Cyber Bullying Classification on Twitter

Nakul Periwal	Ojasva Singh	Parmesh Yadav	Yukti Rao
2020316	2020318	2020319	2020352

1. MOTIVATION

Cyber bullying is a serious and widespread problem that affects millions of people around the world. It is defined as the use of electronic communication to harass, threaten, or harm others, especially on social media platforms. Cyber bullying can have negative impacts on the mental health, well-being, and safety of the victims, as well as the perpetrators and bystanders.
In this project, we will use a dataset of more than 47000 tweets that have been labeled according to the class of cyber bullying they belong to. The dataset is balanced, meaning that it contains approximately 8000 tweets for each class. The tweets are written in English and may contain profanity, slang, or abbreviations.

2. RELATED WORK

2.1 An automatic cyberbullying method to avoid bad consequences of cyber harassment. Four machine learning algorithms are used to identify bullying text and SVM for both BoW and TF-IDF.**[1]**
2.2 The paper presents a method to detect cyberbullying on social media using machine learning. The paper uses two different ways to analyze the words and the emotions in the tweets. The method also uses two different types of classifiers to categorize the tweets as cyberbullying/ not.**[2]**
2.3 literature overview study was done cyberbullying, its consequences and different machine learning techniques for detection. It shown that majority of focus in cyberbullying detection is from text dataset, where this dataset was either from publicly available sites or generated their own dataset and then this dataset is used to train the machine learning algorithm.**[3]**

3. TIMELINE

Week	Task
4-5	Pre-processing and Data Visualization.
6	Naïve Bayes & Random Forest
7	Model Performance Analysis and Midsem report
9-10	SVM & K Nearest
11	Performance analysis
12	Final Report + Presentation

4. INDIVIDUAL TASKS

Tasks	Team Member/s
Data Visualization	Yukti, Parmesh
Data Preprocessing + Feature Selection	Yukti, Parmesh
Naïve Bayes	Ojasva
Random Forest	Nakul
Performance Analysis	Everyone
Midsem Report	Everyone
SVM	Ojasva, Parmesh
K Nearest	Yukti, Nakul
Final Conclusion	Everyone
Final Report + Presentation	Everyone

5. FINAL OUTCOME

The final outcome of this project is a machine learning model that can accurately and efficiently classify tweets based on the type of cyber bullying they contain. The model can also flag potentially harmful tweets as either cyber bullying or not. The model can be used to assist Twitter moderators, users, and researchers in detecting and preventing cyber bullying on the platform.
The project demonstrates the potential and challenges of using machine learning to combat cyber bullying on social media. In order to combat cyber bullying on Twitter, it is important to develop machine learning models that can automatically detect and classify potentially harmful tweets based on the type of cyber bullying they contain.
In this project, we will use a dataset of more than 47000 tweets that have been labeled according to the class of cyber bullying they belong to.

6. REFERENCES

[1] https://www.researchgate.net/publication/351131976_Cyberbullying_Detection_on_Social_Networks_Using_Machine_Learning_Approaches
[2] https://www.researchgate.net/profile/Eslam-Amer-3/publication/333506989_Social_Media_Cyberbullying_Detection_using_Machine_Learning/links/5cf0c852299bf1fb184baf29/Social-Media-Cyberbullying-Detection-using-Machine-Learning.pdf
[3] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4116153