Performance Analysis Report

Project Overview

This project involves classifying online feedback based on six types of offensive content: toxic, abusive, vulgar, menace, offense, and bigotry. Each label is binary and multi-label classification is required as one comment can belong to multiple categories.

Exploratory Data Analysis (EDA)

Several insights were derived during EDA:

- Distribution of each offensive label was visualized to identify class imbalance.

- Sentence length and word distributions were plotted.

- Word clouds were generated to observe common terms in offensive comments.

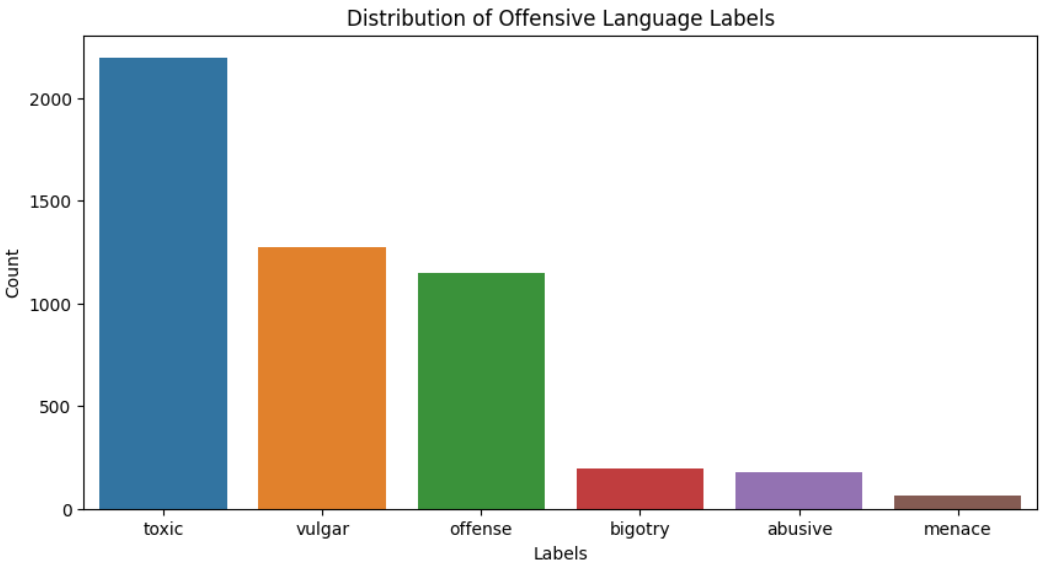

Figure 1: Visualization from EDA

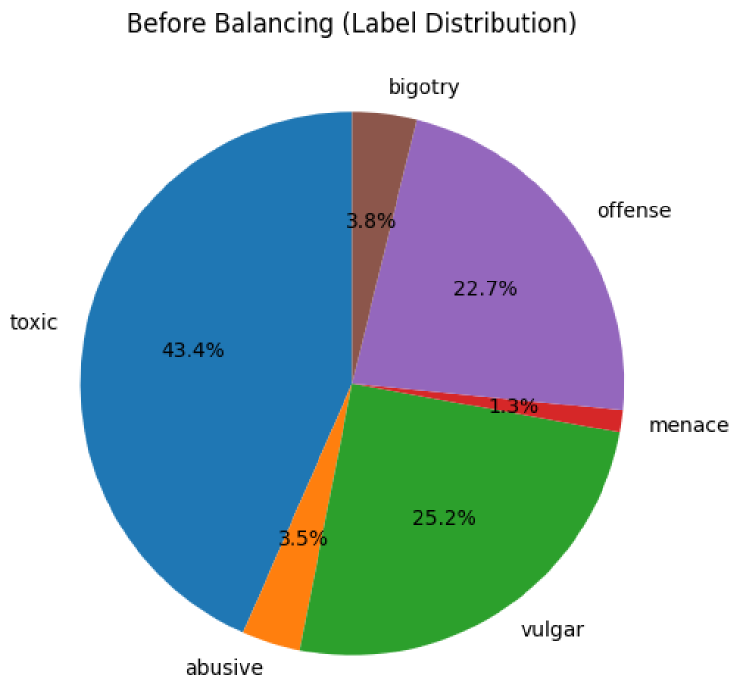

Figure 2: Visualization from EDA

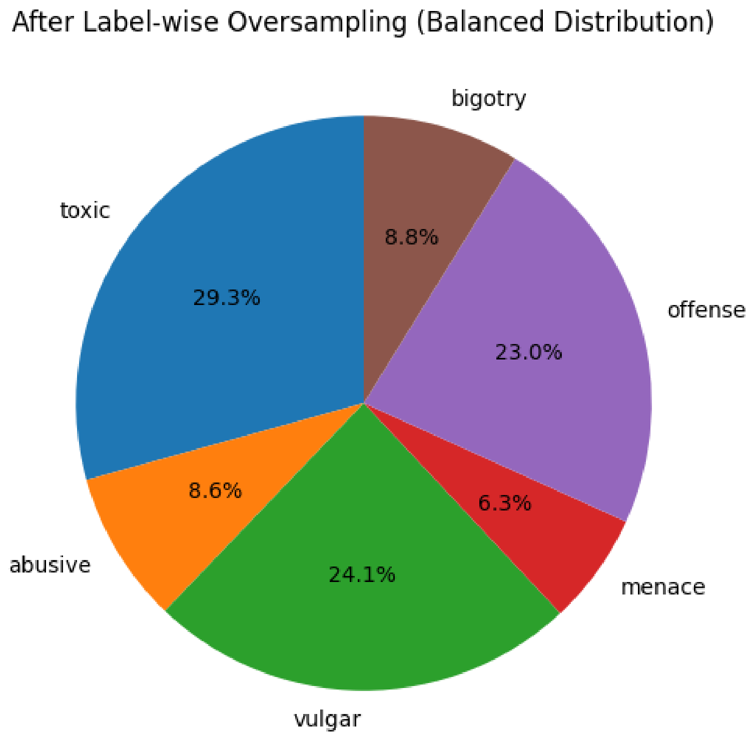

Figure 3: Visualization from EDA

Text Preprocessing

The preprocessing pipeline included:

- Lowercasing all text

- Removing punctuation and special characters

- Removing stopwords using NLTK

- Lemmatization with NLTK's WordNetLemmatizer

- Tokenization using `nltk.word_tokenize`

Model 1: Logistic Regression & LSTM

Two models were implemented in the first notebook:

- Logistic Regression using TF-IDF vectors as features.
- LSTM network using Keras Embedding layer with sequence padding and one LSTM layer.

Logistic Regression served as a fast baseline, while LSTM captured sequential patterns in text.

Model 2: Transformer (BERT)

The second notebook used the `bert-base-uncased` model from HuggingFace Transformers.

- Tokenization using `BertTokenizer`
- Used `BertForSequenceClassification` with sigmoid for multi-label classification
- Fine-tuned using AdamW optimizer and early stopping

Model Evaluation

The models were evaluated using:

- Accuracy

- Precision, Recall, F1-score (micro & macro)

- ROC-AUC scores per label

- Confusion matrix visualizations

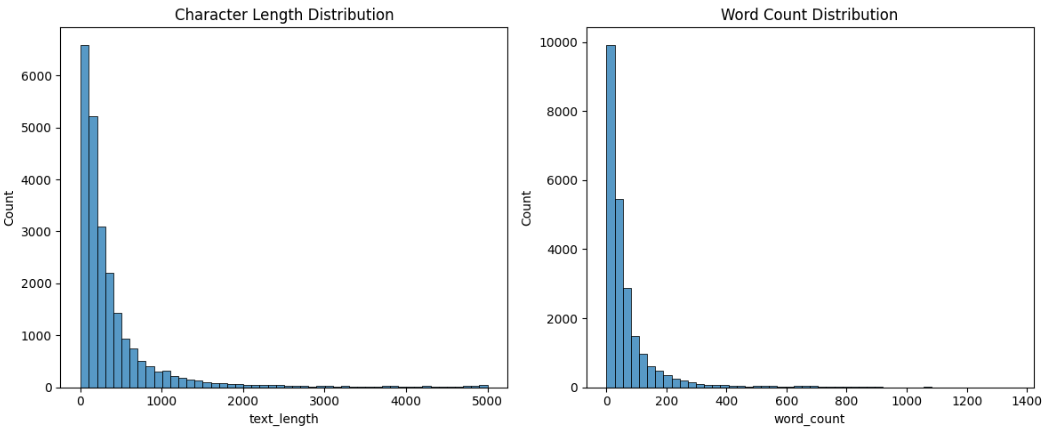

Figure 4: Evaluation plot

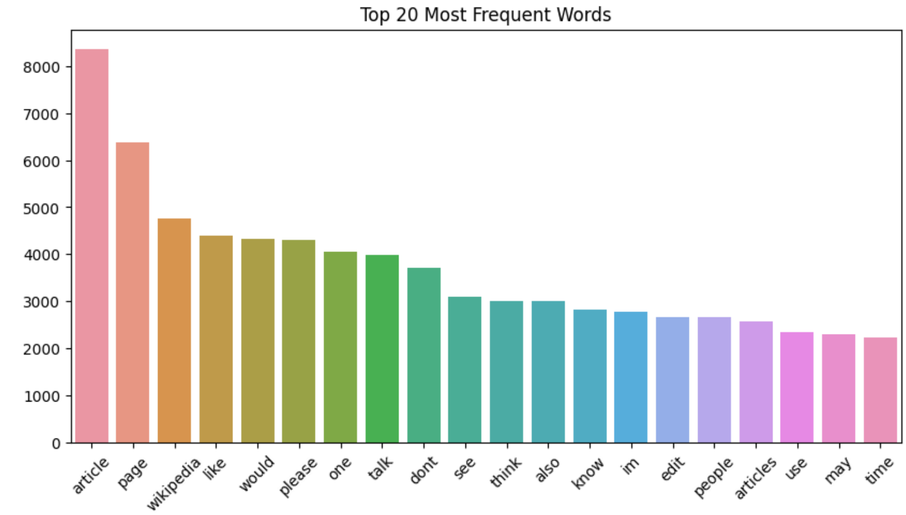

Figure 5: Evaluation plot

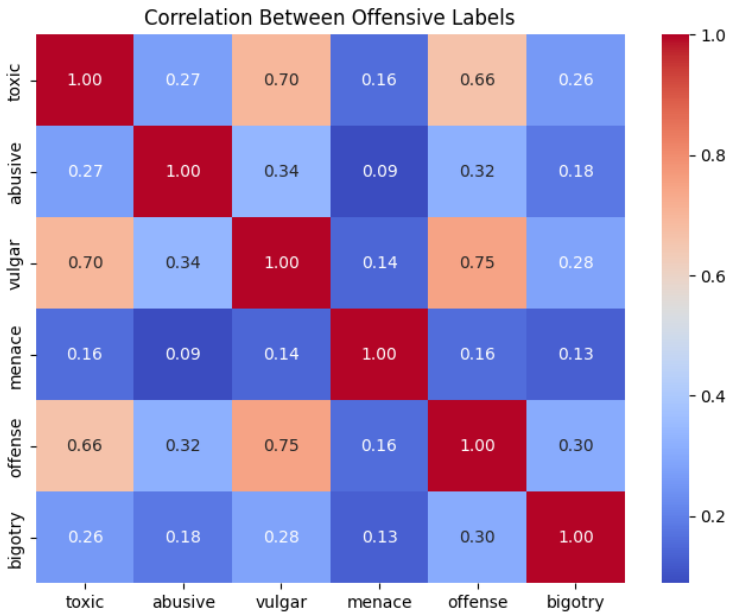

Figure 6: Correlation between offensive labels

Model Performance Summary

Model	Micro F1-score	Macro F1-score	ROC-AUC	Notes
Logistic Regression	0.92	0.89	0.87	Baseline using TF-IDF
LSTM	0.82	0.84	0.88	Sequential pattern learning
BERT	0.89	0.86	0.94	Transformer-based contextual model

Conclusion and Recommendation

Based on performance metrics and evaluation, BERT was the top-performing model. It is recommended for real-world deployment due to its robust understanding of context and significantly higher F1 and ROC-AUC scores.