nos deparamos com ocasiões nos quais nos forçamos a tomar decisões, bem como, das quais não temos a certeza, exclusivamente, temos indicativos e indícios dos quais nos possibilitam determinar alguma probabilidade de acertarmos.

Consta-se assinalar ainda, conforme entendimento de Montgomery e Runger (2012) que se pode definir a probabilidade como sendo bem mais que uma vertente matemática, mas também, um instrumento de suma importância ao estudo de diversos modelos de Pesquisa Operacional, em que, seus conceitos e aplicabilidades são instruídos cada vez mais no ensino médio e superior objetivando assim, tornar mais compreensível um procedimento tão respeitável, bem como também tão utilizado no nosso dia a dia.

Magalhães e Lima (2008) afirma que, o uso do conceito probabilístico adequa uma visão mais ampla e compreensiva, bem como, futura de um apontado evento, sem esquecer que esta não é uma ciência exata, contudo contribui em diversos estudos como embasamento de um ensaio futuro.

Por fim, é perceptível compreender que em algum determinado tempo, fazemos previsões, e com isso, temos a compreensão de que podemos conferir um número no qual quantifica a oportunidade de algo vir a ocorrer. Quando relacionada à matemática, pode-se dizer que o cálculo das probabilidades é um ramo no qual examina os fenômenos aleatórios, assim, observações ou experimentações nos quais, a partir do momento em que são concretizados, não proporcionam resultados distinguidos previamente (MAGALHÃES; LIMA, 2008).

PROBABILIDADE CONDICIONAL E TEOREMA DE BAYES

De acordo com Moreira (2015) o Teorema de Bayes constitui-se como sendo uma fórmula matemática no qual tem seu principal uso para o cálculo da probabilidade de um evento tendo em vista que outro evento já sucedeu, conhecida como probabilidade condicional.

Contudo, Moreira (2015) assinala que o principal ponto em relação ao Teorema de Bayes é a necessidade de se ter alguma informação anterior, assim, é necessário ter conhecimento de que um apurado evento já sucedeu, bem como ainda, qual a probabilidade desse evento.

Deste modo, Moreira (2015) alude que, tendo em vista essa dedução bayesiana no qual se origina o termo “grau de crença”, no qual é essa certeza em determinado evento anterior, tal proposição primitiva. Assim sendo, para que se possa ter uma maior compreensão acerca do cálculo da probabilidade de um

evento A dado que um evento B ocorreu, " $P(A|B)$ ", pelo Teorema de Bayes temos que:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Ou seja, precisamos de alguns dados, que são:

- $P(A|B)$: Probabilidade do evento A acontecer;
- $P(B|A)$ Probabilidade de B acontecer, dado que A já ocorreu;
- $P(A)$: probabilidade de A ocorrer;
- $P(B)$: probabilidade de B ocorrer.

PROBABILIDADE CONDICIONAL

Smole e Diniz (2010) diz que a probabilidade condicional ou probabilidade condicionada pode ser definida como sendo um conceito da matemática no qual abrange dois eventos (A e B) em um espaço amostral (S) finito e não vazio.

Em se tratando de Espaço Amostral e Eventos Smole e Diniz (2010) ainda assinala que o "espaço amostral" configura-se como sendo o conjunto de prováveis resultados conseguidos a partir de um evento ou fenômeno casual. Contudo, em relação aos subconjuntos de um espaço amostral, estes são conhecidos como "eventos".

Para Martins (2017) considere-se um espaço de resultados S e uma probabilidade p nesse espaço. Dados dois acontecimentos A e B , com $P(B) > 0$, define-se probabilidade condicional de A se B (ou dado B , ou sabendo B , ou condicional à ocorrência de B), e representa-se por $P(A | B)$, como sendo:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Mais concretamente, ao falarmos na probabilidade condicional, nomeadamente probabilidade do acontecimento A dado B , estamos a restringir

o espaço de resultados em que estamos a trabalhar ao definido pelo acontecimento B. Assim, a probabilidade de A em B é calculada dividindo a probabilidade da interseção de A com B, pela probabilidade de B.

Conforme entendimento de Moreira (2015) a probabilidade condicional faz menção à probabilidade de que um evento possa incidir com embasamento em um evento precedente. Contudo, incontestavelmente, esses dois eventos necessitam ser conjuntos não vazios pertencentes a um espaço amostral finito.

Já conforme explana Moreira (2015) a probabilidade condicional constitui-se como sendo a probabilidade de ocorrência de um segundo evento de um espaço amostral no qual sobrevém em um evento posteriormente quando já possa ter sucedido o primeiro.

TEOREMA DA DECISÃO

Segundo Moreira (2015) pode-se definir a árvore de decisão como um rabisco coerente e direcionado, sendo esta composta por ramos de ações e ramos de acontecimentos. Em se tratando dos ramos de ações, pode-se afirmar que os mesmos simulam decisões e procedem de nós quadrados, estabelecem alternativas; já no que se refere aos ramos de acontecimentos, estes procedem de nós circulares, são nós de improbabilidade, de dúvidas e concebem eventos aleatórios.

Moreira (2015) diz que a cada um destes ramos de acontecimentos encontra-se consecutivamente unificada uma possibilidade objetiva ou subjetiva e, no arremate de cada ramo terminal, deparam-se os valores ou conveniências das concernentes implicações. Contudo, na edificação da árvore, as probabilidades recomendadas são dependentes aos estados de ciência cômodas a cada nó, e assim, consta-se ainda complementar que em cada ramificação as probabilidades necessitam ter soma unitária.

Moreira (2015) alude que tal instrumento proporciona diversas vantagens, podendo-se destacar entre elas, a de revelar-se clara e explicitamente as decisões nos quais precisarão ser tomadas, bem como ainda, a ordem em que cada decisão eventualmente necessitará ser tomada, além ainda, das decorrências e resultados de cada decisão e por fim, as fontes de insegurança nos quais comprometem cada decisão.

Entretanto, de acordo com Moreira (2015) pode-se dizer que em situações mais simples, quase todos os dias, senão em todos, tomamos decisões em nosso cotidiano, e em grande parte, não percebemos que nos encontramos fazendo escolhas, por envolver circunstâncias triviais. Contudo, pode-se citar como exemplo, se vamos a um aniversário, se vamos comprar ou não um presente, o que comprar, qual roupa vestir, entre outros, tudo requer uma decisão.

Deste modo, pode-se concluir que a Análise da Decisão nos permite ter a solução de problemas mais complicados e difíceis. Além disso, essa ferramenta possibilita reunir um conjunto de conceitos e técnicas quantitativas nos quais tornam mais fácil o tratamento lógico de circunstâncias que possam envolver incerteza, possibilitando que deste modo, possamos tomar decisões.

DIAGRAMA DE ÁRVORE DE DECISÃO

Segundo entendimento de Gollapudi (2016) quando se trata da finalidade de qualquer árvore de decisão, pode-se dizer que é elaborar um padrão viável no qual antecipe e pressagie o valor de uma variável de destino, tendo como principal embasamento no conjunto de variáveis de entrada. Com isso, é importante lembrar que as árvores de decisão podem ser sobrepostas a problemas de regressão e de classificação.

Segundo Lauretto (2020, p. 2), uma árvore de decisão é uma coleção de elementos chamados nós, dentre os quais um é distinguido como uma raiz, juntamente com uma relação de “paternidade” que impõe uma estrutura hierárquica sobre os nós. Formalmente, uma árvore pode ser recursivamente da seguinte maneira:

1. Um único nó é uma árvore. Este nó é também a raiz da árvore.
2. Suponha que t seja um nó e T_1, T_2, \dots, T_k sejam árvores com raízes t_1, t_2, \dots, t_k respectivamente. Podendo construir uma nova árvore transformando t em pai dos nós t_1, t_2, \dots, t_k . Nessa árvore, t será a raiz e T_1, T_2, \dots, T_k serão as sub-árvores ou ramos da raiz. Os nós t_1, t_2, \dots, t_k são chamados filhos do nó t . se t_1, t_2, \dots, t_k é uma sequência de nós em uma árvore tais que t_{i+1} para $1 \leq i < k$ então esta sequência é denominada um caminho do nó t_1 até o nó t_k .

Se t_1, t_2, \dots, t_k é uma sequência de nós em uma árvore tais que t_i é o pai e t_{i+1} para $1 \leq i < k$, então esta sequência é denominada um caminho do nó t_1 até o nó t_k .

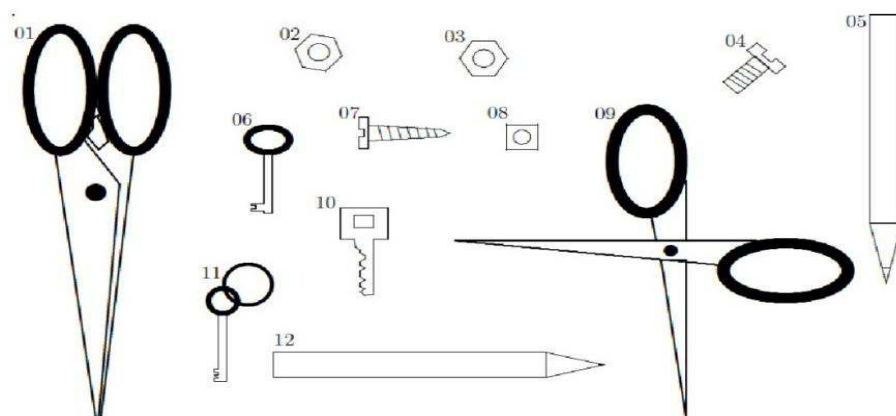
Se existe um caminho do nó a ao nó b , então a é um ancestral de b e b é um descendente de a . Todo nó é ancestral e descendente de si mesmo.

Dado um nó $t \in T$ a sub-árvore ou ramo T_t de T consistirá do nó t (que será a raiz de T_t) juntamente com todos os descendentes de t em T . Todos os nós que não possuem filhos são chamados nós terminais ou folhas. Os nós que contêm filhos são chamados nós não-terminais ou nós internos. O conjunto das folhas de T será denotado por \bar{T} .

Para Marcelo S. Lauretto (2010), uma Árvore de Decisão seria um nó folha (ou nó resposta) que iria conter o nome de uma classe ou um símbolo nulo, ou um nó interno (ou nó de decisão) que conteria o nome de um atributo, para cada possível valor desse atributo, iria corresponder um ramo para outra árvore de decisão. Essa árvore de decisão iria possuir a seguinte estrutura: apresentar nós internos rotulados com atributos; folhas rotuladas com classes, ramos rotulados com valores (atributos categóricos) ou intervalos (atributos numéricos).

Para entender a construção da árvore decisão observe os objetos do exemplo abaixo. Os objetos foram enumerados de 01 a 12, e pertencem a cinco classes, que são elas: porca, parafuso, chave, caneta, tesoura. Agora suponha que esses objetos sejam mostrados a um sistema de visão. E que suas silhuetas são capturadas por uma câmera e processadas por um programa de processamento de imagens.

Figura 1: Exemplo de construção de uma árvore



(a)

Fonte: Lauretto (2010)

Figura 2: Exemplo de construção de uma árvore

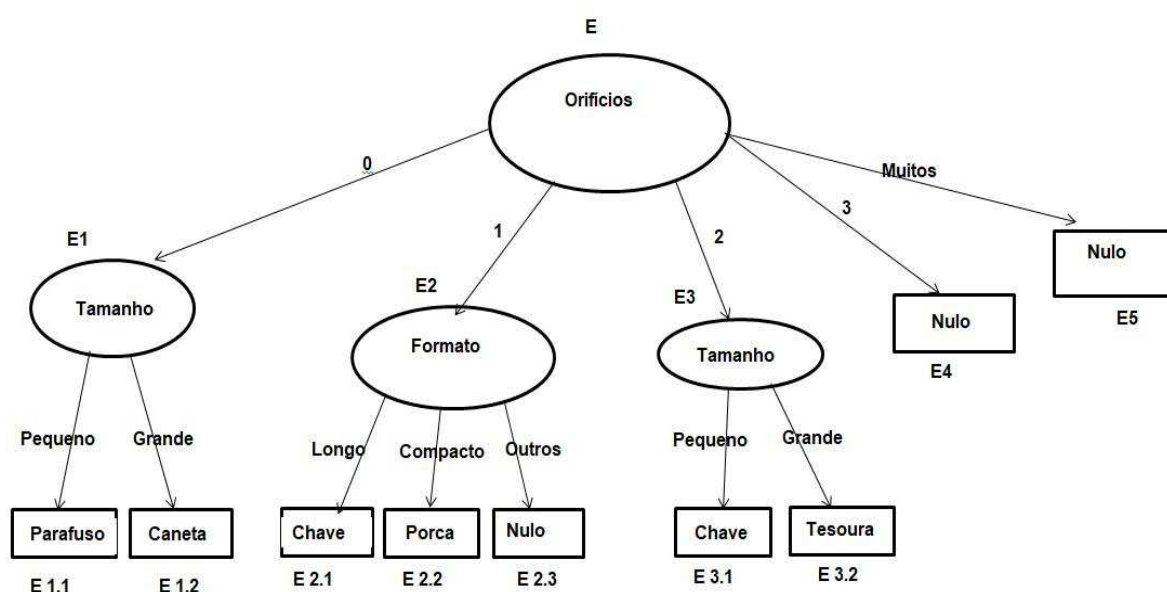
Objeto No.	TAMANHO	FORMATO	ORIFICIO	CLASSE
01	Grande	Longo	2	Tesoura
02	Pequeno	Compacto	1	Porca
03	Pequeno	Compacto	1	Porca
04	Pequeno	Longo	0	Parafuso
05	Grande	Longo	0	Caneta
06	Pequeno	Longo	1	Chave
07	Pequeno	Longo	0	Parafuso
08	Pequeno	Compacto	1	Porca
09	Grande	Longo	2	Tesoura
10	Pequeno	Longo	1	Chave
11	Pequeno	Longo	2	Chave
12	Grande	Longo	0	Caneta

(b)

Fonte: Laretto (2010)

Observe que o programa pode extrair alguns valores de atributos de cada imagem, neste caso, o tamanho, o formato e o número de orifícios de um objeto. Também é possível saber os valores desses atributos como, o tamanho, o Formato, e o Orifícios. Assim, cada objeto pode ser representado através de seu vetor de atributo (item b da figura).

A figura abaixo é uma representação de uma árvore de decisão induzida a partir dos exemplos descrito acima.

Figura 3: árvore de decisão induzida a partir dos exemplos das figuras 1 e 2.

Fonte: Laretto (2010)

Lauretto (2010), descreve as etapas para a construção da árvore com apenas um nó (a raiz) e associando a esse nó o conjunto E . Ele ainda fala que como os exemplos não estão na mesma classe, decidi expandir a árvore. Para isso, escolher um atributo (na ilustração, Orifícios), com o qual consideramos a raiz. Dividimos o conjunto E em subconjuntos, agrupando exemplos de mesmo valor de Orifícios em cada subconjunto. Para cada um desses subconjuntos, criamos um nó-filho do nó Orifícios. Neste caso, os subconjuntos de E seriam:

$$\begin{aligned} E_1 &= \{x \in E | X_{\text{Orifícios}} = 0\} \\ E_2 &= \{x \in E | X_{\text{Orifícios}} = 1\} \\ E_3 &= \{x \in E | X_{\text{Orifícios}} = 2\} \\ E_4 &= \{x \in E | X_{\text{Orifícios}} = 3\} \\ E_5 &= \{x \in E | X_{\text{Orifícios}} = \text{muitos}\} \end{aligned}$$

Agora será analisado o nó referente ao subconjunto E_1 . Uma vez que E_1 contém exemplos de classes distintas, além disso, o nó deve ser dividido e rotulado, passando a ser chamado de nó com um novo atributo, no caso, tamanho, e assim criar novos filhos para esse nó e dividimos E_1 conforme o Tamanho (pequeno; grande).

$$\begin{aligned} E_{1,1} &= \{x \in E_1 | X_{\text{Tamanho}} = \text{Pequeno}\} \\ E_{1,2} &= \{x \in E_1 | X_{\text{Tamanho}} = \text{Grande}\} \end{aligned}$$

Observe que todos os exemplos em $E_{1,1}$ pertencem à mesma classe, neste caso apenas rotulamos o nó correspondente com a classe à qual aqueles exemplos pertencem (parafuso) e declaramos esse nó como uma folha, passando a examinar outros nós que ainda não tenham sido devidamente tratados. O procedimento é análogo para os demais subconjuntos.

Para Lauretto (2010), a ideia que norteia o algoritmo de construção é constituído da seguinte maneira: expansão da árvore através de sucessivas partições do conjunto de treinamento até que a condição de parada seja satisfeita; Eliminação de algumas partes inferiores da árvore, através de reagrupamentos dos subconjuntos da partição Algoritmo.

Os dados necessários para o algoritmo seria um conjunto de aprendizado E , uma condição de parada $P(E)$, uma função de avaliação score (E ; atributo), uma função de classificação classe (T), uma função de categorização cat_{eg} (E ; atributo). Se o conjunto E satisfaz a condição de parada $P(E)$, então a árvore resultante é um único nó t rotulado conforme a regra classe (t).

HISTÓRICO DO ENEM

Conforme Brasil (1996) no ano de 1998, na gestão do até então presidente, Fernando Henrique Cardoso, foi aplicada a primeira edição do Exame Nacional do Ensino Médio (ENEM), no qual fazia parte de uma política de reestruturação do Ensino Médio, o qual deu início com a LDB, Lei 9.394.

Contudo, levando em consideração que tiveram embasamentos em confiabilidades e desenvolturas, bem como também, regularizadas em um discurso de interdisciplinaridade, o ENEM em sua primeira aplicação, apresentou-se como uma prova ao qual era constituída por 63 questões multidisciplinares. O exame nacional, de início, contou, com cerca de 151.221 inscritos e 115.575 participantes, deste modo, em relação a taxa de abstenção, pode-se dizer que esta foi de 26,5% dos inscritos, sendo a mesma aplicada em 184 municípios, abrangendo e compreendendo todas as capitais e municípios com densidade expressiva de matrículas no Ensino Médio, segundo deliberação e decisão sobreposta através da Portaria MEC 438/1998, no qual estabelece o mencionado Exame (INEP, 1998).

Entretanto, segundo Brasil (1996) a partir do momento em que fazemos um comparativo em relação ao número de inscritos do ano de 1998 com a edição de 2016, no qual contabilizou cerca de 8,6 milhões de inscrições (INEP, 2016), podemos perceber um aumento significativo, bem como ainda, é possível compreender o quanto a primeira edição foi limitada. Apesar disso, considera-se que o fato do total menor de participantes da primeira edição do ENEM pode ser esclarecido e ilustrado por sua aplicação se limitar e abreviar-se às capitais, bem como também, às grandes cidades, porém, além disso, através dos alvos esquematizados de início por meio da Portaria MEC 438/1998, ao qual pode ser melhor observado a seguir:

Artigo 1º - Instituir o Exame Nacional do Ensino Médio – ENEM, como procedimento de avaliação do desempenho do aluno, tendo por objetivos:

I – Conferir ao cidadão parâmetro para autoavaliação, com vistas à continuidade de sua formação e à sua inserção no mercado de trabalho;

II – criar referência nacional para os egressos de qualquer das modalidades do ensino médio;

III – fornecer subsídios às diferentes modalidades de acesso à educação superior;

IV – constituir-se em modalidade de acesso a cursos profissionalizantes pós-médio (BRASIL, 1998).

De tal modo, ao analisar a citação acima mencionada, pode-se perceber que segundo a Portaria MEC 438/1998, o primeiro ENEM operava como referência de autoavaliação do aluno ao qual encontrava-se em fase de conclusão da Educação Básica, bem como ainda, municiar documentos e elementos de parâmetros nacionais para o governo federal e Instituições de Ensino Superior (IES) no que se diz respeito à qualidade do Ensino Médio. De tal modo, pode-se dizer ainda que, as médias do ENEM proporcionavam um modo de avaliação singular e de rede, sem expansão de sua utilização para promoção e entrada ao Ensino Superior ou até mesmo outros programas e projetos oferecidos pelo governo, como viria a ser em seguida.

No ano de 2001, houve um crescimento significativo nas inscrições do ENEM, chegando a atingir pela primeira vez, o total de 1 milhão de inscritos, tendo em vista que neste ano houve a gratuidade do Exame para os alunos nos quais frequentaram todo o Ensino Médio em escolas da rede pública (INEP, 2017).

Já no ano de 2005, esse número teve um novo aumento, visto que neste ano chegou-se a marca de 3 milhões de inscritos, tendo em vista ainda que estes alunos encontravam-se vinculados à concepção do Programa Universidade para Todos (ProUni), no qual passou a ser vigorado através da medida provisória 213/04, bem como também, institucionalizado através Lei 11.096/05 no ano de 2004 (INEP, 2017).

Conforme Santos, Souza e Silva (2018) o MEC no ano de 2009, recomendou e sugeriu à Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior uma reestruturação do ENEM, ao qual o mesmo passaria a ter seu uso como sistema unificado de seleção para acesso e sendo

este, porta de entrada para as Instituições Federais de Ensino Superior (IFES). Assim sendo, seria desempenhado através do ENEM sua função de redemocratizar o ingresso ao Ensino Superior. Com isso, passou-se a seguir a proposta mencionada anteriormente e o ENEM passou por uma reestruturação em seu núcleo de confiabilidades e disposições, tornando-se mais próximo das Diretrizes Curriculares Nacionais, além dos Parâmetros Curriculares Nacionais (PCNs) e também, dos currículos escolares, passando-se a ser uma ferramenta de reformulação curricular do Ensino Médio (SANTOS; SOUZA; SILVA, 2018). Ainda de acordo com Santos, Souza e Silva (2018, p. 115):

Na premissa apresentada pela LDB, Lei 9.394/96, a escola deve educar o aluno para que possa alcançar os mais altos níveis de conhecimento, pautando-se na preparação para o mundo do trabalho e cidadania. Essa preparação deveria basear-se na contextualização (PCNs), porém, apesar disto, as avaliações de acesso ao Ensino Superior mantinham-se conteudistas. Assim, o aluno, ao término da Educação Básica, seria avaliado com base na capacidade de armazenar informações e encontrar a alternativa correta para questões isoladas em seu espaço disciplinar.

Contudo, é importante ainda lembrar que a base legal dessa reestruturação do ENEM passou a ser defendida por meio da Portaria MEC 109/09, tal qual apresentava-se consigo as consequentes modificações aos objetivos do Exame expostos e evidentes na Portaria 438/98:

II - estruturar uma avaliação ao final da educação básica que sirva como modalidade alternativa ou complementar aos processos de seleção nos diferentes setores do mundo do trabalho;

III - estruturar uma avaliação ao final da educação básica que sirva como modalidade alternativa ou complementar aos exames de acesso aos cursos profissionalizantes, pós-médios e à Educação Superior (BRASIL, 2009).

Assim, tendo em vista tais categorias, ficou explícito através da Portaria MEC 109/09 a finalidade de poder operar como avaliação final da Educação Básica, confiando e também esperando que seu núcleo de parâmetro, em união com a reestruturação na organização da prova do ENEM.

O exame no ano de 2017, passou por modificações concernente aos dias que tinha aplicação da prova, e com a consulta pública, no ano citado, foi

possível aplicar a prova em dois domingos sucessivos, ou contínuos, no entanto, foi conservada a sua estrutura no que se diz respeito ao número de questões por área e a aplicação da redação (INEP, 2017).

De acordo com o INEP (2020, p.6) para atender a demanda dos usuários sobre informações específicas, são disponibilizadas as provas, os gabaritos, as informações sobre os itens, e as notas e o questionário respondido pelos inscritos no Enem. Os bancos de dados atendem aos parâmetros da Lei de Acesso à Informação (LAI) e são disponibilizados na aba Dados Aberto/Microdados/Enem do portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), em formato “.csv” (formato de arquivo que contém valores separados por delimitador com ponto-e-vírgula) e os inputs para a leitura desses arquivos foram elaborados utilizando os softwares SAS, SPSS e R. Não obstante, se o usuário desejar, poderá não usar os inputs para abrir as bases, pois o formato “.csv” também foi adotado por facilitar a leitura dos arquivos, independentemente do software estatístico utilizado.

Para facilitar a utilização dos resultados, o arquivo principal contém os questionários respondidos pelos participantes, congregando em um único arquivo as informações disponibilizadas sobre os participantes do Enem. Nela, para cada tipo de prova, foram carregadas as informações de posição do item na prova, habilidade, cor da prova, área de conhecimento e gabarito (MICRODADOS ENEM, 2019).

DADOS ABERTOS

Segundo Brasil (2020) a Política de Dados Abertos do poder executivo federal, ao qual passou a ser estabelecida no dia 11 de maio de 2016, através do Decreto nº 8.777, tem como principal finalidade de possibilitar a publicação e dispersão de dados aos quais estão contidos em bases de órgãos e entidades da administração pública federal direta, autárquica e fundacional.

De tal modo, consta-se assinalar que as informações de veemência coletiva, nos quais anteriormente encontravam-se de modo limitados ao domínio da administração pública, sobrevêm a ter maior acessibilidade por parte de toda sociedade no formato de dados abertos, objetivando deste modo

o aperfeiçoamento da cultura da transparência, bem como também, do controle social e da inovação (BRASIL, 2020).

Assim sendo, conforme Brasil (2020) o Ministério da Mulher, da Família e dos Direitos Humanos, em cumprimento ao art. 5º, parágrafo 2º do decreto acima mencionado, através deste mesmo documento formaliza o seu Plano de Dados Abertos (PDA).

Sendo o PDA uma ferramenta importante, no qual operacionaliza a Política de Dados Abertos do Ministério, de maneira a solidificar as orientações em relação ao processo de planejamento, registro, abertura, monitoramento e ateuo da utilização de suas bases de dados, tornando-se assim mais fácil a compreensão do público em se tratando dos documentos que se encontram disponibilizados (BRASIL, 2020).

Contudo, pode-se dizer que, depois de um trabalho de sensibilização e envoltura, inúmeros campos e domínios disponibilizaram suas bases de informações para compor o inventário de dados do MMFDH.

MATERIAIS E MÉTODOS

Para realização do estudo, além da consulta às diversas fontes bibliográficas relacionadas à problemática, utilizamos como abordagem, uma pesquisa quantitativa de caráter exploratório (Gil,2002).

O banco de dados que foi utilizado neste trabalho contempla informações dos microdados do Enem 2019 que são o menor nível de desagregação de dados recolhidos por meio do exame. Eles atendem à demanda por informações específicas ao disponibilizar as provas, os gabaritos, as informações sobre os itens, as notas e as respostas do questionário socioeconômico respondido pelos inscritos no ENEM. A referida base contém informações de 5.095.270 candidatos inscritos no ENEM 2019, relativa a 100 variáveis. Contudo, visando atender aos objetivos propostos do trabalho, delimitamos nosso estudo aos dados dos estudantes residentes no município de Cuité-PB, o equivalente a 756 inscritos, relativos a 11 variáveis, dentre as disponibilizadas na base de dados original, conforme quadro 1.

Quadro 1. Nome e descrição das variáveis utilizadas no processo de análise.

Nome da variável	Descrição
SG_UF_RESIDENCIA	Sigla da Unidade da Federação de residência
NU_IDADE	Idade
TP_SEXO	Sexo
TP_ESTADO_CIVIL	Estado Civil
TP_COR_RACA	Cor/raça
NO_MUNICIPIO_RESIDENCIA	Nome do município de residência
NU_NOTA_REDACAO	Nota da prova de redação
TP_STATUS_REDACAO	Situação da redação do participante
Q006	Qual é a renda mensal de sua família? (Some a sua renda com a dos seus familiares.)
Q024	Na sua residência tem computador?
Q025	Na sua residência tem acesso à Internet?

Fonte: INEP/2021

Como ferramentas de auxílio para a realização deste trabalho foram utilizados, basicamente, dois softwares: R (R CORE TEAM, 2021) e Microsoft Excel 2019 (MICROSOFT, 2019). Os pacotes computacionais disponíveis para serem executados no software R e que foram utilizados neste trabalho são: rpart (THERNEAU et al., 2013) para indução das árvores de decisão e data.table (MATT DOWLE E ARUN SRINIVASAN, 2021) útil para manipulação de grandes bases de dados (1 GB para 100 GB na memória RAM).

RESULTADOS

ANÁLISE DESCRITIVA

Na figura 4 e quadro 2 podemos observar uma análise descritiva de todas as variáveis explicativas disponíveis na base de estudo (756 observações), segmentadas em categorizadas e numéricas, respectivamente.

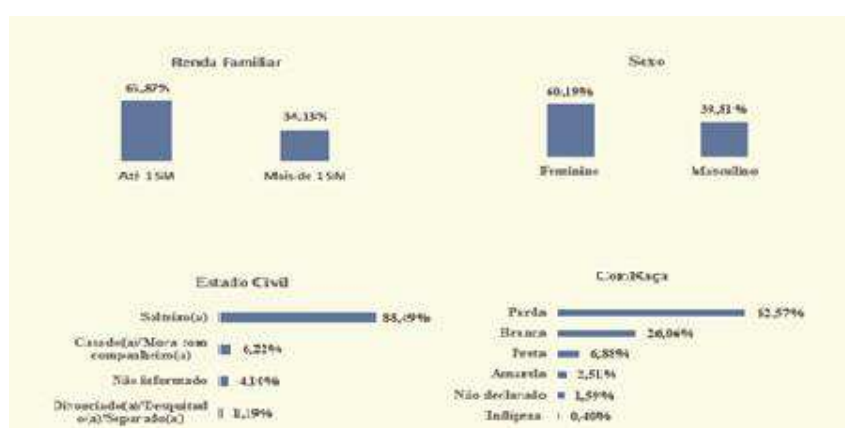
Para as variáveis categorizadas, figura 6 observa-se a proporção da variável resposta para cada categoria. Analisando a variável renda familiar, pode-se perceber que grande parte dos inscritos, ou seja, aproximadamente 66% têm renda familiar de até um salário-mínimo. Segundo Silva et al., (2020) a

renda familiar, mostra-se relacionada ao desempenho dos candidatos em todos os estudos encontrados em uma vasta revisão bibliográfica. Tais estudos analisaram microdados de diversas edições do Exame Nacional do Ensino Médio (ENEM) e utilizaram diversas técnicas estatísticas para fazerem a associação entre as variáveis essas socioeconômicas e o desempenho dos candidatos (FIGUEIRÊDO; NOGUEIRA; SANTANA, 2014; VICENTE, 2014; MENDES; KARRUZ, 2015; PIRES, 2015; LOBO; CASSUCE; CIRINO, 2016; CARVALHO, 2017; SILVA et al., 2017).

Estes mesmos autores relatam que atualmente, o acesso a bens e serviços de tecnologia pode ser um incremento considerável para a educação. No que diz respeito ao acesso à internet e a computador em casa, e citam os achados bibliográficos que trouxeram estudos empíricos que revelaram associação entre essa condição e a nota dos participantes (VICENTE, 2014; MENDES; KARRUZ, 2015; LOBO; CASSUCE; CIRINO, 2016). Destacamos que para a base de dados avaliada apesar do percentual significativo de candidatos pontuando baixa renda familiar, temos que aproximadamente 71% afirmaram ter internet em casa, porém, acredita-se que o acesso a esta ferramenta tecnológica seja realizada de forma precária, pois, aproximadamente 60% dos candidatos pontuaram não ter computador em casa.

Por fim, destacamos que 72% dos candidatos se declararam de cor/raça parda, preta, amarela ou indígena, para Silva, et al., (2020) a raça/cor da pele é uma das variáveis socioeconômica considerada de impacto significativo no desempenho dos candidatos no exame. Nessa variável, o desempenho educacional reproduziria injustiças étnico-raciais históricas da sociedade brasileira.

Figura 4: Distribuição de percentuais das variáveis categóricas utilizadas na pesquisa, referente aos candidatos que prestaram ENEM 2019 e declararam residência no município de Cuité-PB





Base: 756 (número de inscrito no ENEM 2019 e que declararam residência no município de Cuité-PB)

Fonte: INEP/2021

Quadro 2. Distribuição de medidas de posição e dispersão das variáveis quantitativas utilizadas na análise dos candidatos que prestaram ENEM 2019 e declararam residência no município de Cuité-PB.

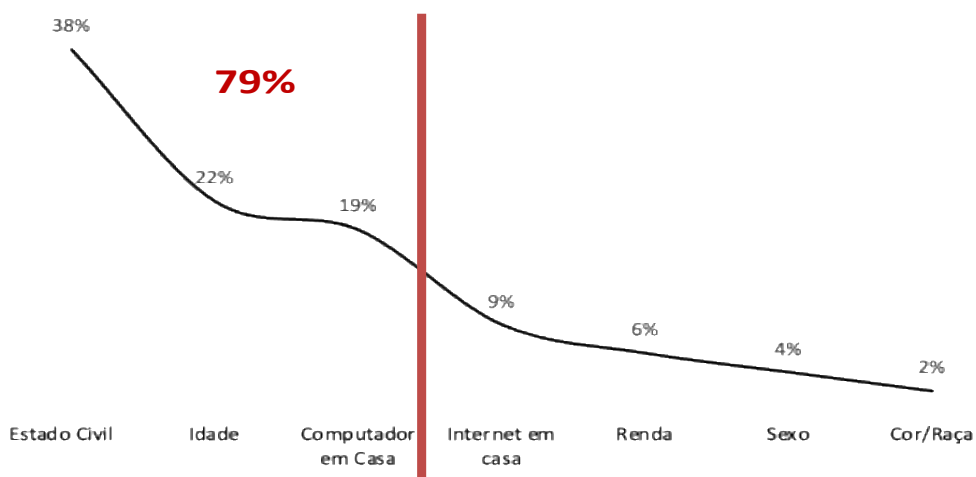
Medidas	Idade	Redação
Média	21,61	545,74
Mediana	20,00	560,00
Moda	18,00	620,00
Desvio-padrão	6,01	178,18

Base: 756 (número de inscrito no ENEM 2019 e que declararam residência no município de Cuité-PB)

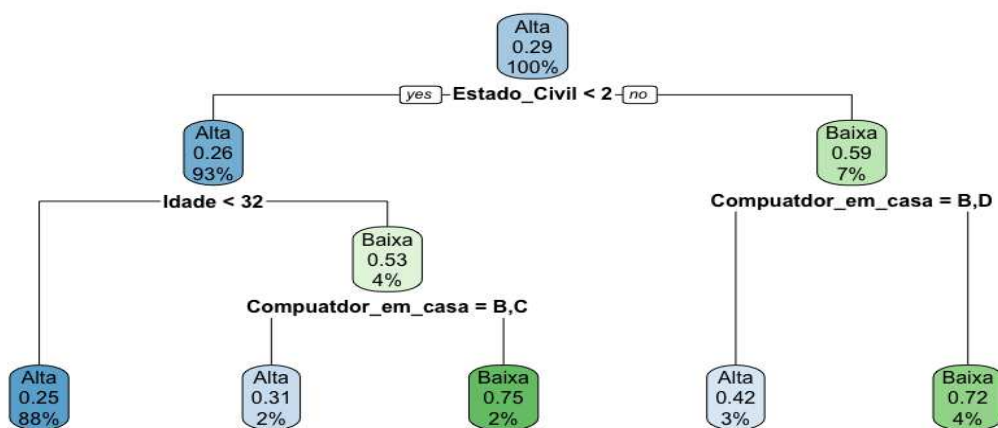
Fonte: INEP/2021

ÁRVORE DE DECISÃO

A análise empírica sobre as variáveis que utilizamos para tentar explicar realmente o fenômeno de classificação de conceito alto e baixo na redação do ENEM concentra 79% desta significância nas variáveis Estado Civil, Idade e Computador em casa, pois as variáveis explicativas devem ser selecionadas não somente, pela correlação forte com a variável dependente, mas sim por explicar adequadamente o fenômeno. De acordo com Breiman, e Friedman (1984), as variáveis mais utilizadas para explicar fenômenos de competência, através de análise desagregada, são as variáveis socioeconômicas.

Figura 5: Índice de importância da variável para o modelo

Fonte: Própria

Figura 6. Árvore de decisão para matriz de dados dos candidatos do ENEM/2019 com residência em Cuité/PB.

Fonte: Própria

No topo da árvore, temos a probabilidade geral do candidato tirar uma nota igual ou maior que 500 na redação. Podemos observar a proporção de candidatos que obtiveram 29% deste escore.

Seguindo, os caminhos da árvore, temos a classificação dos candidatos em dois grupos, 93% têm idade abaixo de 32 anos, sendo 26% a probabilidade de que candidatos deste grupo pontuem uma nota maior ou igual a 500.

Para aproximadamente 4% dos candidatos com computador em casa, 31% têm chance de tirar uma nota maior ou igual a 500 na redação. Por outro lado, 75% dos que não têm pode tirar uma nota menor de 500 na redação.

Com uma acurácia de aproximadamente 75%, o modelo previu corretamente na simulação 96 candidatos com nota maior ou igual a 500, mas classificou 19 com nota menor que 500, quando na realidade eles tiraram nota maior ou igual a 500. Por analogia, o modelo classificou erroneamente 10 candidatos com nota maior ou igual a 500 enquanto eles estavam com nota abaixo do escore.

Quadro 3. Matriz de confusão para os dados de teste dos candidatos do ENEM/2019 com residência em Cuité/PB.

Atributos	Maior ou igual a 500	Menor que 500
Maior ou igual a 500	96	19
Menor que 500	10	27

Fonte: Própria

CONCLUSÃO

Para os resultados obtidos, válidos para os dados contextualizados neste estudo, podemos pontuar que árvores de decisão são algoritmos de aprendizado versáteis que podem realizar tarefas de classificação e regressão. Podendo serem considerados algoritmos significativamente poderosos, capazes de ajustar conjuntos de dados complexos de forma simples, dado que não é necessária uma pré-seleção de variáveis, uma vez que o algoritmo busca as variáveis mais significativas disponíveis. Outra vantagem vista para este modelo é sua fácil apresentação e interpretação dos resultados, podendo ser lido por pessoas sem um avançado conhecimento técnico.

Considerando as premissas do modelo, apenas três variáveis foram consideradas significativa na plotagem da árvore de decisão, que foram: estado civil, idade e computador em casa.

Na simulação de previsão do modelo, foi verificado uma precisão de acerto de 75%. Diante do exposto, apesar do modelo de previsão obtido neste estudo se mostrar adequado, entendemos que outros modelos devem ser

validados em trabalhos futuros buscando melhorar esta precisão inserindo-se novas variáveis a análise.

REFERÊNCIAS

AHMED, I. et al. Fast content-based file type identification. In: SPRINGER. **IFIP International Conference on Digital Forensics**. [S.l.], 2011. p. 65–75.

BELL, J. Machine Learning: Hands-on for developers and technical professionals. [S.l.]: **John Wiley & Sons**, 2014.

BORGES, Wagner de Souza. Probabilidades. **Revista Primus Vitam**. Nº 2 , 1º semestre de 2011.

BRASIL. Ministério da Mulher, da Família e dos Direitos Humanos. **Dados abertos**. Publicado em: 30/08/2018 19h28. Atualizado em 01/09/2020. Disponível em: <https://www.gov.br/mdh/pt-br/acesso-a-informacao/dados-abertos-1>. Acesso em: 01 de março de 2021.

_____. **Portaria MEC nº 438, de 28 de maio de 1998**. Institui o Exame Nacional do Ensino Médio – ENEM. 1998. Disponível em: <http://www.crmariocovas.sp.gov.br/pdf/diretrizes_p0178-0181_c.pdf>. Acesso em: 01 de março de 2021.

_____. **Lei nº 9.394, de 20 de dezembro de 1996**. Estabelece as diretrizes e bases da educação nacional Lei de Diretrizes e Bases da Educação Nacional. Disponível em: <http://www.crmariocovas.sp.gov.br/pdf/diretrizes_p0178-0181_c.pdf>. 01 de março de 2021.

_____. **Portaria INEP nº 109, de 27 de maio de 2009. Estabelece a sistemática para a realização do Exame Nacional do Ensino Médio no exercício de 2009**. 2009. Disponível em: <https://www.legisweb.com.br/legislacao/?id=214657>. Acesso em: 01 de março de 2021.

Breiman, L.; Friedman, J.H.; Olshen, R.A. e Stone, C.J., Classification and Regression Trees. Wadsworth International Group, Califórnia (1984)

DANTAS, C. A. B. **Probabilidade**: Um Curso Introdutório. 2ed. São Paulo: Editora da Universidade de São Paulo, 2004.

GIL, Antônio Carlos. Como elaborar projetos de pesquisa. 4. ed. - São Paulo: Atlas, 2002.

GOLLAPUDI, S. Practical Machine Learning. [S.l.]: **Packt Publishing Ltd**, 2016.

GRAÇA MARTINS, M. E., MONTEIRO, C., VIANA, P. V., TURKMAN, M. A. A. (1999) – *Probabilidades e Combinatória*. Ministério da Educação, Departamento do Ensino Superior. ISBN: 972-8417-33-0. Depósito Legal 143440/99.

INEP - INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Exame Nacional do Ensino Médio**: Relatório Final 98. 1998. Disponível em: <http://portal.inep.gov.br/informacao-da-publicacao/-/asset_publisher/6JYIsGMAMkW1/document/id/494207>. Acesso em: 01 de março de 2021.

INEP - INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **ENEM 2016**: resultado individual. 2016. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/downloads/2016/apresentacao_final_resultados_2016.pdf>. Acesso em: 01 de março de 2021.

INEP - INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **MEC e Inep anunciam mudanças no exame em função de consulta pública**. 2017. Disponível em: http://inep.gov.br/artigo2/-/asset_publisher/GngVoM7TApe5/content/mec-e-inep-anunciam-mudancas-noexame-em-funcao-de-consulta-publica/21206?inheritRedirect=false. Acesso em: 01 de março de 2021.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do Enem 2019**. Brasília: Inep, 2019. Disponível em: <<http://portal.inep.gov.br/web/guest/microdados>>. Acesso em: 20 jun. 2020.

JAMES, G. et al. An introduction to statistical learning. [S.l.]: **Springer**, 2013. v. 6.

JAMES, B. R. **Probabilidade**: um curso de nível intermediário. 3ª ed. Rio de Janeiro: IMPA, 2006.

LAURETTO, Marcelo S. **Árvores de Decisão**. EACH-USP, Novembro/2010. Disponível em: https://edisciplinas.usp.br/pluginfile.php/4469825/mod_resource/content/1/ArvoresDecisao_normalsize.pdf. Acesso em: 04 mai. 2021.

MAGALHÃES, MN; LIMA, ACP. **Noções de Probabilidade e Estatística**. São Paulo: EDUSP, 2008.

MONTGOMERY, DC; RUNGER, GC. **Estatística aplicada e probabilidade para engenheiros**. Rio de Janeiro: LTC Editora, 2012.

MOREIRA, Andrea de Paula Machado. **Aplicações da teoria da decisão e probabilidade subjetiva em sala de aula do ensino médio** / Andréa de Paula Machado Moreira. – Campinas, SP : [s.n.], 2015.

MORGADO, A. C. O.; CARVALHO, J. B. P.; CARVALHO, P. C. P. e FERNANDEZ, P. **Análise Combinatória e Probabilidade**. 9ª ed. Rio de Janeiro: SBM, 2006.

ROCHA, Ana Angelita da; RAVALLEC, Carmem Teresa Gabriel. Enem nos documentos: uma leitura pós-fundacional da reestruturação do exame em 2009. **Revista e-Curriculum**, São Paulo, v. 12, n. 3, p.1993-2018, out./dez. 2014.

SANTOS, Jean Mac Cole Tavares; SOUZA, Francisco das Chagas Silva; SILVA, Francisco José Balduino da. Exame Nacional do Ensino Médio: considerações sobre a influência do ENEM na produção do currículo de história. **História & Ensino**, Londrina, v. 24, n. 1, p. 111-133, jan./jun. 2018.

Silva, A. Pinheiro; Pontes Junior, J. A. de F.; Magalhães Júnior, A. G. PERFIL DOS PARTICIPANTES DO EXAME NACIONAL DO ENSINO MÉDIO (ENEM) EM 2017: COMPARAÇÃO ENTRE BRASIL, NORDESTE E CEARÁ. *Educação & Linguagem* · ISSN: 2359-277X · ano 7 · nº Especial 2 · p. 12-24. JUN. 2020.

SMOLE, K. S.; DINIZ, M. I. **Matemática Ensino Médio**. Volume 2.6ª ed. São Paulo: Saraiva, 2010.