



Ericky Dias, Felipe Freitas, Filipe Pereira, Guilherme Farçoni, Guilherme Sitton, Izaque Israel e Julia Issa

**IA na Educação: Transformando a Avaliação para
Combater a Formação Sem Base
A Revolução da IA na Avaliação Educacional:
Personalização para Alunos**

Brasil
2023

Resumo

Estudantes da TripleTen foram desafiados para um hackathon. Alunos da área de dados e web dev, se juntaram formando a organização Sons of Node, para cumprir um desafio. Trazer luz acerca do tema Inteligência Artificial. Neste relatório contém toda a análise desenvolvida pela área de dados, desde a obtenção de dados até seu dicionário, análise exploratória, data vis, e conclusões geradas. Nosso projeto se baseia na defesa do avanço da IA, principalmente no ramo da educação, onde analisamos e entendemos melhor sobre o problema da formação de alunos do ensino médio sem base. (explicar mais a análise)

Palavras-chave

Lista de tabelas, possíveis anexos e abreviaturas (caso seja necessário fazer), posso botar essa página agora no início ou no final do relatório.

Sumário

Introdução	5
Sessão 1: Materiais e Métodos.....	7
Dicionário dos Dados	7
Sessão 2: Análise exploratória dos dados.....	11
Sessão 3: Resultados e Conclusões.....	11
Referências	12

Introdução

- Explicar o que é hackathon

Hackathon é um projeto multidisciplinar na área de tecnologia, feito tanto por instituições da área da educação quanto por empresas para obter novas ideias e revolucionar o mercado. Suas únicas regras são trabalhar por um tempo limitado somente em projetos novos e criativos que não estejam relacionados ao dia a dia da empresa ou da instituição. Muitas coisas que já estão presentes no nosso cotidiano foram criadas em hackathon. Em 2013 o Facebook realizou hackathons para a melhoria de sua plataforma e consequentemente da empresa, funções como **álbuns compartilhados**, figurinhas de like e climas nos eventos foram uma das ideias implementadas deste hackathon. Até mesmo o botão “curtir” saíram destas maratonas. Seus benefícios são incontáveis.

- Objetivo/Problemática

O hackathon da TripleTen tem como objetivo apresentar dados pertinentes relacionados ao tema de inteligência artificial - seus prós e contras. A organização Sons of Node defende o avanço da IA (**explicar melhor a opinião do grupo**) e tem como objetivo aprofundar o tema em relação a educação, combatendo o problema de formação de alunos do ensino médio sem de fato ter o conhecimento esperado da Base Nacional Comum Curricular.

- Explicar o produto

Nosso projeto consiste em fazer uma análise exploratória de dados sobre o desempenho dos alunos em todo o Brasil, usando como métrica o Exame Nacional do Ensino Médio (ENEM), a partir disso tendo noção do desempenho por sexo, idade, raça e até por tipo de escola, se é pública ou particular. Desta forma, a área de web irá criar um site com a apresentação dos dados e com uma prova geral (desenvolvida por IA) sobre a matéria escolhida, gerando ao final dela um feedback personalizado ao aluno descrevendo pontos que ele precisa melhorar para ter o total entendimento da matéria, para desenvolver a educação brasileira como um todo, podendo assim ser facilmente escalável e tendo tanto escolas como o próprio governo como possíveis clientes

- Que dados você tem?

Como dito anteriormente, nós da Sons of Node utilizamos dados do ENEM, disponibilizado pelo próprio governo brasileiro. Tendo em si uma vasta gama de dados sabendo todo o background do aluno. Desde o sexo, idade, raça, qual a situação de conclusão do ensino médio, estado civil, se estudou em escola pública ou privada e etc. **(opinião da descrição)**

- O que você fez?

Começamos nossa análise com uma bela ‘limpeza’ dos dados, pois muita coisa encontrada era desnecessária na nossa análise, além de deixar o dataframe mais pesado e lento ao trabalhar com ele. Tinham colunas que havia mais de 70% de dados ausentes por exemplo, então resolvemos tirar elas para facilitar nossa análise já que não eram colunas do nosso ‘foco’. Após a limpeza começamos a famosa AED, análise exploratória de dados.

- **Quais são suas principais descobertas, resumidamente?**

- Como ler seu relatório/o que você tem no relatório? (irei resumir o conteúdo das sessões do relatório)

O relatório é dividido em sessões. A primeira sessão se chama Materiais e Métodos, mais focada em nossa metodologia de pesquisa, pensamentos, e esclarecimento dos dados de uma forma mais aprofundada. A segunda sessão se chama Análise exploratória dos dados já é sobre nossa análise em si, com nossas descobertas e data vis. Partindo então para a nossa terceira sessão onde compartilhamos nossos resultados e conclusões gerais.

Sessão 1: Materiais e Métodos

- Explicar a metodologia de pesquisa (pensamento e etc.)

O pensamento geral do grupo desde o princípio sempre foi defender o avanço da inteligência artificial. O campo escolhido foi a educação, o que é um campo muito amplo e cheio de oportunidades, mas precisávamos definir um foco, e esse foco foi o desempenho dos estudantes do ensino médio brasileiro. Nossa hipótese é definida por dar mais conhecimento e também direção aos jovens brasileiros para que tenham o entendimento mínimo necessário ao sair do ensino médio, combatendo a formação de alunos sem a devida base da Base Nacional Comum Curricular. Resultando então em um aumento do **índice de educação brasileira**, fazendo com que o país se desenvolva cada vez mais.

- Explicar de onde vem os dados

Para então medir o desempenho dos alunos no ensino médio, nada melhor do que o Enem, que também recolhe dados de pessoas já formadas. Nossos dados foram coletados pelo governo brasileiro e disponibilizados no próprio site². Contendo informações sobre o Enem realizado no ano de 2022.

- Explicar o que contêm nos dados

O dataframe original vem com mais de 3 milhões de linhas e muitas colunas, havia também o dicionário original contendo mais duas páginas sobre outros dados que não foram utilizados neste hackathon. Após a limpeza dos dados criamos nosso próprio dicionário.

Dicionário dos Dados

id: Número de inscrição

gp_idade: Faixa etária

1: Menor de 17 anos

2: 17 anos

3: 18 anos

4: 19 anos

5: 20 anos

6: 21 anos

7: 22 anos

8: 23 anos

9: 24 anos

- 10: 25 anos
- 11: Entre 26 e 30 anos
- 12: Entre 31 e 35 anos
- 13: Entre 36 e 40 anos
- 14: Entre 41 e 45 anos
- 15: Entre 46 e 50 anos
- 16: Entre 51 e 55 anos
- 17: Entre 56 e 60 anos
- 18: Entre 61 e 65 anos
- 19: Entre 66 e 70 anos
- 20: Maior de 70 anos

sexo: Sexo

M: Masculino

F: Feminino

est_civil: Estado Civil

0: Não informado

1: Solteiro(a)

2: Casado(a)/Mora com companheiro(a)

3: Divorciado(a)/Desquitado(a)/Separado(a)

4: Viúvo(a)

raca: Cor/raça

0: Não declarado

1: Branca

2: Preta

3: Parda

4: Amarela

5: Indígena

6: Não dispõe da informação

sit_ens_med: Situação de conclusão do Ensino Médio

1: Já concluí o Ensino Médio

2: Estou cursando e concluirei o Ensino Médio em 2022

3: Estou cursando e concluirei o Ensino Médio após 2022

4: Não concluí e não estou cursando o Ensino Médio

ano_conclusao: Ano de Conclusão do Ensino Médio

0: Não informado

1: 2021

2: 2020

3: 2019

4: 2018

5: 2017

6: 2016

7: 2015

8: 2014

9: 2013

10: 2012

11: 2011

12: 2010

13: 2009

14: 2008

15: 2007

16: Antes de 2007

tp_ensino: Tipo de instituição que concluiu ou concluirá o Ensino Médio

0: Não informado

1: Ensino Regular

2: Educação Especial - Modalidade Substitutiva

tp_escola: Tipo de escola do Ensino Médio

1: Não Respondeu

2: Pública

3: Privada

treineiro: Indica se o inscrito fez a prova com intuito de apenas treinar seus conhecimentos

1: Sim

0: Não

present_cien: Presença na prova objetiva de Ciências da Natureza

0: Faltou à prova

1: Presente na prova

2: Eliminado na prova

present_hum: Presença na prova objetiva de Ciências Humanas

0: Faltou à prova

1: Presente na prova

2: Eliminado na prova

present_port: Presença na prova objetiva de Linguagens e Códigos

0: Faltou à prova

1: Presente na prova

2: Eliminado na prova

present_mat: Presença na prova objetiva de Matemática

0: Faltou à prova

1: Presente na prova

2: Eliminado na prova

nota_cien: Nota em Ciências da Natureza

nota_hum: Nota em Ciências Humanas

nota_port: Nota em Linguagens e Códigos

nota_mat: Nota em Matemática

esp_ing: Tipo de prova de língua estrangeira

0: Inglês

1: Espanhol

nota_redac: Nota da Redação

- Possíveis insights que esses dados podem gerar

Descrever dados agrupados e a nota geral

Sessão 2: Análise exploratória dos dados

- Um mínimo de 50 palavras para cada visualização, explicando, dando um contexto mais amplo e tirando conclusões
- Uma conclusão de pelo menos 50 palavras, indicando quais outras explorações podem ser feitas no conjunto de dados
- Data vis
- Insights gerados

Sessão 3: Resultados e Conclusões

- Quais foram os resultados de fato
- Ideias aceitas e refutadas
- Conclusão Geral

- **PROXÍMOS PASSOS**

Referências

- 1- <https://www.techtudo.com.br/noticias/2014/01/sete-ideias-do-facebook-que-viraram-produtos-em-hackathons-de-2013.ghtml>
- 2- <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>