

SADENEM - PE: Um Sistema Avaliativo de Desempenho do ENEM para o Estado de Pernambuco

Fábio A. Freitas¹, Gabriel Oliveira², Henrique Farias³, Lucas Francisco⁴, Pedro A. X. Silva⁵

¹Departamento de Computação(DC)
Universidade Federal Rural de Pernambuco (UFRPE)
Recife – PE – Brasil

Abstract.

1. Introdução

O Exame Nacional do Ensino Médio (ENEM) é o sistema avaliativo que mede o desempenho dos estudantes no fim da educação básica [Silveira et al. 2015]. Esse exame durante 10 anos, entre 1998 (ano de criação) e 2008, tinha o objetivo único e exclusivo de avaliar as competências e habilidades dos alunos que já finalizaram o ensino médio . Porém, essa prova ganhou maior notoriedade quando boa parte das universidades federais do Brasil decidiram utilizar o ENEM como processo seletivo de ingresso ao ensino superior com a criação, em 2009, do Sistema de Seleção Unificada (SISU) que é uma plataforma para alocar candidatos às vagas mediante sua nota no ENEM.

EIXOS COGNITIVOS (comuns a todas as áreas de conhecimento)

- I. **Dominar linguagens (DL):** dominar a norma culta da Língua Portuguesa e fazer uso das linguagens matemática, artística e científica e das línguas espanhola e inglesa.
- II. **Compreender fenômenos (CF):** construir e aplicar conceitos das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos histórico-geográficos, da produção tecnológica e das manifestações artísticas.
- III. **Enfrentar situações-problema (SP):** selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para tomar decisões e enfrentar situações-problema.
- IV. **Construir argumentação (CA):** relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente.
- V. **Elaborar propostas (EP):** recorrer aos conhecimentos desenvolvidos na escola para elaboração de propostas de intervenção solidária na realidade, respeitando os valores humanos e considerando a diversidade sociocultural.

Figura 1. Eixos Cognitivos, retirado da matriz de referência do ENEM

Desde 2009, onde houve uma grande modificação comparado a anos anteriores, a prova do ENEM estrutura-se em 5 áreas diferentes que são: Linguagens e Códigos e

suas tecnologias (LC), Ciências Humanas e suas tecnologias (CH) , Ciências da Natureza e suas tecnologias (CN), Matemática e suas tecnologias e Redação. Dessas provas 4 delas são objetivas e apenas a prova de Redação é totalmente aberta, pois o candidato deverá escrever um texto dissertativo-argumentativo mediante a um tema escolhido pela banca examinadora. Nas provas objetivas, cada área possui um montante de 45 questões. Por fim, o aluno deverá resolver 180 questões e escrever uma redação. Todas as questões do ENEM são minuciosamente elaboradas baseadas na matriz de referência divulgada pelo ministério da educação do Brasil (MEC), nesse documento está exposta as competências e habilidades que cada questão possui, bem como os eixos cognitivos. Ao todo são 5 eixos cognitivos (que são comuns à todas as áreas de conhecimento), representado na figura 1, 30 competências e 120 habilidades divididas para as áreas objetivas. A prova de redação é avaliada em 5 competências, que são desde domínio pleno da língua portuguesa a elaborar uma proposta de intervenção.

Todos os anos milhões de estudantes realizam a prova do ENEM, em 2018 foram mais de 5.5 milhões de inscrições confirmadas [INEP 2018] e em 2019 mais de 5.1 milhões [MEC 2019]. Esses números são extremamente expressivos e gera diversas informações salvas em Microdados. Segundo o INEP: "Os microdados do Inep se constituem no menor nível de desagregação de dados recolhidos por pesquisas, avaliações e exames realizados", isto quer dizer que o INEP obtém os dados das provas de todos os estudantes e disponibilizam essa base de dados, porém esconde informações que podem identificar o candidato, como: Nome, CPF, Bairro, entre outras características que não podem ser reveladas pelo direito a segurança de dados.

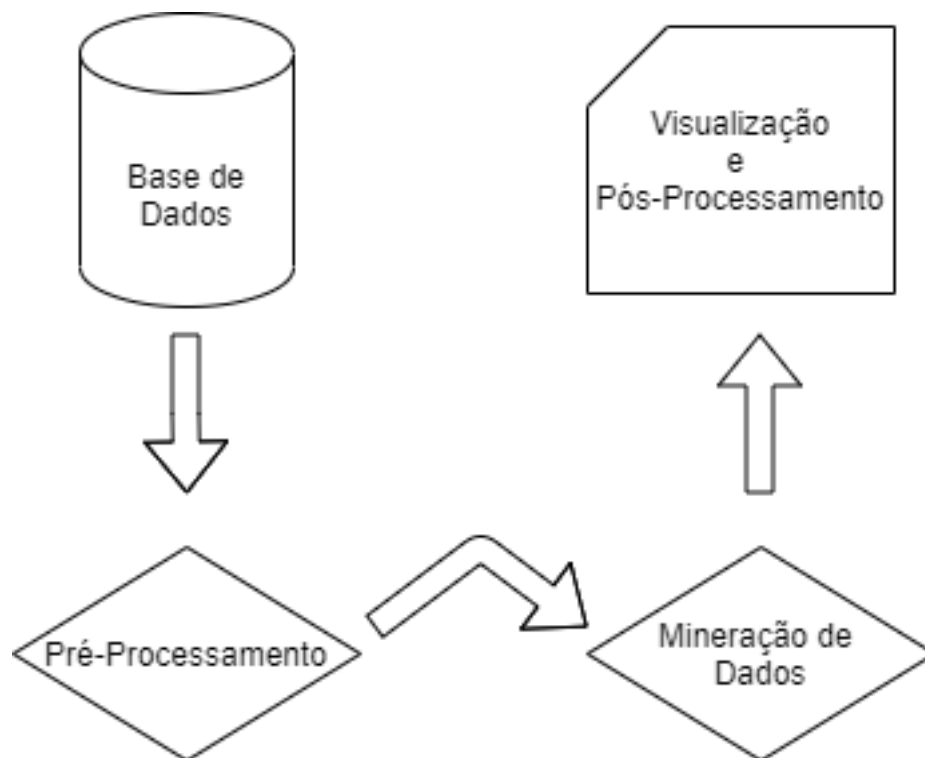


Figura 2. Pipeline Mineração de Dados

A figura 2 representa o arquétipo de qualquer projeto de pesquisa relacionado a

mineração de dados, claro que as etapas variam bastante entre projetos. Basicamente, foi apresentado nessa figura 4 passos Essências: Análise das Base dados, Pré-processamento, Mineração de Dados e Pós-processamento.

Esse relatório visa aplicar técnicas de análise e mineração de dados para extrair informações do microdados do Exame Nacional do Ensino Médio, nos anos de 2016 e 2017, com relação a Habilidades e Competências das provas objetivas no estado de Pernambuco. Além de criar uma ferramenta de visualização gráfica (*dashboard*) para análise dos resultados.

Este documento está organizado da seguinte forma: Na seção 2, Será sobre as bases de dados; na seção 3 são apresentadas técnicas de pré-processamento e limpeza de dados; na seção 4 são descritos os passos da mineração de dados e aprendizagem de máquina; a seção 5 sobre visualização e pós-processamento; finalmente, na seção 6 é mostrado a conclusão do desenvolvimento do projeto para disciplina de mineração de dados educacionais.

2. Pré-processamento

Essa certamente é uma das etapas mais importantes com relação a projetos de mineração de texto. Consiste na extração e preparação dos dados que serão utilizados nas análises posteriores. Neste trabalho, a metodologia utilizada para esta etapa foi a de aquisição de dados e em seguida sua limpeza.

2.1. Aquisição da Base de Dados

A aquisição dos dados relacionados ao anos de 2016 e 2017 do ENEM foi realizada pela plataforma de microdados do Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (INEP) [Inep 2020], que os disponibilizam de forma aberta.

Dentre os arquivos baixados podem ser encontradas as provas, gabaritos, e algumas planilhas que agrupam os dados referentes aos estudantes que prestaram a prova em todo o Brasil naquele ano. Dentre estas planilhas destacam-se a que armazena os dados de cada estudante que prestou o ENEM (Microdados); a que possui a relação entre as respostas, competências e habilidades de cada prova daquele ano (Itens da Prova); e os dicionários de dados, que explica os valores referentes a cada coluna destas dois arquivos mencionados.

Para preparação da fase de limpeza de dados, foram enviados os dicionários de dados, os microdados e os itens prova para uma pasta compartilhada no *Google Drive*, onde seria mais fácil lidar com eles. Em particular os microdados, uma vez que cada arquivo possui pelo menos 3 *gigabyte*.

2.2. Limpeza da Base de Dados

Esta etapa consistiu em compreender as informações contidas nos microdados e nos itens prova, afim de excluir aquelas que não agregariam à fase de mineração de dados. Após a leitura dos dicionários e alinhamento com a proposta desta trabalho, foram implementados algoritmos nas linguagens *Python* e *R* para a resolução desta tarefa, uma vez que seria impraticável realizá-la manualmente.

A limpeza dos *microdados* se deu pela exclusão dos dados que não seriam utilizados posteriormente. Tendo em vista que o objetivo deste trabalho é o de analisar o

desempenho do ENEM para o estado de pernambuco, as informações referentes aos demais estados e unidade federativa não seriam úteis, e portanto foram excluídas. Além disto, estes arquivo é composto por 166 colunas, que agrupam informações dos dados: socioeconômicos, redação, provas objetivas, local de aplicação da prova, certidão do ensino médio, pedido de recursos especializados, dados da escola e dados do participante. Dentre essas informações, foram selecionados os dados das provas objetivas e cidade de residência dos candidatos. Para a realização desta atividade foi escrito um algoritmo na linguagem *python*, e executado no *Google Colab*, devido ao custo computacional que seria processar um arquivo tão grande localmente. Como resultado, foram obtidos dois arquivos de 30 *megabytes* e com 30 colunas cada, para os microdados de 2016 e 2017.

Utilizando a linguagem *R*, foi implementado um algoritmo que retornava o número de acertos e a respectiva nota por prova de um candidato. Este processamento de dados foi importante para a implementação posterior da rede neural responsável pela predição da nota de um estudante.

3. Mineração de Dados

Teve o objetivo de avaliar o desempenho dos estudantes de pernambuco no ENEM nos anos de 2016 e 2017. Para isto, foi utilizado como base a matriz de competências e habilidades requeridas para a resolução e compreensão das questões das provas. Tendo em vista que existem 30 habilidades para cada uma das provas (CH, CN, LC e MT), a mineração de dados se deu pelo cálculo da acurácia de cada habilidade por município do estado e pela implementação de um modelo de rede neural para a predição das notas de um candidato.

Para o cálculo da acurácia, foi implementado um algoritmo na linguagem *R*, por ser uma ferramenta específica para análise de dados. São computadas a quantidade de questões e o respectivo número de acertos de um candidato para cada uma das 120 habilidades. Com divisão do total pelo número de acertos, temos a acurácia de um estudante para cada habilidade. São organizados os dados dos estudantes pelo seu respectivo município e em seguida executada esta computação. Como resultado, temos a acurácia de cada habilidade por município.

Para a implementação da rede neural, foi utilizada a linguagem *Python* juntamente com a biblioteca linguagem *Scikit Learn*, que já contém implementações de diversos algoritmos de redes neurais artificiais. Utilizando o arquivo que agrupa a relação entre a quantidade de questões acertadas e a respectiva nota de um candidato, foi montada a estrutura da rede neural com o algoritmo *random forest*. O objetivo desta modelo era o de prever a nota de cada prova dada uma lista com a quantidade de acertos por prova.

4. Visualização e Pós-processamento

Nessa etapa da pesquisa Utilizasse das mais variadas técnicas de visualização gráficas e análise de dados. Para isso, criamos um *Dashboard* interativo com todos os dados processamos nos passos anteriores. Basicamente, esse *dashboard*, desenvolvido na linguagem *Outsystem* apresenta o gráfico de taxa de acerto das habilidades nas provas objetivas do ENEM por município do estado de Pernambuco. A figura 3 representa a tela inicial da aplicação do *dashboard* interativo, nesse exemplo o gráfico representa a taxa de acerto

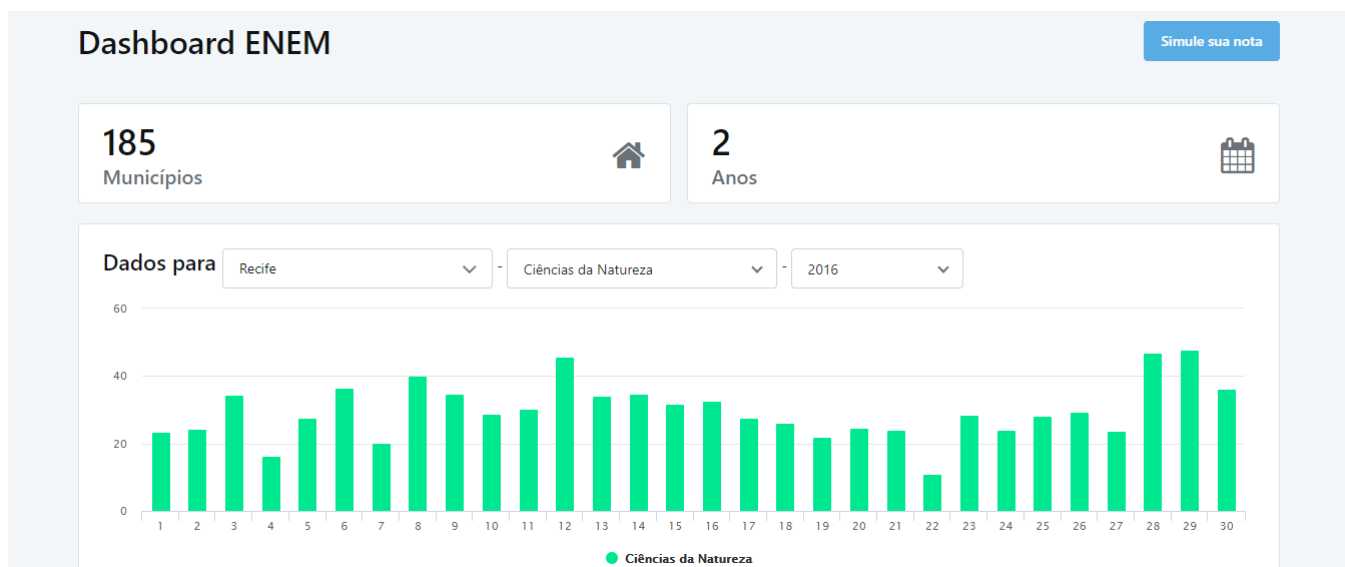
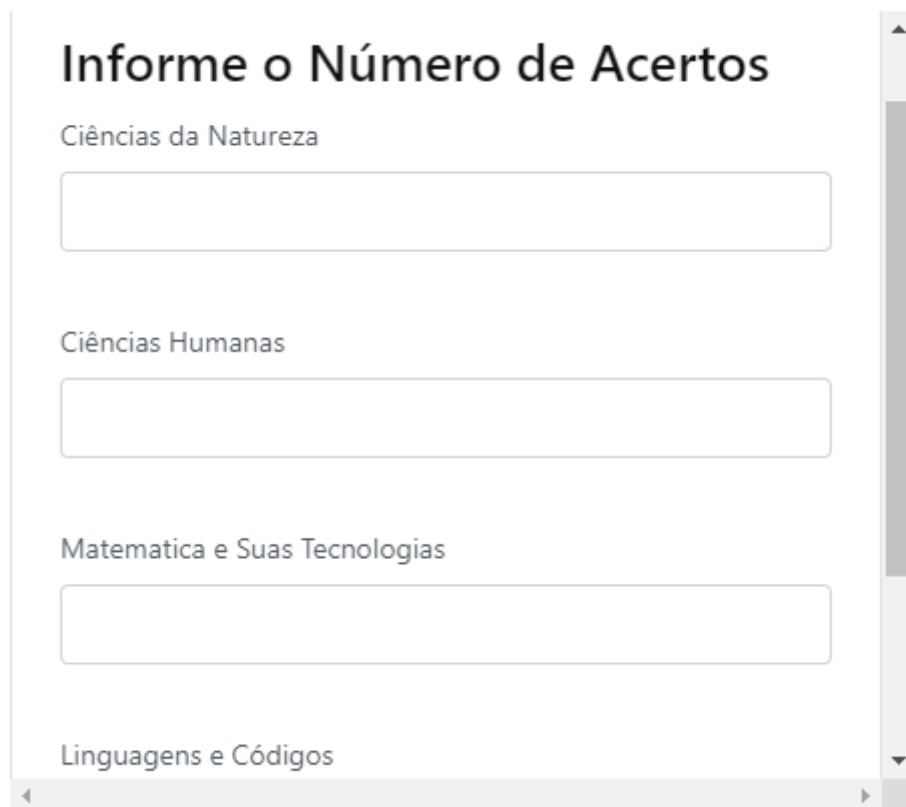


Figura 3. Tela inicial da Aplicação

das habilidades de Ciênciais da Natureza na cidade de Recife no ENEM de 2016. Pode-se mudar o ano, município e a prova objetiva que o gráfico muda completamente. Ao todo são 185 municípios, 4 provas objetivas e 2 anos de ENEM (2016 e 2017). A figura 4 representa a lista de habilidades da prova objetiva extraída da matriz de referência, assim, para efeito de análise, o usuário pode ver quais são as habilidades que tem maior taxa de acerto. Por fim, o *dashboard* possui um simulador de nota baseado na quantidade de acertos, representado na figura 5. Utilizou-se nessa etapa técnicas de aprendizagem de máquina para gerar as notas.

Habilidades Ciências da Natureza	
	Reconhecer características ou propriedades de fenômenos ondulatórios ou oscilatórios, relacionando-os a seus usos em diferentes contextos. Habilidade nº1
	Associar a solução de problemas de comunicação, transporte, saúde ou outro, com o correspondente desenvolvimento científico e tecnológico. Habilidade nº2
	Confrontar interpretações científicas com interpretações baseadas no senso comum, ao longo do tempo ou em diferentes culturas. Habilidade nº3
	Avaliar propostas de intervenção no ambiente, considerando a qualidade da vida humana ou medidas de conservação, recuperação ou utilizaç... Habilidade nº4
	Dimensionar circuitos ou dispositivos elétricos de uso cotidiano. Habilidade nº5

Figura 4. Tela de Habilidades da aplicação



Informe o Número de Acertos

Ciências da Natureza

Ciências Humanas

Matemática e Suas Tecnologias

Linguagens e Códigos

Figura 5. Simulador de Nota

5. Conclusão

Estes mecanismos de dados permite a análise das habilidades mais carentes de cada município, e portanto, serve como uma recomendação para os estudantes e instituições de identificarem as habilidades que precisam de um foco mais apurado. Em [UFRPE 2020a] está disponível todo o código fonte do projeto e em [UFRPE 2020b] o *dashboard* interativo.

Referências

- INEP (2018). Enem 2018 tem 6,7 milhões de inscritos. <https://bit.ly/34JVmP6>.
- Inep (2020). Microdados inep. <http://inep.gov.br/microdados>.
- MEC (2019). Mais de 3,9 milhões de candidatos participam do primeiro dia do enem 2019. <https://bit.ly/3eaku4I>.
- Silveira, F. L. d., Barbosa, M. C. B., and Silva, R. d. (2015). Exame nacional do ensino médio (enem): uma análise crítica. *Revista Brasileira de Ensino de Física*, 37(1):1101.
- UFRPE (2020a). Código fonte. <https://github.com/fabioafreitas/MineracaoDadosEducacionais>.
- UFRPE (2020b). Sadenem - pe: Dashboard. <https://henriquefps.outsystemscloud.com/ENEMAi/>.