UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
BACHARELADO EM CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL

FINE-TUNING DE UM MODELO DISTILBERT PARA ANÁLISE DE SENTIMENTOS EM AVALIAÇÕES DE FILMES

Disciplina: Processamento de Linguagem Natural
Docente: Yuri de Almeida Malheiros Barbosa
Discente: Maxwel de Andrade Barbosa

João Pessoa – PB
2025

SUMÁRIO

LISTA DE FIGURAS

1 APRESENTAÇÃO DO PROBLEMA

A classificação de sentimentos em avaliações de filmes é uma tarefa essencial em Processamento de Linguagem Natural (PLN) e aprendizado de máquina. Milhares de avaliações são postadas todos os dias em plataformas como Amazon, IMDb e Rotten Tomatoes, compreender de maneira otimizada se essas avaliações são positivas ou negativas pode auxiliar estúdios e plataformas de streaming na análise de feedbacks e no direcionamento de campanhas publicitárias.

Entretanto, classificar sentimentos em textos é desafiador, pois envolve interpretar sarcasmo, subjetividade e múltiplos estilos de escrita. Este projeto visa construir o pipeline de preparação de dados necessário para treinar um modelo que tenta prever o sentimento de uma avaliação de filme.

O trabalho atual visa encarar esse problema utilizando métodos de PLN para classificar os comentários de usuários postados na plataforma IMDb. A partir do fine-tuning de um modelo DISTILBert.

2 OBJETIVOS

Objetivo Geral

Realizar o fine-tuning do modelo baseado na arquitetura transformer, DistilBERT, para classificar avaliações de filmes em categorias de sentimento (positivo ou negativo), com base nas características textuais obtidas a partir da base de dados.

Objetivos Específicos

- Realizar o pré-processamento do dataset de avaliações de filmes, preparando os dados para entrada no modelo;
- Implementar um pipeline de tokenização e vetorização dos textos;
- Treinar a rede neural para tarefas de classificação binária, usando uma arquitetura apropriada para o problema;
- Avaliar o desempenho do modelo utilizando métricas como acurácia, perda de validação e matriz de confusão.

3 DADOS UTILIZADOS E PRÉ-PROCESSAMENTO DOS DADOS

O conjunto de dados usado neste projeto é o "IMDB Dataset of 50K Movie Reviews", disponível no Kaggle. Este dataset contém 50.000 avaliações de filmes extraídas do site IMDb, cada uma rotulada como "positiva" ou "negativa". As avaliações são divididas igualmente entre os dois sentimentos, com 25.000 exemplos para ambas as classes, e foram originalmente coletadas para pesquisas em Processamento de Linguagem Natural (PLN) e aprendizado de máquina. O dataset é amplamente utilizado como benchmark em tarefas de análise de sentimentos, devido à sua qualidade e equilíbrio entre as classes.

Para preparar os dados para o modelo foi realizado o pré-processamento. Inicialmente, as avaliações foram convertidas para letras minúsculas para garantir uniformidade. Em seguida, foram removidas tags HTML e caracteres especiais, como pontuações e símbolos, utilizando expressões regulares. Esse processo de limpeza visa eliminar ruídos que possam interferir na análise semântica dos textos. Após a limpeza, os textos foram tokenizados e vetorizados, transformando cada token em um identificador numérico único. Para garantir que todas as sequências tivessem o mesmo comprimento, foi aplicada a técnica de padding, preenchendo as sequências menores com zeros. Essas etapas são fundamentais para que os dados possam ser adequadamente utilizados por modelos de redes neurais, que requerem entradas numéricas de tamanho consistente.

4 METODOLOGIA

Este projeto emprega o fine-tuning do modelo DistilBERT que é uma versão compacta e eficiente do BERT para a classificação automática das resenhas de filmes presentes no conjunto IMDb, com o objetivo de reconhecer padrões de polaridade (positiva ou negativa) nos textos. A metodologia aplicada segue um fluxo completo que vai desde a limpeza e tokenização dos dados, passa pelo ajuste fino do transformador pré-treinado em um conjunto reduzido de treino/validação, até a avaliação final do modelo por meio de métricas como acurácia e matriz de confusão.

4.1 Técnica utilizada

O DistilBERT foi selecionado por ser um dos modelos com menor quantidade de parâmetros (66M) se comparado à outros modelos como Mistral 7B, Llama 2 7B dado o poder computacional limitado disponível e também devido à sua capacidade de capturar dependências semânticas de longa distância por meio do mecanismo de auto-atenção bidirecional dos transformers, mantendo 97 % do desempenho do BERT original com cerca de metade dos parâmetros e, portanto, oferecendo um treinamento mais rápido e computacionalmente econômico (em termos de consumo de GPU). No notebook, o modelo DistilBertForSequenceClassification pré-treinado foi ajustado sobre o corpus IMDb, sendo configurado para identificar a polaridade das resenhas em duas classes: negativa (0) ou positiva (1). A etapa de tokenização utilizou o DistilBertTokenizer, que converte cada texto em pares input_ids / attention_mask com truncamento, preservando a bidirecionalidade do contexto. Sobre a representação gerada pelo encoder que é composto por seis blocos transformer com camadas de atenção multi-cabeças e feed-forward posicionais foi acoplado um perceptron denso de saída com ativação soft-max, responsável por mapear o vetor contextual para as probabilidades das duas classes. Durante o treinamento, foi utilizado o otimizador Adam com weight decay 0.01, taxa de aprendizado inicial de 2×10^{-5} garantindo atualização estável dos pesos do transformer. Foram executadas três épocas com batch size 32, gradientes reiniciados a cada atualização. Dessa forma, o fine-tuning do DistilBERT integrou o conhecimento linguístico amplo adquirido na fase pré-treino com as nuances domínios-específicas das resenhas de filmes, resultando em um classificador eficiente.

4.2 Experimento para avaliar a técnica utilizada

Para medir o impacto do fine-tuning no DistilBERT, primeiro reduzimos a quantidade de amostras de treinamento. Inicialmente eram 50 mil exemplos que foram reduzidos a um subconjunto estratificado de 10 mil, sendo 50% de cada classe objetivando diminuir o tempo de treinamento sem comprometer a representatividade. Esse recorte foi, então, dividido aleatoriamente em 80 % para treino (8000 exemplos) e 20 % para validação-teste (2000 exemplos), preservando a proporção de classes em ambos os conjuntos. Antes de qualquer ajuste, passamos o corpus de validação pelo DistilBERT pré-treinado obtendo um baseline de 0.5 de acurácia, valor esse que serviu como referência para as melhorias subsequentes. Em seguida, realizamos o

fine-tuning por 3 épocas e durante o treinamento, o desempenho do modelo foi monitorado através da acurácia e da função de perda. A técnica de dropout interno do próprio DistilBERT (p = 0,1), aliada à estratificação dos dados, assegurou melhor generalização. Após o ajuste, o modelo atingiu acurácia de 0.89 e loss 0.26 no conjunto de teste, uma melhoria considerável em comparação com o baseline. Por fim, geramos a matriz de confusão e inspecionamos exemplos em que a rede ainda confunde ironia ou negação, possibilitando a identificação de futuras linhas de aprimoramento, como aumento do corpus ou fine-tuning por mais épocas com *early-stopping*. Encontra-se abaixo a matriz de confusão para o teste do modelo antes e após o treinamento:

Figura 1: Matriz de confusão antes do fine-tuning

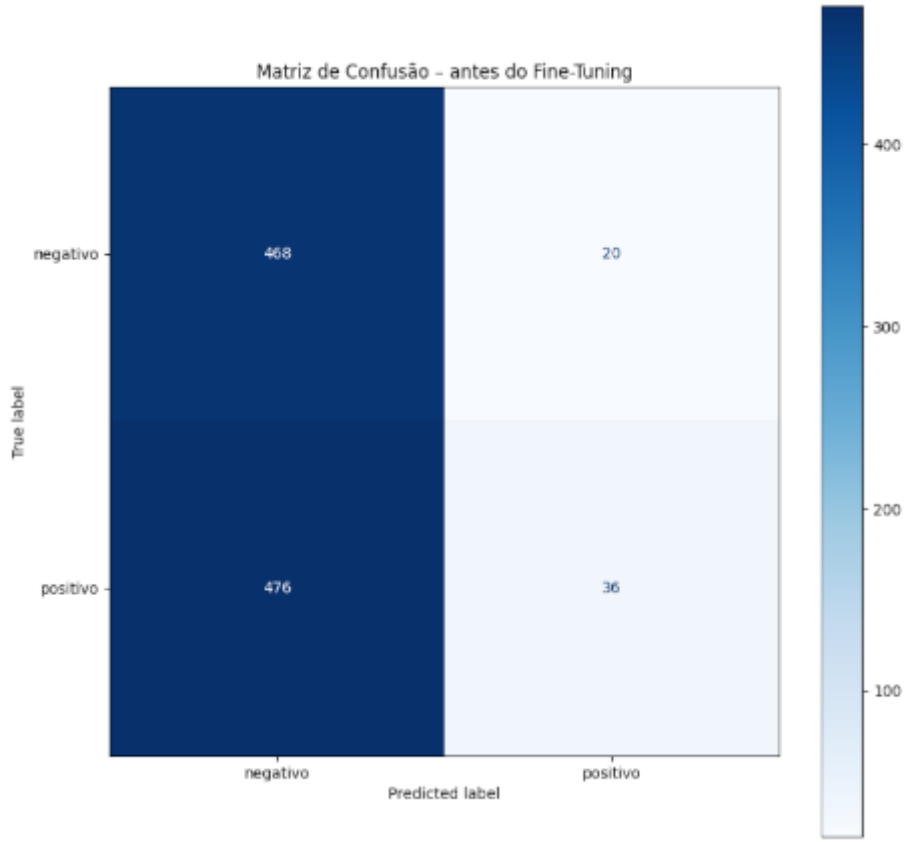
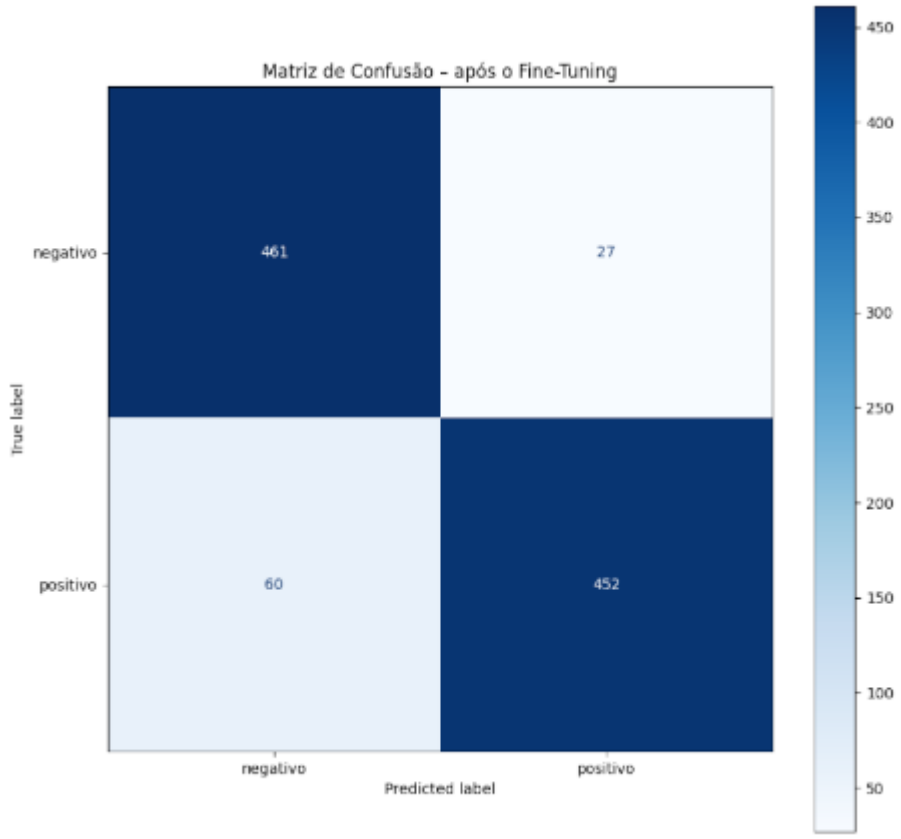


Figura 2: Matriz de confusão após o fine-tuning



5 RESULTADOS

O modelo DistilBERT após o treinamento alcançou 0.89 de acurácia no conjunto de teste, mostrando boa capacidade de identificar a polaridade das resenhas. Ainda assim, o modelo confunde alguns textos com linguagem coloquial, ironia ou entusiasmo exagerado, indicando que exemplos desse tipo são sub-representados no treino. Esses erros sugerem reforçar o corpus com amostras que tragam gírias, sarcasmo e expressões informais a fim de melhorar a robustez do classificador.

Figura 3: Perda do treino e da validação ao longo das épocas

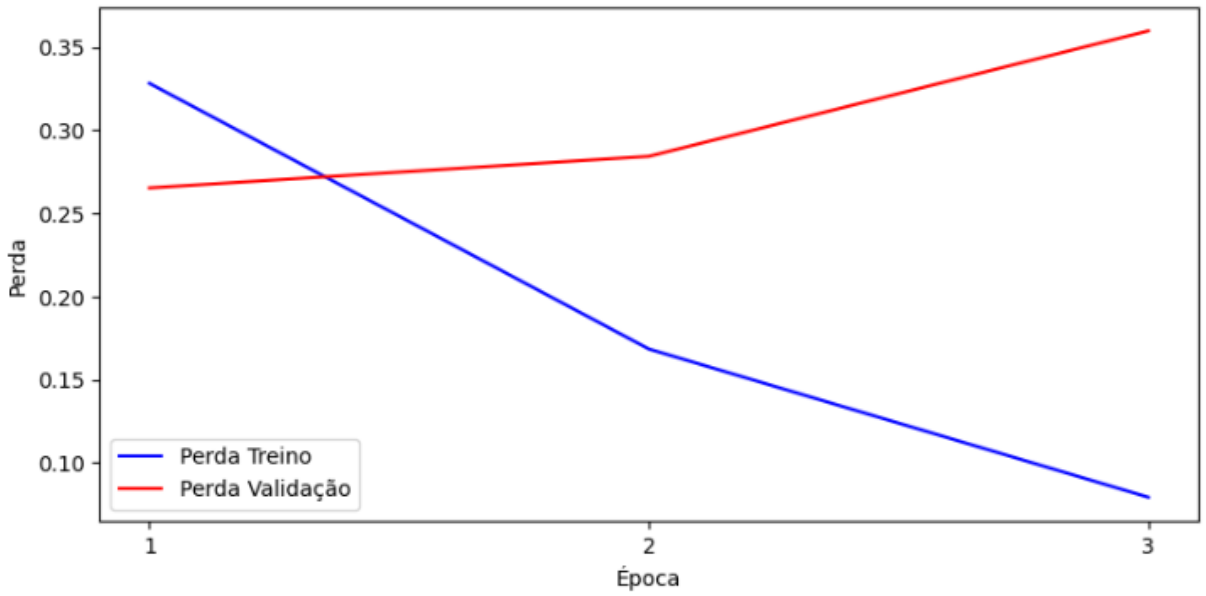
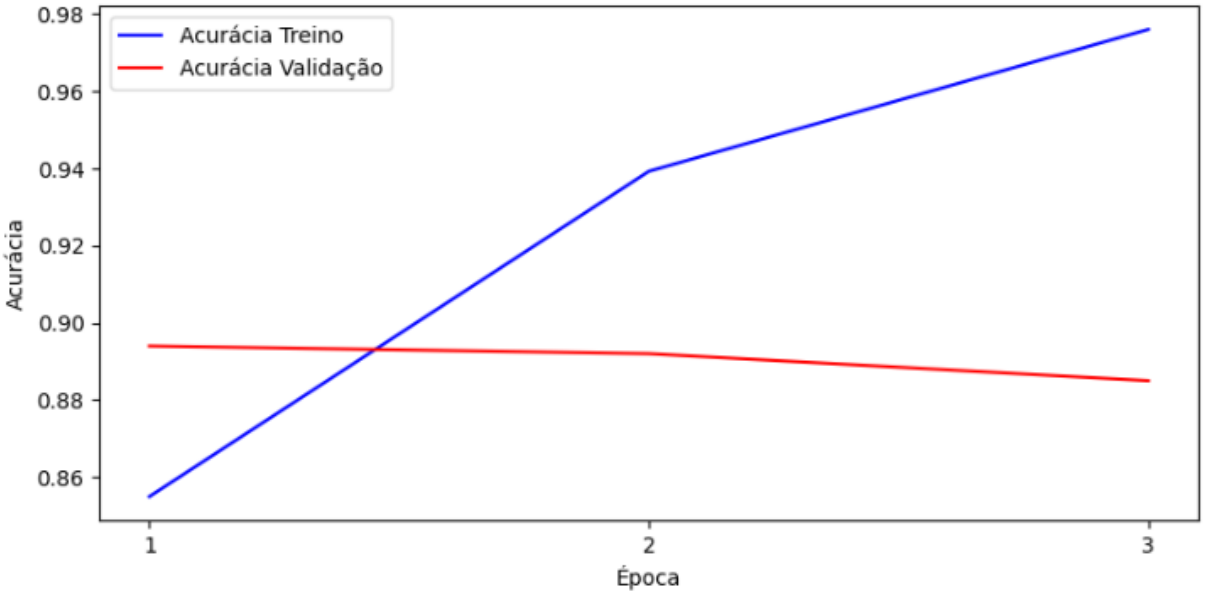


Figura 4: Acurácia do treino e da validação ao longo das épocas



6 REFERÊNCIAS

HUGGING FACE. Pretrained models. 2020. Disponível em: https://huggingface.co/transformers/v2.9.1/pretrained_models.html. Acesso em: 28 abr. 2025.

HUGGING FACE. DistilBERT: model documentation. 2020. Disponível em: https://huggingface.co/transformers/v2.9.1/model_doc/distilbert.html. Acesso em: 28 abr. 2025.