Title : Project Registration & Progress Review **FF No. 180**

Department: Multidisciplinary Engineering	**Academic Year:** 2022-23	
Semester: I	**Batch :** 3	**Group No. :** 2
Project Title: Cyberbullying analysis in social media		
Project Area: Artificial Intelligence and Machine Learning, Natural Language Processing		

Group Members Details:

Sr. No.	Class & Div.	Roll No.	G.R.No.	Name of Student	Contact No.	Email ID
1	AI-B	45	12110152	Samved Patil	8999638550	samved.patil21@vit.edu
2	AI-B	48	12111408	Siddhi Patil	8605194573	siddhi.patil211@vit.edu
3	AI-B	55	12110717	Vaishnavi Petkar	9325397073	vaishanavi.petkar21@vit.edu
4	AI-B	56	12110650	Mrunmayee Phadke	8767581918	mrunmayee.phadke21@vit.edu

Name of Internal Guide: Puja A. Cholke
Contact No. & Email ID: 9028175205 & puja.cholke@vit.edu

Project approved / Not approved

29/9/22

Guide **Project Coordinator** **Head of Department**

Project Synopsis

As people spend more time utilising technology that keeps them constantly connected to other people, cyberbullying is becoming increasingly common. Cyberbullies can communicate with their victims in a variety of methods, including text messaging, social networking websites, and instant messaging through the internet. Cyberstalking and harassment are just a few examples of the various types of cyberbullying. Cyberbullying is a significant issue that, like traditional bullying, may make the victim feel inadequate and unduly self-conscious and even lead to suicidal thoughts.

It is paramount to detect and ensure safety across social platforms. Thus, we aim to build a model using Artificial Intelligence and Machine Learning (AI&ML) and Natural Language Processing (NLP) to detect various types of cyberbullying on social media. The categories that will compile to form the model are slang corpus, hate comments, racism, sexism, body shaming, religious discrimination, age based discrimination, sentiment analysis, offensive messages, spam detection and overall hate-speech detection. To proceed with the same we first will research through various research papers and articles. Then we will collect data (which should be unbiased, thus including bullying and non-bullying content), process it using different data pre-processing techniques and finally use multiple algorithms for the analysis.

We expect the tentative outcome to analyse the variety of content on social media platforms and detect the parameters mentioned above that make the crux of cyberbullying.

Batch No.	3	Group No.	2		
Activity			Review Schedule	Progress Review Report submitted	Signature of Guide
Review 1			Mid Sem. Semester	Yes	
Review 2			End of Semester	Yes / No	

Format of Progress Review Report:

Review No.: 1 **Group No.: 2** **Date: 10/10/22**

Progress Review Report

We read various research papers and review papers and gained knowledge of work done on the topic till date and gauged their drawbacks. We made a literature survey. We learnt about the tools and technologies required to solve the problem statement. This includes having a good knowledge of Natural Language Processing (NLP) and the libraries required like nltk and spacy. We created a model for spam detection. For that we searched rigorously for an unbiased and large dataset. We finally chose a cited dataset from UCI and used various preprocessing techniques on it like tokenization and stemming. Then we created a bag of words model followed by the algorithm, Naive Bayes Classifier model where we split the dataset into train and test datasets (4:1). The confusion matrix showed great results and the model had an accuracy of 98.4753%. We are now working on finding suitable datasets, data preprocessing techniques and algorithms for the remaining parts of the problem statement.

Signature of Guide:

Review No.: 2	Group No.: 2	Date:16/12/21

Progress Review Report

Samved Patil- Reviewed more research papers and implemented a binary-based model using Linear Regression with accuracy of 78.82%. Also, in the data preprocessing techniques like lemmatization and stemming.

Siddhi Patil- Gathering of datasets and their compilation. Worked on Preprocessing techniques like tokenization and using RE functions for the same. Also, in the modelling of cyberbullying detection.

Vaishnavi Pethkar- Worked on data preprocessing techniques and word embeddings that were used in the project like TF-IDF. Also, in the implementation of model using UI .

Mrunmayee Phadke- Collecting of data sets and worked in modelling for detection. Implemented the GUI using Streamlit and worked in this deployment of the model.

Photo:

Signature of Guide: 13/12/22