ANALYZING USER FEEDBACK FOR ERROR DETECTION

-ARAVIND MORIGADI
SANDEEP SANGVI

Table Of Content

Introduction

This project simplifies how developers handle user feedback and log data by using NLP technique. Our goal is to speed up error detection and resolution, helping developers fix problems faster and more effectively which speeds up error detection and resolution, improving app quality and user satisfaction.

Research

Evaluate both traditional machine learning models and state-of-the-art transformer models like DistilBERT.

Motivation

Sentiment analysis helps in understanding user feedback, enabling businesses to refine services.

Problem Statement

How do transformer-based models like DistilBERT compare with traditional ML models in sentiment classification?

Literature Review

- The literature highlights key advancements in software engineering using NLP and machine learning techniques. Guzman et al. (2014) explored sentiment analysis to interpret user emotions in bug reports and reviews, enhancing software quality, while Antoniol et al. (2008) introduced automated bug report categorization, enabling faster issue resolution through NLP techniques. He et al. (2018) proposed hybrid approaches combining statistical methods and machine learning for anomaly detection in software monitoring, ensuring efficient issue identification.
- Traditional models like Logistic Regression and SVM, paired with word embeddings such as GloVe, provide strong baseline performance, but recent developments in transformer-based models like DistilBERT, fine-tuned on domain-specific data using HuggingFace tools, significantly improve context understanding and accuracy in software-specific tasks.

Corpus Acquisition & Approach

01

Corpus Acquisition

The project uses Kaggle's "Reviews for Classification" dataset with 26,966 labeled records, categorizing sentiments as positive, negative, or neutral based on customer reviews.

02

Data Collection

The dataset comprises 26,966 customer reviews sourced from Kaggle, labeled for sentiment analysis into positive, negative, or neutral categories, providing a robust foundation for training and evaluation.

03

Data Preprocessing

Preprocessing steps include lowercasing, tokenization, removing stopwords, and feature extraction using GloVe embeddings and TF-IDF.

04

NLP Techniques& LLM Models

Traditional models like Logistic Regression, Random Forest, SVM, and XGBoost are combined with transformer-based techniques, such as fine-tuning DistilBERT for superior contextual understanding. Evaluation metrics include Accuracy, Precision, Recall, and F1-Score.

Implementation

1. Libraries Used

Python: Pandas, sklearn, HuggingFace's transformers.
Pretrained Models: GloVe for embeddings and DistilBERT for transformer-based classification.

Key Steps for DistilBERT

1. Fine-tuned DistilBERT on the dataset for 26966 samples (for computational efficiency).
2. Tokenized text with HuggingFace tokenizer
3. Used DistilBERT's classification head for predictions.

2. Github code uploaded link

Presentation is a public speaking activity that aims to convey a message or information. The things conveyed can be in the form of ideas, ideas,

Experimental Setup

01

Test

Compare traditional models with DistilBERT in predicting sentiments (positive, neutral, and negative).

02

Experiment

Measure performance on the same dataset split for both traditional ML models and DistilBERT.

03

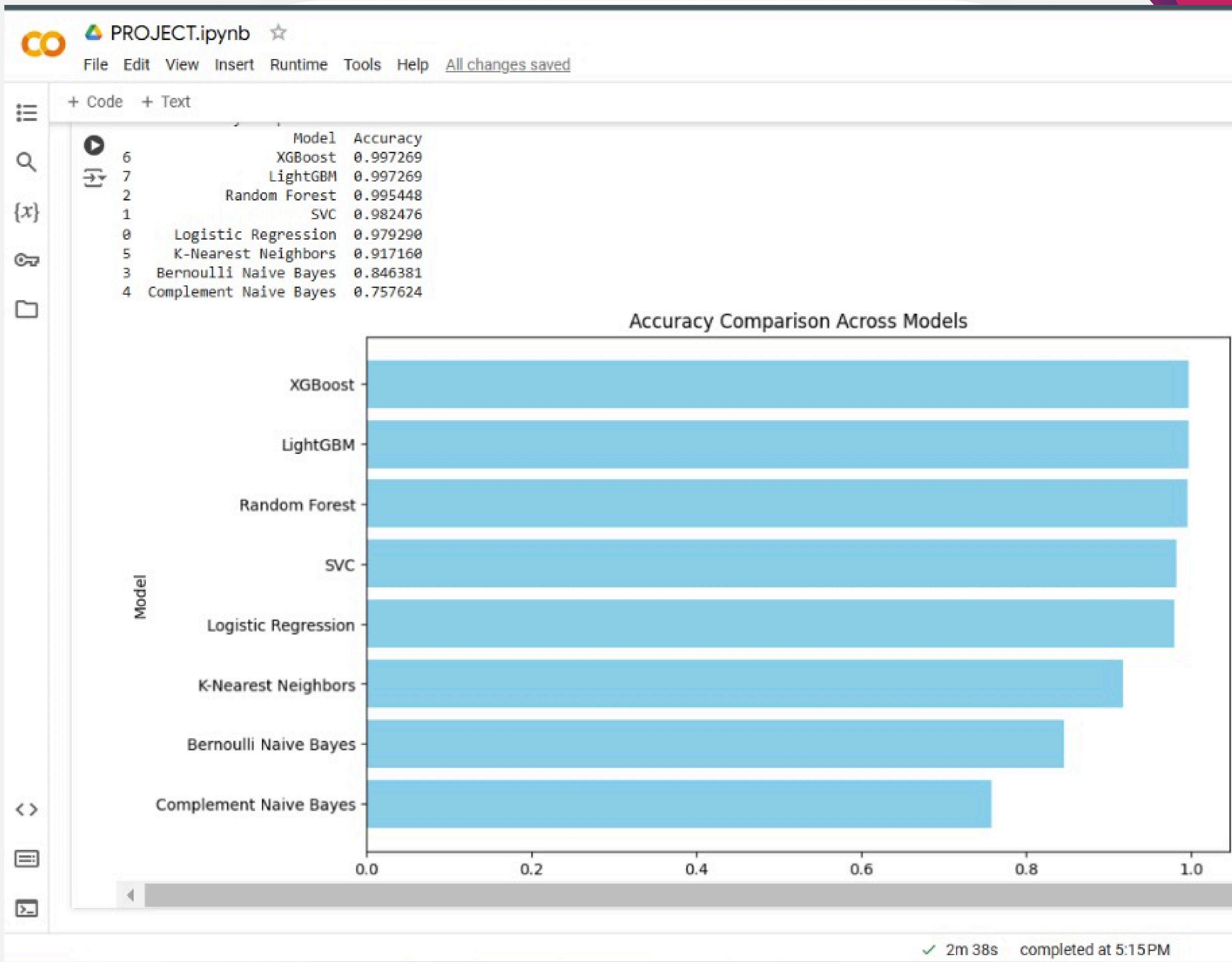
Procedure

For DistilBERT: Used fine-tuning on a subset of data (26966 records) due to resource constraints.
For ML models: Trained using extracted features (GloVe embeddings).

Results:

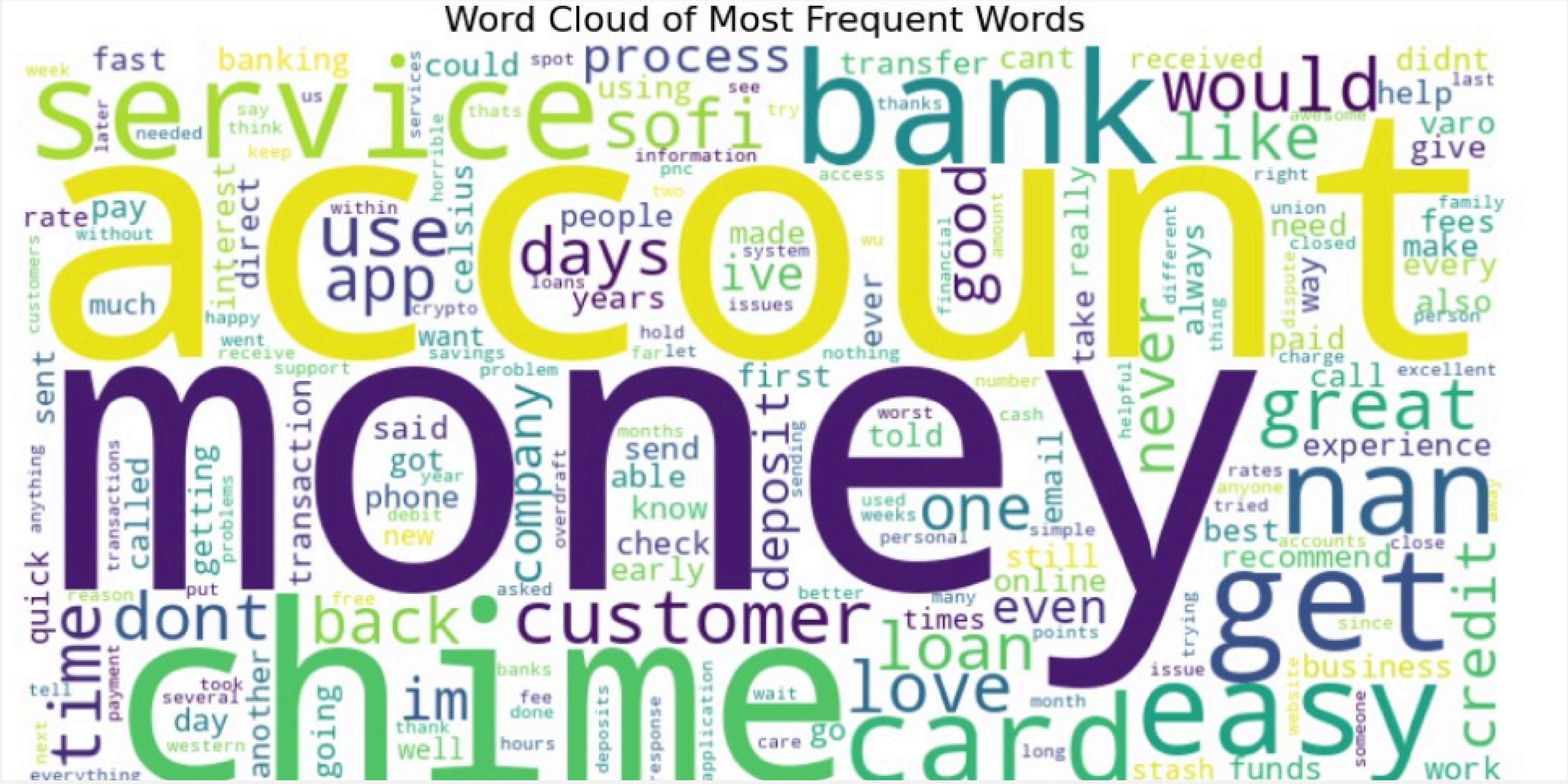
1. Machine Learning Approaches

Presentation is a public speaking activity that aims to convey a message or information. The things conveyed can be in the form of ideas, ideas, programs, products, or services and are usually carried out in a discussion or forum.



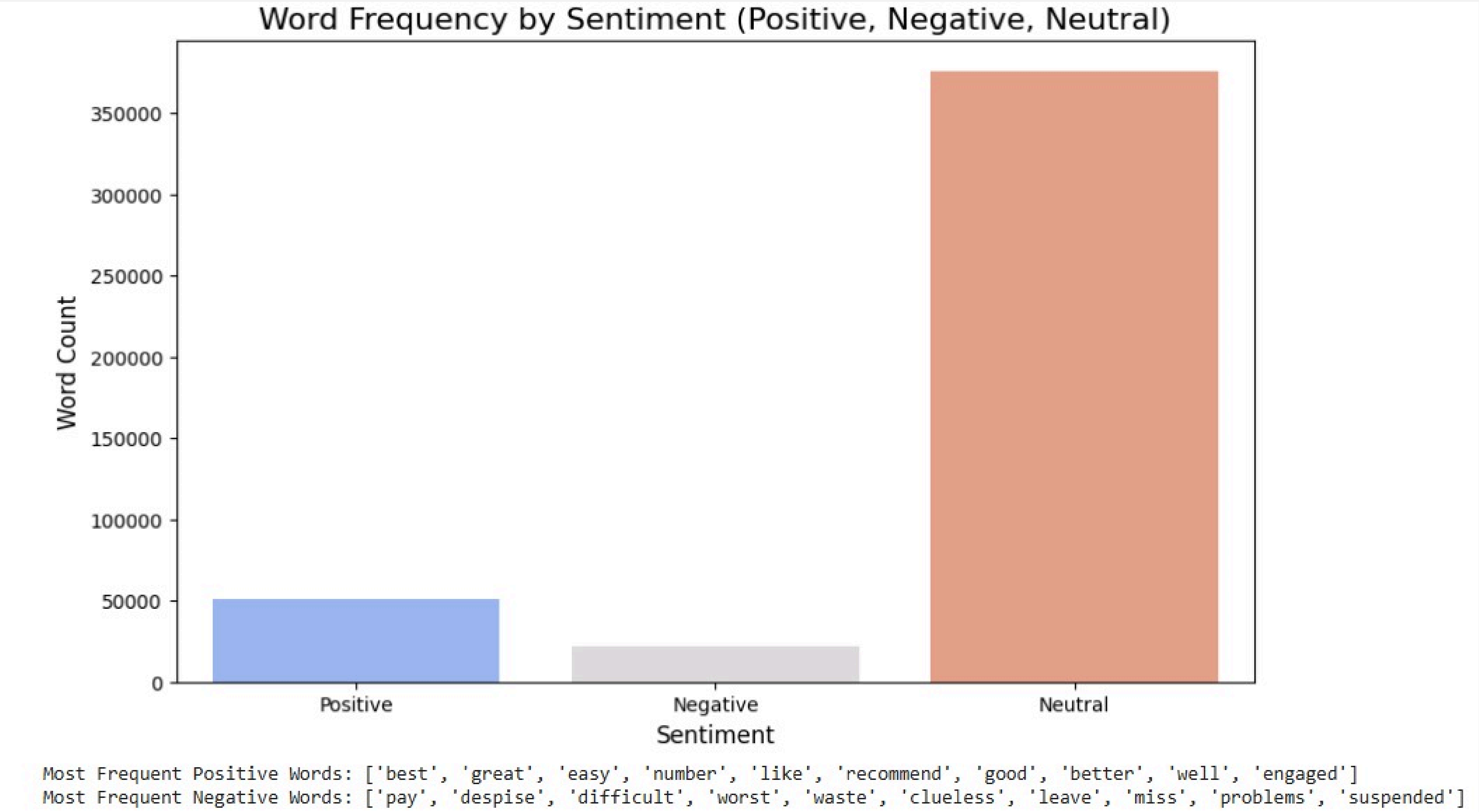
2. Word Cloud

Presentation is a public speaking activity that aims to convey a message or information. The things conveyed can be in the form of ideas, ideas, programs, products, or services and are usually carried out in a discussion or forum.



3. Sentiment Analysis

Presentation is a public speaking activity that aims to convey a message or information. The things conveyed can be in the form of ideas, ideas, programs, products, or services and are usually carried out in a discussion or forum.



Discussion & Analysis of Results

Insights

- DistilBERT captures nuanced meanings, making it robust for handling imbalanced datasets and complex texts.
- Traditional ML models are cost-effective and interpretable but lack deep contextual understanding.

Challenges

- Resource constraints during DistilBERT fine-tuning required dataset size reduction.
- Overhead in deploying transformer models compared to traditional ones.

Future Work:

- Apply fully trained BERT or domain-specific transformers (e.g., FinBERT for financial sentiment).
- Real-time sentiment analysis integration in business tools.
- Optimize computational resource usage to fine-tune on larger datasets.

Meet Our Team

Aravind Morigadi

Done with the Data preprocessing and integration of GloVe embeddings.

Sandeep Sangvi

Done with the Implementation and evaluation of ML models & Fine-tuning DistilBERT and result analysis.

THANK YOU

FOR YOUR ATTENTION& TIME