



UNIVERSIDADE FEDERAL DO ACRE
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÕES

PREVISÃO DO DESEMPENHO DOS PARTICIPANTES DO ENEM NO ESTADO
DO ACRE A PARTIR DE CARRACTERÍSTICAS SOCIOECONÔMICAS E
ACADÊMICAS

RIO BRANCO
2022

LUCAS DE LIMA CHAVES

**PREVISÃO DO DESEMPENHO DOS PARTICIPANTES DO ENEM NO ESTADO
DO ACRE A PARTIR DE CARRACTERÍSTICAS SOCIOECONÔMICAS E
ACADÊMICAS**

Projeto de monografia apresentado
como exigência parcial para
obtenção do grau de bacharel em
Sistemas de Informação da
Universidade Federal do Acre.

Prof. Orientador: Dr. Manoel Limeira
de Lima Júnior

RIO BRANCO

2022

SUMÁRIO

1. Introdução.....	4
2. Problema.....	5
3. Objetivos	7
3.1 Objetivo Geral.....	7
3.2 Objetivos Especificos	7
4. Justificativa	9
5. Fundamentação Teórica	10
5.1 Knowledge Discovery in Databases	10
5.1.1 Regressão	13
5.1.2 Algoritmos de Regressão.....	14
5.1.2.1 Linear Regression.....	14
5.1.2.2 Instanced Based Learning	14
5.1.2.3 SMOReg.....	15
5.1.2.4 Randon Forest	15
5.1.2.5 M5P	16

5.2 Exame Nacional do Ensino Médio	16
6. Metodologia.....	18
7. Esboço do capítulo e seções	20
8. Cronograma	22
Referências.....	23

1 INTRODUÇÃO

O Exame Nacional do ensino Médio (ENEM) tem como objetivo avaliar o desempenho escolar ao final da educação básica. realizado anualmente desde 1998, o exame permite o acesso ao ensino superior por meio de programas como o Sistema de Seleção unificada (SISU), o programa Universidade para Todos (PROUNI) e o programa de Financiamento Estudantil (FIES). No entanto, a atuação do participante é crucial para isso, pois as vagas são limitadas, permitindo apenas a entrada do primeiro classificado (INEP, 2019a).

Os dados dos participantes do ENEM são coletados por meio de inscrições e provas e atendem à demanda de informações específicas, como questões de provas, respostas, informações sobre os elementos, notas e perfil socioeconômico dos membros (INEP, 2019b). Esses dados e resultados do ENEM podem ser usados para pesquisas e indicadores educacionais (INEP, 2019a). Segundo Junior (2018), esses dados são fornecidos pelo INEP e pode haver um conjunto de dados embutidos que pode ajudar, por exemplo, a entender os fatores associados ao desempenho dos participantes que realizaram o exame. E dentre estes fatores temos os socioeconômicos que, segundo Silveira, Barbosa e Silva (2015), são responsáveis por 66% da variância da nota dos participantes.

Portanto, para a extração do conhecimento a partir deste conjunto dados, que possa auxiliar no objetivo de prever o desempenho do candidato com base em seus atributos socioeconômicos, faz-se necessário uma atividade específica e adequada para este objetivo. Assim sendo proposto a mineração de dados, que será executada através de um processo de descobrimento de conhecimento, na base de dados do Enem 2021, no que tange aos participantes do estado do Acre.

Por fim, tomando como base tais informações, foi escolhida a mineração de dados como atividade principal, a fim de se analisar os dados dos participantes Acreanos do ENEM 2021, tendo como objetivo prever o desempenho dos candidatos com base nos seus atributos socioeconômicos.

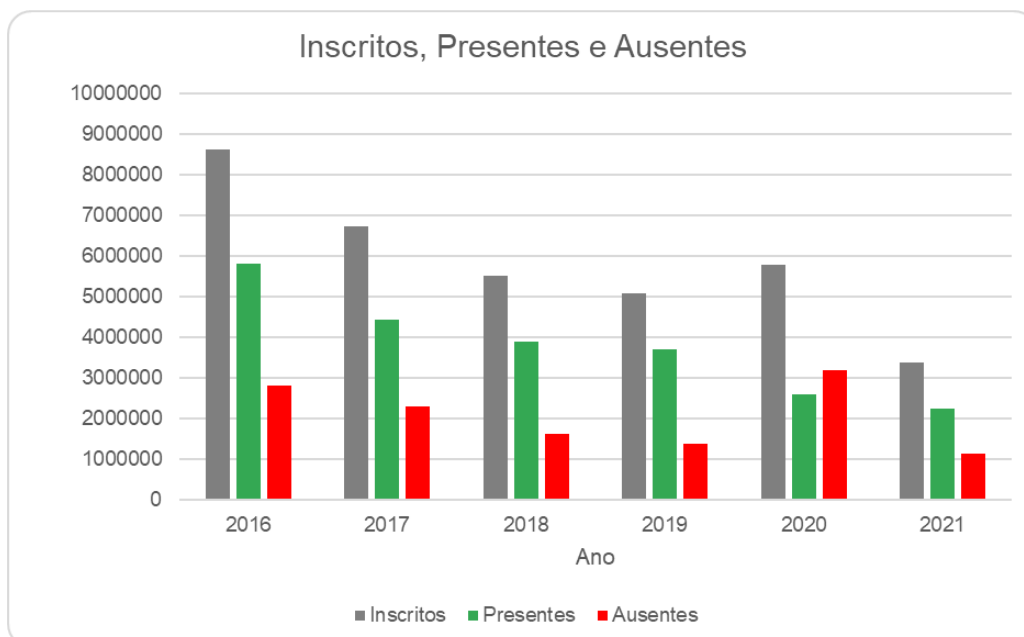
2 PROBLEMA DA PESQUISA

O Enem (Exame Nacional do Ensino Médio) de 2021, segundo o jornal O Globo em uma matéria realizada no mesmo ano, teve 3,1 milhões de inscritos confirmados, o menor número desde 2005. Vários são os motivos que podem ser estipulados como causadores desse fenômeno, dentre eles podemos citar a pandemia de Covid-19 atrelada a uma crise econômica além do pagamento da taxa de R\$ 85,00 para realização da prova (Oglobo, G1, 2021).

Conforme a Figura 1 tendo como eixo y e x, respectivamente o número de candidatos inscritos em milhões e o ano de realização do Enem, além de apresentar por meio da cor cinza o total de inscritos em cada ano, a cor verde o total de presentes nos dois dias de provas e por fim os ausentes representados no gráfico em vermelho, nota-se que nos últimos anos a busca dos jovens por ensino superior público através do ENEM, que atualmente é a principal forma de ingresso em universidades federais no Brasil, vem diminuindo juntamente a uma significativa taxa de ausência.

Outro fator que se pode analisar na figura 1 foi a quantidade de candidatos ausentes nas edições do Enem 2020 e 2021, onde se observa respectivamente que mais da metade dos inscritos não compareceram para a realização das provas, e que mesmo após o fim do período de isolamento, situação advinda da pandemia de COVID-19, os candidatos não compareceram em maioria se comparado a quantidade de ausentes para a realização das provas evidenciando assim que a pandemia e sua crise econômica pode não serem os únicos contribuintes para o desinteresse dos candidatos para a participação no Enem no decorrer dos últimos anos .

Figura 1 - Relação de candidatos inscritos por unidade da federação.



Fonte – Adaptado de INEP

Segundo pesquisa coordenada pelo Atlas das Juventudes, em apoio a fundação Roberto Marinho, realizada com jovens de diferentes faixas etárias nas cinco regiões do Brasil, somente 3 a cada 10 pretendem realizar o próximo ENEM, a mesma pesquisa ainda revela que uma maioria de jovens que estavam trabalhando (8 a cada 10), dos quais um terço começou a trabalhar devido à pandemia. Tais informações demonstram o vínculo entre educação e fatores socioeconômicos no desempenho dos estudantes.

Almeida e Moreira (2015), discorrem sobre a importância da preocupação com a qualidade na educação assim como no incentivo por parte do estado, para o sucesso acadêmico, crescimento econômico e redução das desigualdades sociais de seus estudantes.

Trazendo essa questão para o âmbito estadual, segundo lista oficial pública pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) o estado do Acre obteve a segunda menor quantidade de inscritos, se comparado aos outros estados brasileiros, evidenciando a importância do estado de fomentar e investir em políticas de apoio à educação e da prestação de incentivo aos jovens que

estão saindo do ensino médio para buscarem uma melhor qualificação no que tange a educação superior pública.

Dessa forma, buscaremos prever o desempenho dos candidatos, com base em seus respectivos dados socioeconômicos, preenchidos em um formulário anexado no ato da inscrição, além de analisar quais destes atributos socioeconômicos e acadêmicos têm mais influência nas notas obtidas pelos candidatos do Enem 2021 no estado do Acre.

Em resumo este trabalho tem como objetivo analisar a problemática “Prever o desempenho dos candidatos do Enem 2021 no Estado do Acre, a partir de seus atributos socioeconômicos.”, gerando assim um conhecimento útil para as autoridades competentes atuarem com uma maior eficiência e eficácia no foco das dificuldades passadas no dia-a-dia dos estudantes como um todo na educação do estado do Acre.

3 OBJETIVOS DA PESQUISA

Nesta seção estão definidos os objetivos desta pesquisa, tanto de forma geral quanto de forma específica.

3.1 OBJETIVO GERAL

Prever o desempenho dos alunos utilizando atributos socioeconômicos, como renda, estrutura familiar, tipo de ensino dentre outros, presentes no questionário respondido pelos inscritos no Exame Nacional do Ensino Médio - ENEM 2021 no estado do Acre.

3.2 OBJETIVOS ESPECÍFICOS

1. Extrair e processar dados a partir da base de dados do Enem 2021 presentes na base de dados abertos do Ministério da Educação, disponibilizada pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais);
2. Aplicar técnicas e algoritmos de mineração de dados para prever o desempenho dos alunos;

3. Analisar padrões encontrados expondo os atributos mais significativos;
4. Responder às questões levantadas como hipóteses na pesquisa.

4 JUSTIFICATIVA DA PESQUISA

Durante a Conferência “Desigualdades no Acesso à Educação Superior por Populações Desfavorecidas da Região da América Latina e do Caribe no Contexto da Pandemia COVID-19”, realizada, virtualmente, pelo Instituto Internacional da UNESCO para a Educação Superior na América Latina e no Caribe (IESALC), vinculado à Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), foram levantados dados que mostram a relação entre o acesso ao ensino superior e perfil social e econômico dos ingressantes, onde percebeu-se que o percentual de matriculados mais pobres entre 2000 e 2018, foi de 5%, subindo para 10% em 2018; e entre os mais ricos, no mesmo período, o percentual foi de 22%, se situando em 22% para 2018. Piorando ainda mais com a pandemia de COVID-19, na qual acarretou em um cenário mais acentuado de exclusão.

No âmbito estadual, segundo estudo realizado pela Fundação Getúlio Vargas (FGV) no ano de 2021, o estado do Acre conta com 45,53% pessoas (do total da população do estado) vivendo com renda per capita até R\$ 497 reais mensais, fazendo assim o estado ocupar o 13º lugar no ranking dos estados mais pobres do país.

Portanto, o presente trabalho visa trazer benefícios para a sociedade e, principalmente, para o departamento de educação do estado Acre, pois com a previsão e a informação dos atributos que mais influenciam no desempenho dos estudantes em um exame de suma importância como o Enem, existe a possibilidade de reagir ao problema, ou seja, as autoridades competentes podem traçar metas e diretrizes para melhorar políticas de seguridade social das pessoas de baixa renda e seu acesso à educação pública de qualidade, principalmente aqueles em estado de vulnerabilidade social.

5 FUNDAMENTAÇÃO TEÓRICA

Nesta seção serão abordados alguns conceitos fundamentais para realização deste trabalho tais como: Knowledge Discovery in Databases (KDD), tarefa de regressão, algoritmos de regressão, Enem.

5.1 KNOWLEDGE DISCOVERY IN DATABASES (KDD)

Conforme aponta Fayyad, et al. (1996), historicamente, o conceito de encontrar padrões úteis em dados recebeu vários nomes, dentre eles mineração de dados, extração de dados, descoberta de dados, mineração de dados, arqueologia de dados e mineração de padrões. O termo mineração de dados tem sido usado principalmente por estatísticos, analistas de dados e neste artigo, usamos o termo "padrão" para nos referirmos a um padrão ou modelo de extração de dados.

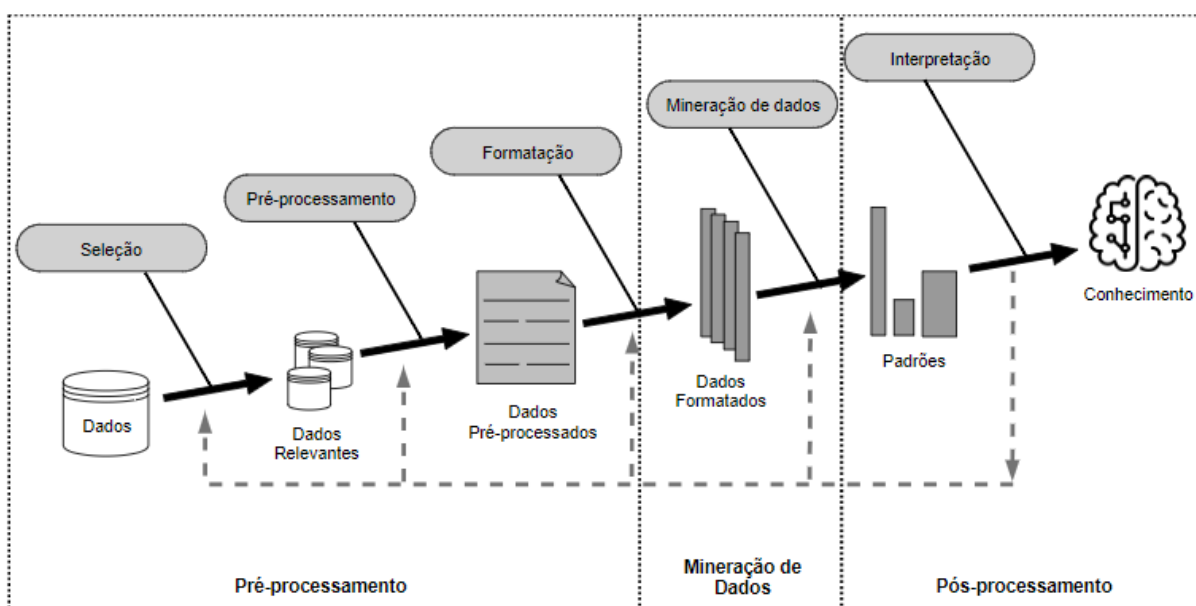
O termo KDD foi cunhado no primeiro workshop de KDD em 1989 (Piatetsky, Shapiro 1990) para enfatizar que "conhecimento" está no fim do processo de descoberta. Entrando mais a fundo, o processo de KDD filtra e identifica padrões nos dados analisados, que uma vez processados, geram informações válidas e estratégicas auxiliando na tomada de decisões mais efetivas. (Fay et al., 1996). Segundo Piatetsky e Shapiro (1990), a atividade de KDD tem cinco tarefas que se relacionam entre si, que são:

1. **Seleção:** etapa que é realizada a escola dos dados, é analisado sua relevância e disponibilidade.
2. **Pré-processamento:** processo de preparação dos dados para a aplicação das técnicas de mineração.
3. **Transformação:** delimitação das normas básicas de representação dos dados utilizados, determinado sua padronização na base a ser gerada como base de dados resultado.
4. **Mineração de dados:** aplicação das técnicas e algoritmos de mineração de dados inteligentes, segundo o objetivo da pesquisa, algumas dos algoritmos utilizados são: redes neurais, árvores de decisão, análise de clustering, dentre outros.

5. **Interpretação:** esta etapa está relacionada á análise dos dados e a verificação se os resultados obtidos com a mineração de dados fazem sentido ou possuem alguma validade para o problema proposto.

Já para Bravin et al. (2019) a atividade de KDD, consiste em três fases que podem ser observadas na figura 2, ou sub processos que são: pré-processamento, mineração de dados e pós-processamento.

Figura 2 - Etapas do KD



Fonte – Bravin et al. 2019

A primeira etapa inclui atividades relacionadas à coleta, organização e processamento de dados e visa preparar os dados para os algoritmos que os irão consumir na etapa seguinte.

Sucessivamente a etapa de pré-processamento de dados, a fase de mineração de dados depende muito de aprendizado de máquina, reconhecimento de padrões, otimização, simulação, estatísticas e técnicas de análise multivariada para descobrir padrões nos dados.

As técnicas de mineração de dados são operações usadas para encontrar padrões em grandes quantidades de dados. Esses padrões podem ser explicativos, para descrever relacionamentos entre dados, ou preditivos, para prever valores futuros com base em dados passados (Matos G. 2006).

Dentro do processo de KDD, a tarefa de mineração de dados pode ser dividida em duas categorias:

- I. **Tarefas Preditivas:** Segundo Júnior (2017), preveem valores futuros com base em dados passados. Dentro da categoria preditiva, ainda temos como principais atividades as tarefas de:
 - a. Classificação, constituídas de variáveis discretas.
 - b. Regressão, quando a variável dependente é numérica.
- II. **Tarefas Descritivas:** Descrevem relacionamentos entre dados, suas tarefas mais conhecidas são:
 - a. Regras de associação e agrupamento ou *Clustering* (Júnior, 2017).

Este trabalho fará uso da tarefa de regressão, pois é adequada para a manipulação de valores não discretos, ou seja, numéricos, essa tarefa será detalhada na próxima subsecção.

Por fim, a derradeira etapa se constitui do tratamento e análise do conhecimento obtido com o objetivo de viabilizar o que foi descoberto (Goldshmidt & Passos, 2005). As etapas operacionais são descritas a seguir:

- a. Pré-processamento: é a fase onde os dados são selecionados e organizados (Han, et al., 2012).
- b. Mineração de Dados: é a fase onde os dados são processados por meio de algoritmos e técnicas pré-estabelecidas de acordo com o que se busca como objetivo e geram o conhecimento (Goldshmidt & Passos, 2005).
- c. Pós-processamento: a etapa final que visa a análise, interpretação e visualização do modelo de conhecimentos gerado pela etapa de mineração de Dados (Goldshmidt & Passos, 2005).

5.1.1 REGRESSÃO

As tarefas de regressão são semelhantes às tarefas de classificação, nas quais busca-se uma função que associa os registros do banco de dados a uma série de valores reais. A principal diferença é que a propriedade de destino recebe um valor numérico. Algumas situações em que a regressão pode ser usada são: Prever o equivalente de lucro ou demanda com base em condições anteriores, como condições sazonais ou climáticas (NOGARE, 2016). As tarefas de regressão usam estatísticas, redes neurais e outras técnicas que fornecem recursos para sua implementação (EVANGELISTA, POLETO, 2020).

Assim como nas tarefas de classificação, que se enquadram no tipo descritiva, a regressão e sua categoria de predição, usam de dois conjuntos de dados para a confecção de um modelo capaz de fazer as predições das variáveis dependentes (PIN, STEINBACH, 2014):

1. Base de Treinamento: Parte da base de dados onde a variável que buscar-se-á gerar é conhecida, a fim de treinar o modelo a achar a variável dependente.
2. Base de Teste: Conjunto onde de fato o modelo é posto à prova, aqui a variável dependente não é conhecida.

Segundo HAN et al (2019), existem algumas maneiras de se dividir os conjuntos de teste e treinamento, sendo as mais utilizadas:

1. Holdout: As bases são divididas de forma aleatória, usualmente, se divide 66% da base para treino e os outros 33% são voltados a atividade de teste do modelo.
2. Random Subsampling: Funciona basicamente da mesma forma que o holdout quando se trata de divisão da base de dados, mas na avaliação, ele consiste de uma média de repetição de k vezes o Holdout.
3. Cross Validation: A base é dividida em k partes, todas de tamanhos semelhantes, e em cada iteração são executados testes e treinamentos, de forma que a todo momento parte das instancias sejam utilizadas para treinamento enquanto a outra está para teste.

4. Dessa forma, cada item ou instancia de ambas as partes é utilizada um mesmo número de vezes por cada tipo de base, seja teste ou treinamento.

5.1.2 ALGORITMOS DE REGRESSÃO

Os algoritmos de regressão utilizados para realizar a previsão do desempenho dos candidatos são: Linear Regression, IBK, SMOReg, Random Forest e o M5P.

5.1.2.1 LINEAR REGRESSION

A regressão é chamada de linear quando a relação entre a variável preditora e a resposta segue um comportamento linear. Neste caso, você pode criar um modelo onde o valor de y é uma função linear de x . Exemplo: $y = b + wx$. O mesmo princípio pode ser usado para modelos com várias variáveis preditoras (CAMILO, SILVA 2009).

5.1.2.2 ALGORITMO INSTANCE BASED LEARNING (IBK)

O IBK tem como base a família de algoritmos IBL, que são advindos do método matemático Nearest Neighbours (vizinhos mais próximos), mas mesmo sendo da mesma classificação de algoritmos, ou seja, mesma família o IBL tem como

característica ser incremental, assim ele prioriza aumentar a acurácia da classificação de novas instâncias Daelemans e Bosch (2005).

5.1.2.3 ALGORITMO SMOREG

Amplamente conhecido e utilizado, o algoritmo SMOReg é desenvolvido a partir do algoritmo de classificação SMO, que também é uma implementação do SVM, método de *machine learning* utilizado para a tarefa de classificação Shevade et al. (2000).

5.1.2.4 ALGORITMO RANDOM FOREST

No algoritmo de aprendizagem *ensemble* random forest, utilizado para classificação e regressão existe a combinação de vários modelos que o torna mais robusto e complexo para a realização de algumas atividades.

Nele, cada classificador base é uma árvore de decisão, ou seja, são constituídas partir de instancias retiradas do conjunto de treinamento, essas amostras ou instancias da

base, são retiradas de forma aleatória, a fim de possibilitar que as instancias selecionadas possam ser repetidas na seleção BREIMAN (2001).

5.1.2.5 ALGORITMO M5P

Esse algoritmo tem a sua base no algoritmo M5, que se utiliza do conceito de arvores de modelo, uma estratégia parecida com arvores de decisão, mas que no caso desse em específico, suas folhas armazenam modelos de regressão linear, assim o M5P tem capacidade de realizar tarefas de predição com alto grau de dimensionalidade além de permitir que o tamanho das arvores sejam reduzidos ao ponto de não haver muitas perdas, criando assim modelos de compressão mais simples Wang e Witten (1997).

5.2 EXAME NACIONAL DO ENSINO MÉDIO

Em 1998, o exame nacional do ensino médio (ENEM), foi criado com a premissa de avaliar o desempenho de alunos ao fim de seu ensino médio, tendo uma reformulação no ano de 2009, com o objetivo de democratizar as chances ao acesso ao ensino superior público, além de ser esse um movimento de estopim para uma nova reestruturação de grades curriculares presentes no ensino médio brasileiro Andriola (2011).

Atualmente o foco do Enem continua sendo analisar o desempenho escolar e acadêmico ao final do ensino médio, possibilitando também a inscrição em programas do ministério da educação. INEP (2018). Para Travitzki (2013), desde o seu lançamento, o ENEM vem crescendo ano a ano, demonstrando sua relevância

nacional e aprovação social, agora fortemente consolidada e focando mais no raciocínio do que no conteúdo em si.

A Figura 3 mostra alguns dos momentos importantes das mudanças do ENEM de forma mais simples para se entender sua origem e suas principais mudanças:

Figura 3 - Marcos na História do Enem

ANO	Marcos na História do Enem
1998	Criação do ENEM.
2000	Algumas adotam o ENEM como parte do critério de seleção.
2001	Criação da taxa de Isenção
2004	Criação do ProUni
2005	ProUni é Anexado ao ENEM
2006	Médias ENEM por Escola
2007	Criação do Reuni
2009	Criação do SISU
2010	ENEM pode ser utilizado como certificação para ensino médio
2011	Enem obrigatório a candidatos do FIES

Fonte – Adaptado de Travitzki (2013).

6 PROCEDIMENTOS METODOLÓGICOS

A metodologia de pesquisa conforme Wazlawick (2009), consiste na sequência de etapas necessárias para demonstrar que o proposto foi alcançado, por tanto, se as etapas do método forem realizadas, os resultados obtidos devem ser conclusivos e satisfatórios. Além disso, Wazlawick também define algumas classificações de pesquisa, dentre elas podemos apontar o tipo de pesquisa que será decorrente neste projeto, caracterizada por ele como “apresentação de algo novo”, que consiste na apresentação de uma maneira diferente de se mostrar uma resolução de um problema pertinente a áreas atuais.

O método de pesquisa que será utilizado nesse trabalho, será o quantitativo, que segundo Wainer, J. (2007), esse método de pesquisa, se caracteriza pela correlação entre variáveis em um grande conjunto de dados, essa definição se aplica bem a quantidade massiva de dados que se pretende analisar nesta pesquisa.

Além disso, em relação aos objetivos, será uma pesquisa exploratória, pois a utilização de Mineração de Dados, tendo como alvo a base de dados do Enem, nunca foi realizada com os participantes do Enem no estado do Acre, se tornando assim um canal onde pode acontecer a descoberta de conhecimento importante para a tomada de decisões para o benefício da população local.

Na realização desta pesquisa, será utilizado a base de dados disponibilizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) referente ao ano de 2021, na qual consta as notas, atributos socioeconômicos de cada participante que realizou a prova nesse ano e outros dados referentes ao exame.

Para utilizar as tarefas de Mineração de Dados serão realizados os seguintes passos, associado às etapas apresentadas no processo de Knowledge Discovery in Databases (KDD):

1. **Seleção de um conjunto de dados:** será utilizada a base de dados do Enem do ano de 2021.

2. **Pré-processamento e transformação dos dados:** serão corrigidas as eventuais inconsistências e limpando a base de dados para remover apenas os atributos dispensáveis para realização desta pesquisa.
3. **Mineração de dados:** a tarefa de regressão será realizada através da aplicação de algoritmos e técnicas, visando gerar regras de regressão entre os fatores socioeconômicos e as notas dos participantes ao aplicá-la nos dados obtidos.
4. **Interpretação dos resultados:** os resultados obtidos serão organizados e mostrados, elencando quais atributos influenciam na nota do candidato.

7 ESBOÇO DOS CAPÍTULOS E SEÇÕES

6. Introdução

7. Problema

8. Objetivos

3.1 Objetivo Geral

3.2 Objetivos Especificos

9. Justificativa

10. Fundamentação Teórica

5.1 Knowledge Discovery in Databases

5.1.1 Regressão

5.1.2 Algoritmos de Regressão

5.1.2.1 Linear Regression

5.1.2.2 Instanced Based Learning

5.1.2.3 SMOReg

5.1.2.4 Randon Forest

5.1.2.5 M5P

5.2 Exame Nacional do Ensino Médio

6. Metodologia

7. Esboço do capítulo e seções

8. Cronograma

Referências

8 CRONOGRAMA

Nesta seção, será apresentado o cronograma de desenvolvimento desta pesquisa, listando as atividades necessárias para realização deste trabalho, tendo como base as etapas fundamentais para os processos *Knowledge Discovery in Databases* (KDD), descritas nos procedimentos metodológicos, seção 6.

Quadro 1 – Cronograma com distribuição de tarefas

Tarefas	MÊS I	MÊS II	MÊS III	MÊS IV
Seleção de um conjunto de dados acadêmicos	X			
Pré-processamento e transformação dos dados	X	X		
Mineração de dados e realização de experimentos			X	
Interpretação e avaliação				X
Escrita do TCC		X	X	X
Avaliação da banca - Defesa do TCC				X

Fonte: Elaboração própria.

REFERÊNCIAS

INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). **ENEM**. 2019. Disponível em: <<http://portal.inep.gov.br/web/guest/Enem>> Acesso em: 09 de out. 2022.

INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). **Nota Explicativa ENEM 2015 por Escola**. 2015. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2015/nota_explicativa_ene2015_por_escola.pdf> Acesso em: 19 de ago 2022.

JÚNIOR, Wilton Moreira de Santana. **Mineração em dados do ENEM para a predição do desempenho acadêmico no âmbito da Rede Federal de Educação Tecnológica**. Recife: Universidade Federal de Pernambuco, 2018. Disponível em: <[https://repositorio.ufpe.br/bitstream/123456789/29994/1/DISSERTA%c3%87%c3%83O %20Wilton%20Moreira%20de%20Santana%20J%c3%banior.pdf](https://repositorio.ufpe.br/bitstream/123456789/29994/1/DISSERTA%c3%87%c3%83O%20Wilton%20Moreira%20de%20Santana%20J%c3%banior.pdf)> Acesso em: 07 de dez. 2019.

SILVEIRA, Fernando Lang da; BARBOSA, Marcia Cristina Bernardes; SILVA, Roberto da. **Exame Nacional do Ensino Médio (ENEM): uma análise crítica**. [S.l.]. Revista Brasileira de Ensino de Física, v. 37, n. 1, 2015. Disponível em: <<http://www.scielo.br/pdf/rbef/v37n1/1806-1117-rbef-S1806-11173710001.pdf>> Acesso em: 18 de ago. 2019.

WAINER, J. (2007). **Métodos de pesquisa quantitativa e qualitativa para a ciência computação**. In T. KOWALTOWSKI e K. BREITMAN (Org.), Atualização em Informática da Sociedade Brasileira de Computação, pp. 221-262.

Mais de 45% da população do Acre está abaixo da linha da pobreza, diz estudo da FGV. Folha de Globo, Acre, 07 de jul de 2022. Disponível em: < <https://g1.globo.com/ac/acre/noticia/2022/07/02/mais-de-45percent-da-populacao-do-acre-esta-abaixo-da-linha-da-pobreza-diz-estudo-da-fgv.ghtml>>. Acesso em: 17 de ago. de 2022.

Walter Daelemans; Antal van den Bosch (2005). **Memory-Based Language Processing**. Cambridge University Press.

SHEVADE, S. K. et al. **Improvements to the smo algorithm for svm regression.** **IEEE transactions on neural networks**, IEEE, v. 11, n. 5, p. 1188–1193, 2000

Acesso dos mais pobres ao ensino superior é desafio a ser enfrentado na América Latina e Caribe. Ielsac Unesco, 20 de nov. de 2019. Disponível em: < <https://www.iesalc.unesco.org/2020/11/19/acesso-dos-mais-pobres-ao-ensino-superior-e-desafio-a-ser-enfrentado-na-america-latina-e-caribe/>>. Acesso em: 17 de out. de 2022.

FERRARI BRAVIN, G.; LEE, L.; RISSINO, S. das D. MINERAÇÃO DE DADOS EDUCACIONAIS NA BASE DE DADOS DO ENEM 2015. Brazilian Journal of Production Engineering, [S. l.], v. 5, n. 4, p. 186–201, 2019. DOI: 10.0001/%x. Disponível em: <https://periodicos.ufes.br/bjpe/article/view/27361>. Acesso em: 18 out. 2022.

ANDRIOLA, Wagner Bandeira. **Doze motivos favoráveis à adoção do Exame Nacional do Ensino Médio (ENEM) pelas Instituições Federais de Ensino Superior (IFES).** **Ensaio: avaliação e políticas públicas em educação**, v. 19, n. 70, 2011

GOLGHER, A. **Diálogos com o ensino médio 6: o estudante de ensino médio no Brasil analisado a partir de dados do INEP**. Belo Horizonte: UFMG/Cedeplar, 2018.

Almeida, Moreira. (2015). **A influência dos fatores socioeconômicos no desempenho Acadêmico dos estudantes de ensino médio integrado do Ifba/campus barreiras**. Disponível em: <
http://www.equidade.faced.ufba.br/sites/equidade.oe.faced.ufba.br/files/jucinara_pinto_robinson_tenorio_-_a_influencia_dos_fatores_socioeconomicos_no_desempenho.pdf > Acesso em: 20 de out. 2022.

FERRARI BRAVIN, G.; LEE, L.; RISSINO, S. das D. **MINERAÇÃO DE DADOS EDUCACIONAIS NA BASE DE DADOS DO ENEM 2015**. Brazilian Journal of Production Engineering, [S. l.], v. 5, n. 4, p. 186–201, 2019. DOI: 10.0001/%x. Disponível em: <https://periodicos.ufes.br/bjpe/article/view/27361>. Acesso em: 18 out. 2022.

Y. Wang and I. H. Witten. 1997. **Induction of model trees for predicting continuous classes**. In Poster papers of the 9th European Conference on Machine Learning. Springer.

Leite et all, Aruany. **Juventudes e a Pandemia do Coronavírus**: Copyright Conselho Nacional da Juventude (CONJUVE) 2020, Disponível em: <
<https://sinapse.gife.org.br/download/juventudes-e-a-pandemia-do-coronavirus> >
 Acesso em: 20 out. 2022.

MATOS, Guillermo; CHALMETA, Ricardo y COLTELL, Oscar. **Metodología para la Extracción del Conocimiento Empresarial a partir de los Datos**. Inf. tecnol. [online]. 2006, vol.17, n.2, pp.81-88. ISSN 0718-0764. <http://dx.doi.org/10.4067/S0718-07642006000200011>.

de Lima Júnior, M. L. **Previsão De Integradores E Tempo De Vida De Pull Requests**. Tese de Doutorado, Universidade Federal Fluminense, 2017.

EVANGELISTA, POLETO, Tobias. **Algoritmos e técnicas para a mineração de dados**. Editora: Fundação Educacional do Município de Assis, Instituto Municipal de Ensino Superior de Assis, São Paulo-SP, Brasil 2020. Disponível em: < <https://docplayer.com.br/168316101-Algoritmos-e-tecnicas-para-a-mineracao-de-dados.html> >

NOGARE, Diego; ZAVASCHI, Thiago. **Análise preditiva com azuremachine learning E R**. São Paulo: B2u, 2016.

Jiawei Han, Micheline Kamber, J. P. Data Mining. **Concepts and Techniques, 3rd Edition, 3rd edition ed. The Morgan Kaufmann Series in Data Management Systems**. Morgan Kaufmann, 2011.

Camilo, Silva, Cassio silva, Joao carlos. Mineração de Dados: **Conceitos, Tarefas, Métodos e Ferramentas**, Instituto de Informática UFG, 28 de fev 2012. Disponível em: < <https://ww2.inf.ufg.br/node/355> >

Breiman, L. **Random forests**. Machine Learning 45, 1 (2001), 5–32.

TRAVITZKI, Rodrigo. ENEM: **limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar**. 2013. tese de doutorado. universidade de São Paulo.