



Nome do projeto: APLICAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE LINGUAGEM NATURAL PARA AVALIAÇÃO AUTOMÁTICA DE QUESTÕES DISSERTATIVAS

Aluno: Lucas Mateus de Moraes

Orientador: Prof. Dr. Charles Henrique Porto Ferreira (cferreira@fei.edu.br)

## RESUMO

No contexto educacional, a correção de avaliações, principalmente as dissertativas, demanda um tempo considerável dos docentes. Este trabalho propõe um algoritmo para gerar avaliações automáticas de respostas a questões dissertativas, utilizando uma média ponderada de diferentes fatores extraídos do texto através de técnicas de Processamento de Linguagem Natural (NLP) e Inteligência Artificial (AI). Cada fator recebe um peso específico, determinado por regressão linear treinada com bases em diferentes idiomas, resultando em pesos distintos para cada versão do algoritmo.

Os resultados demonstraram acurácia de 63,43% para a versão em português (*dataset* de questões de Biologia), 83,58% para a versão em espanhol (*dataset* de questões de Literatura) e 81,63% para a versão em inglês (*dataset* de questões de Ciência da Computação). De forma geral, o algoritmo apresentou desempenho regular, gerando avaliações próximas às realizadas por docentes na maior parte dos casos. Os resultados indicam que, com aprofundamento, essa abordagem pode ser promissora para soluções relacionadas à problemática abordada neste trabalho.

## OBJETIVOS

O objetivo final do trabalho foi desenvolver e implementar o algoritmo para uma abordagem automatizada de avaliação para respostas dissertativas.

A proposta teve como alvo simplificar o processo de correção manual dessas respostas, promovendo uma análise dos fatores extraídos do texto e sua relação com as avaliações geradas. As metas planejadas no trabalho foram especificadas nos seguintes tópicos:

A) Desenvolver um algoritmo para gerar avaliações para respostas dissertativas com uma resposta padrão de um professor como referência.

B) Considerar os fatores de frequência de termos, distância de Levenshtein e similaridade semântica, através do uso dos modelos de linguagem BERT, BERTimbau e BETO, usados para Inglês, Português e Espanhol, respectivamente.

C) Validar a eficácia do algoritmo em testes, comparando-o com dados de avaliações já corrigidas.

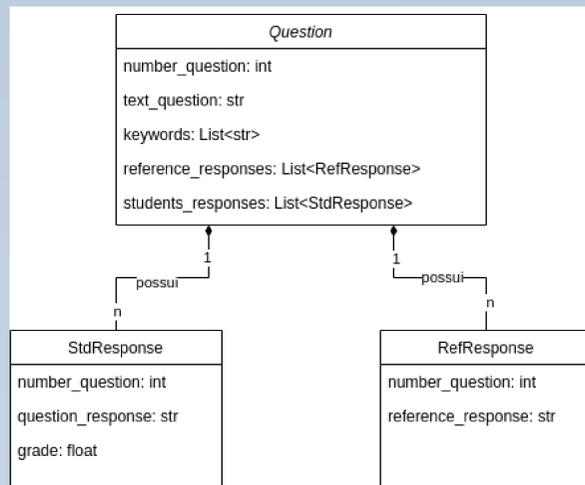
D) Como última meta, foi feito um protótipo para testes práticos, servindo como um exemplo de aplicação da técnica que foi proposta no trabalho.

## METODOLOGIA

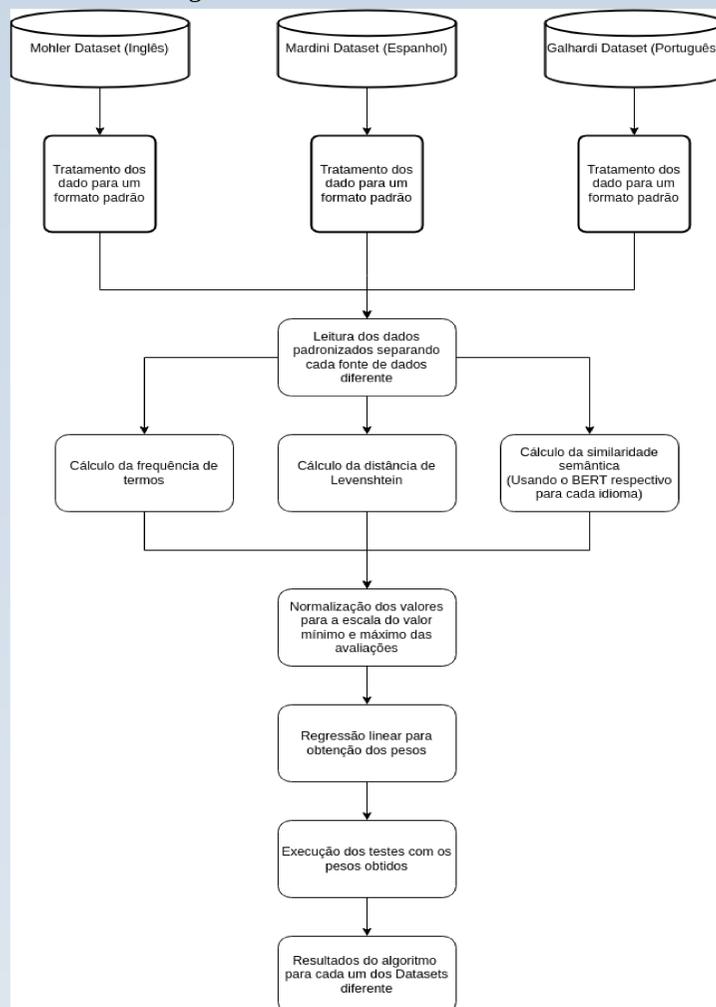
Na metodologia, de forma geral, foram retirados dos textos disponíveis nas bases de dados os três fatores que são usados pela regressão linear e posteriormente nos testes de geração de avaliações. A métrica utilizada na metodologia pode ser matematicamente descrita como uma média ponderada:

$$Métrica = \frac{fator1 \times peso_1 + fator2 \times peso_2 + fator3 \times peso_3}{\sum_{i=1}^3 peso_i}$$

A primeira etapa envolve a formatação dos dados oriundos de diferentes bases, convertendo-os em um modelo comum que pode ser representado no diagrama abaixo:



Após a formatação, os dados são processados pelo algoritmo para a extração dos fatores que influenciam a avaliação. Com os fatores normalizados, a próxima etapa é a aplicação de uma regressão linear. Após isso, utilizando os pesos determinados pela regressão linear, o algoritmo gera avaliações automáticas para um conjunto de dados separados. Finalmente, as avaliações geradas pelo algoritmo são comparadas com as notas originais fornecidas pelos professores na base de dados. Todo o processo da metodologia foi modelado no fluxograma abaixo:

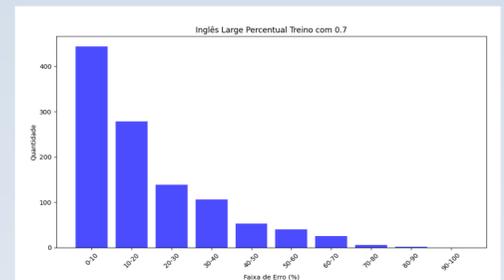
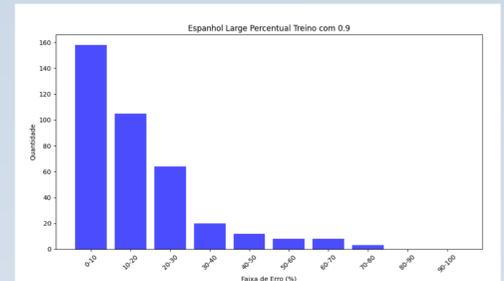
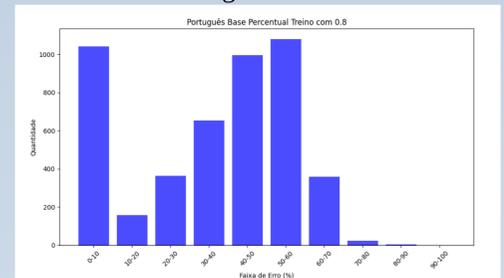


## RESULTADOS

Os resultados obtidos indicam uma melhoria no desempenho do algoritmo à medida que mais dados de treinamento são utilizados. No entanto, mesmo com uma quantidade substancial de dados de treinamento, ainda há espaço para melhorias.

Para o conjunto de dados em português, avaliando a acurácia média, o melhor resultado obtido foi de 63,43%, enquanto para o conjunto de dados em espanhol, foi de 83,58% e para o conjunto de dados em inglês foi de 81,63%.

As faixas de variação de erro percentual podem ser vistas nos gráficos abaixo:



## CONCLUSÃO

Pode-se concluir, deste trabalho, que, embora com limitações, os resultados obtidos apontam para uma direção promissora.

Uma explicação para os casos onde houverem erros em faixas percentuais altas é a falta de compreensão de mais fatores do texto que o algoritmo ainda não possui capacidade de avaliar profundamente.

Tendo isso em vista, como possibilidade de trabalhos futuros para melhoria dos resultados, algumas opções podem ser consideradas, como por exemplo:

A) Inclusão de mais fatores: Identificação de paráfrases e mais características da semântica do texto podem ajudar a alcançar uma métrica melhor.

B) Uso de outros modelos ou redes neurais: Comparar a regressão linear com outros modelos ou redes neurais, pode fornecer boas considerações sobre qual abordagem é mais eficaz para determinação dos pesos na métrica.