

Final Project Presentation

Orkun Kinay

July 30, 2024

Team Member's Details

- ▶ **Group Name:** Orkun
- ▶ **Name:** Orkun Kinay
- ▶ **Email:** orkunkinay@sabanciuniv.edu
- ▶ **Country:** Turkey
- ▶ **College/Company:** Sabancı University
- ▶ **Specialization:** NLP

Model Performance Summary

- ▶ **Logistic Regression:** 89.26
 - ▶ A simple yet effective linear model for binary classification tasks.
 - ▶ Provides a solid baseline accuracy and is computationally efficient.
- ▶ **Random Forest:** 99.71
 - ▶ Demonstrated the highest accuracy among the tested models.
 - ▶ Handles imbalanced data well and provides feature importance.
- ▶ **Gradient Boosting:** 89.40
 - ▶ Effective but performance slightly lower than Random Forest.
 - ▶ Known for improving prediction accuracy through its iterative approach.
- ▶ **Support Vector Machine (SVM):** 89.65
 - ▶ Achieved good accuracy with a linear kernel.

Interpretation of Results

- ▶ **Highest Accuracy:** Random Forest model achieved the highest accuracy at 99.71
- ▶ **Consistency:** Both Logistic Regression and SVM provided consistent results with accuracies close to 89
- ▶ **Gradient Boosting:** Showed promise with slightly lower accuracy. Could benefit from further hyperparameter tuning.

Recommended Models for Technical Users

Based on the insights from the EDA, the following models are recommended for the hate speech detection task:

- ▶ **Logistic Regression:** Effective for binary classification tasks and computationally efficient.
- ▶ **Random Forest:** High accuracy and robustness in handling imbalanced data.
- ▶ **Gradient Boosting:** Effective for improving prediction accuracy.
- ▶ **Support Vector Machines (SVM):** Powerful for text data and useful when the data is linearly separable.
- ▶ **Deep Learning Models:** BERT can capture the contextual meaning of words and improve classification performance.

Summary

The EDA has provided valuable insights into the dataset, highlighting the importance of text length and common words in distinguishing hate speech from non-hate speech tweets. The recommended models offer a range of approaches to effectively classify hate speech, with options for both traditional machine learning and deep learning methods.

GitHub Repo Link

<https://github.com/orkunkinay/Hate-Speech-Detection/tree/main>