Student Id ;23241085
Student Name ; Omor Bin Amjad Chowdhury
Course Code:424
Section:01
Paper Title: An Ensemble-Based Multi-Classification Machine Learning Classifiers Approach to Detect Multiple Classes of Cyberbullying
Paper Link: https://www.mdpi.com/2504-4990/6/1/9

1.1 Motivation/Purpose/Aims/Hypothesis:

The objective of this research is to develop a system for detecting six different types of cyberbullying (age, gender, religion, race, ethnicity, and general bullying) so as to deal with the escalating challenge of cyberbullying on social media, specifically Twitter. The main aim is to enhance the precision and efficacy of current detection models using ensemble machine learning approaches that effectively deal with multiclass datasets by combining various classification models.

1.2 Contribution:

Two ensemble models are presented in this paper that use Decision Trees, Random Forests, and XGBoost classifiers. Stacking and voting methods are applied to combine them. It is shown that when it comes to multi-class classification of cyber bullying tweets, these ensemble classifiers do better than traditional machine learning models on a dataset with six different types of bullying.

1.3 Methodology:

- **Dataset**: The dataset comprises 48,000 tweets drawn from Twitter that had been classified into six distinct cyberbullying categories through which the study was conducted.
- **Feature Extraction**: For feature extraction, the authors employed Term Frequency-Inverse Document Frequency (TF-IDF) with unigrams, bigrams, and trigrams.
- **Classifiers**: The three machine learning classifiers tested were: Decision Trees (DT), Random Forest (RF) and XGBoost. These classifiers were combined using two ensemble methods- stacking and voting.
- **Evaluation**:Models were evaluated on accuracy, F1-score, precision, recall and area under curve (AUC). Stacking classifier achieved the highest accuracy of 90.71%.

1.4 Conclusion:

The stacking classifier, one of the ensemble-based approaches, was found to be more effective in detecting different forms of cyberbullying than conventional models in terms of its precision and performance. Moreover, it was shown that the use of ensemble methods together with proper feature extraction can greatly enhance multi-class classification

2. Critiques/Limitations:

2.1 First Critique/Limitation:

One major limitation of this model is that it only works based on English tweets from Twitter; hence, it cannot be generalized to other languages or other social network sites. Multilingual datasets or platform-specific data could be considered in future research through the same methodology used in this work.

2.2 Second Critique/Limitation:

In spite of the fact that they employed a hefty amount of information, there are only six classifying kinds of cyber bullying. There were no other facets (for example, intimidation or pretending to be someone else) that may have impaired this model's capability to detect various forms of harassment in daily life.

2.3 Third Critique/Limitation:

Thus, TF-IDF does not take into account the contextual meaning of the words as it is based only on frequencies and hence is not appropriate for such applications. Even though TF-IDF has proven effective in beginning texts analyses, other techniques such as Word2Vec or BERT seem promising in terms of better understanding Twitter user context.

3. Synthesis: Potential Applications and Future Work

3.1 First Potential/Idea for a New Paper:

One possible next research paper might investigate the performance of this ensemble-based technique when applied to multilingual data sets, thus widening its use for detecting cyberbullying in different tongues. Additionally, it may involve the employment of context-dependent features extraction methods such as BERT to improve efficiency.

3.2 Second Potential/Idea for a New Paper:

In the future, the development of a real-time cyberbullying detection system could be looked into as yet another avenue for future research, which would involve using streamed data from various social media platforms. This would be a way to assess the scalability and performance of ensemble model under real time conditions.