

Tackling Cyberbullying on Twitter: A Comparative Analysis of Machine Learning Algorithms for Sentiment Analysis

Social Computing

UMASS Lowell

Sharmin Sultana

01339552

Anaadumba Raphael

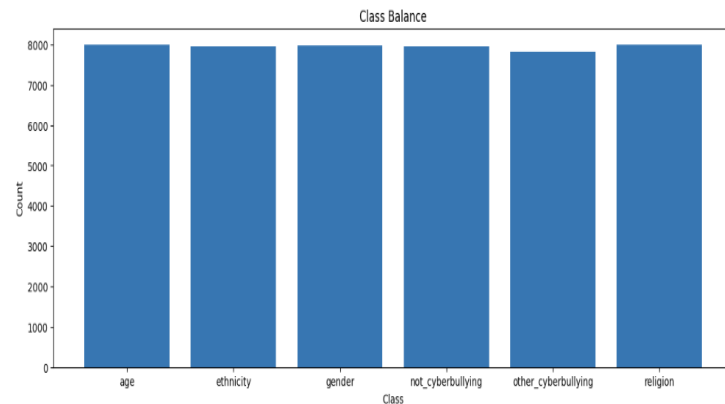
02019997

Introduction & Motivation

- This article discusses the issue of cyberbullying on social media platforms like Twitter and explores the potential of using sentiment analysis to address this problem.
- The study's results can aid in the creation of better cyberbullying detection tools, which can lead to a safer and more positive online space.
- In addition, addressing the problem of cyberbullying, will create safe social media environment for all users.
- This study findings provide researchers which machine learning algorithm is the most effective in detecting cyberbullying sentiment on Twitter.

Dataset & Evaluation Method

- The study utilized the Cyberbullying dataset, which contains over 47,000 raw tweets categorized into various types of cyberbullying, with a balanced distribution of 8000 instances per class.
- The study used a pre-processed dataset of 40,000 tweets, divided into 80% training and 20% testing data
- Vectorised features using the tfidf feature extraction method.
- The aim of the study was to perform sentiment analysis on the dataset using machine learning techniques.
- The performance of the models was evaluated using various metrics, including accuracy, recall, precision, and F1 score.



Method

- The study utilized various methods to identify and classify cyberbullying tweets, including data pre-processing, feature extraction, and machine learning algorithms.
- Data pre-processing: Pre-processing aims to transform data for better analysis by removing unnecessary information, such as special characters and links, in order to generate a cleaner dataset.
- Feature extraction: We used TF-IDF in our cyberbullying project to extract features from tweet text.
- Machine Learning Algorithms: We used several machine learning algorithms including Logistic Regression, Support Vector Machines, Random Forest, and Multilayer Perceptron for tweet classification.

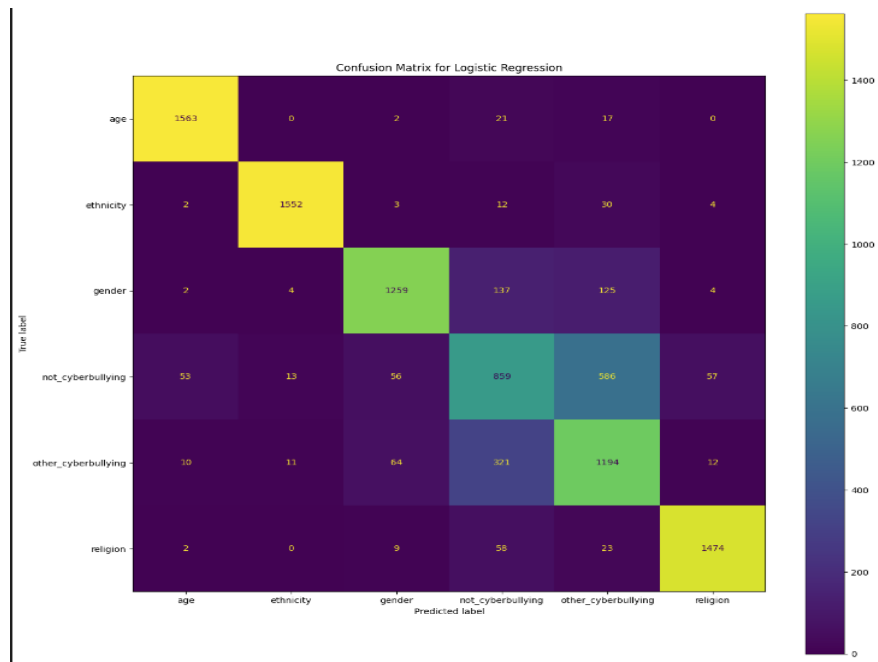
Results and Insights

- The Logistic Regression model performed best with accuracy of 0.83, followed by MLP and SVM with 0.82 and 0.825 accuracy respectively. Random Forest had the lowest accuracy of 0.81.

Table 1: Performance of Machine Learning Models

| Model | Accuracy | Recall | Precision | F1 Score |
|------------------------|----------|--------|-----------|----------|
| Logistic Regression | 83.00% | 83.48% | 82.94% | 83.01% |
| Random Forest | 81.00% | 81.04% | 80.92% | 80.93% |
| Support Vector Machine | 82.5% | 84.02% | 83.51% | 83.45% |
| Multi-layer Perceptron | 82.0% | 81.88% | 81.85% | 81.84% |

- We also visualized the confusion matrix diagram for a better understanding of each model.



References

- [1] R. Suhas Bharadwaj, S. Kuzhalvaimozhi, and N. Vedavathi. A novel multimodal hybrid classifier based cyberbullying detection for social media platform. In Radek Silhavy, Petr Silhavy, and Zdenka Prokopova, editors, *Data Science and Algorithms in Systems*, pages 689–699, Cham, 2023. Springer International Publishing.
- [2] Md Manowarul Islam, Md Ashraf Uddin, Linta Islam, Arnisha Akter, Selina Sharmin, and Uzzal Kumar Acharjee. Cyberbullying detection on social networks using machine learning approaches. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pages 1–6, 2020.
- [3] Xing Li, Zhanqi Xu, Fan Yang, and Yunbo Li. Multi-objective hybrid evolution with information entropy awareness for controller placement. In *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 135–140, 2021.
- [4] Noviantho, Sani Muhamad Isa, and Livia Ashianti. Cyberbullying classification using text mining. In *2017 1st International Conference on Informatics and Computational Sciences (ICICoS)*, pages 241–246, 2017.