1. Title:

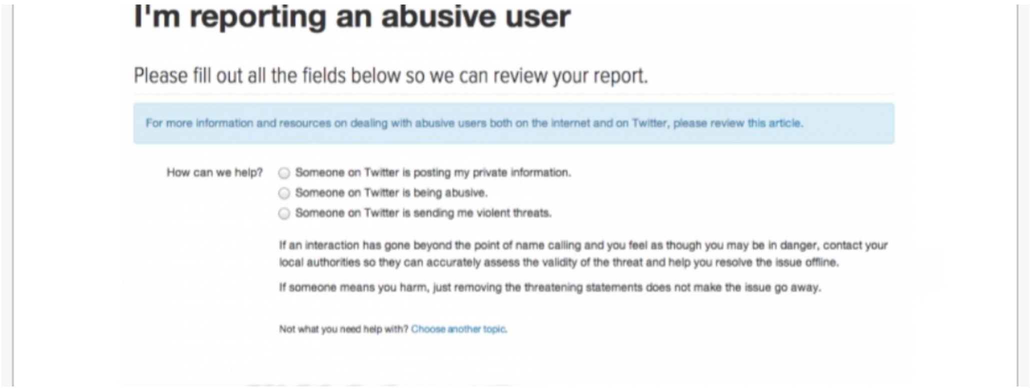
Predicting Cyberbullying on Social Media Using Machine Learning Techniques

2. Project Statement:

The project aims to address the escalating concern of cyberbullying on social media platforms (such as Twitter (or X), Instagram, Facebook, etc.) by utilising machine learning and deep learning algorithms, such as Support Vector Machine (SVM), Convolutional Neural Networks (CNN), Random Forest etc to predict and report them.
To improve the accuracy of the solution, the project will leverage the Natural Language Toolkit for data preprocessing and feature extraction. Then, using the ML techniques, models will be built and evaluated to effectively distinguish cyberbullying on social media. This project will be helpful for timely detection of bullying episodes and providing assistance to victims.

3. Outcomes:

- Real Time Detection of Cyberbullying Episodes: Creation of a real-time system that monitors social media data and alerts authorities or organisers of potential harassment or bullying on social media.
- Understanding of Types of Cyberbullying: Researchers can also gain insights on the kind of cyberbullying such as harassment over Age, Religion, Ethnicity, Gender, etc. The authorities can then take necessary steps based on the type of cyberbullying.
- Deployment and integration: Researchers can focus on the deployment and integration of cyberbullying tweet prediction models into existing social media platforms. This can provide real-time feedback to users and contribute to a safer and more inclusive online environment.
- Contribution to Cyber Safety: The ultimate outcome of such a project would be to contribute to cyber safety and security by providing a tool that can detect harassment on social media. Cyberbullying is a grave issue with severe consequences and such ML models can provide promising solutions to combat it.



I'm reporting an abusive user

Please fill out all the fields below so we can review your report.

For more information and resources on dealing with abusive users both on the internet and on Twitter, please review this article.

How can we help?
○ Someone on Twitter is posting my private information.
○ Someone on Twitter is being abusive.
○ Someone on Twitter is sending me violent threats.

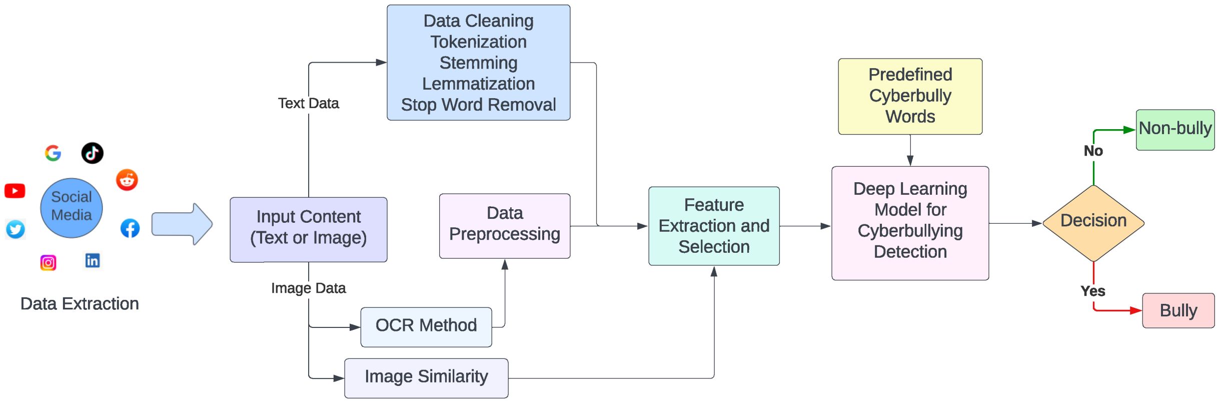
If an interaction has gone beyond the point of name calling and you feel as though you may be in danger, contact your local authorities so they can accurately assess the validity of the threat and help you resolve the issue offline.

If someone means you harm, just removing the threatening statements does not make the issue go away.

Not what you need help with? Choose another topic.

Modules to be Implemented:

1. Data Ingestion
2. Exploratory Data Analysis (EDA)
3. Data Preprocessing using NLP techniques
4. Machine Learning Models (Random Forest, SVM, CNN, etc.)
5. Evaluation and Comparative Analysis of Models
6. Project Presentation & Documentation



Week-wise Implementation Plan of Modules:

Milestone 1: Week 1-2

Module 1: Data Collection and Importing Relevant Libraries

- Understand the problem statement
- Gather Twitter (or X) data from relevant sources
- Import relevant libraries on Python

Module 2: Exploratory Data Analysis (EDA)

The goal is to perform EDA on the raw data and provide data visualisations in the form of charts. Examples below:

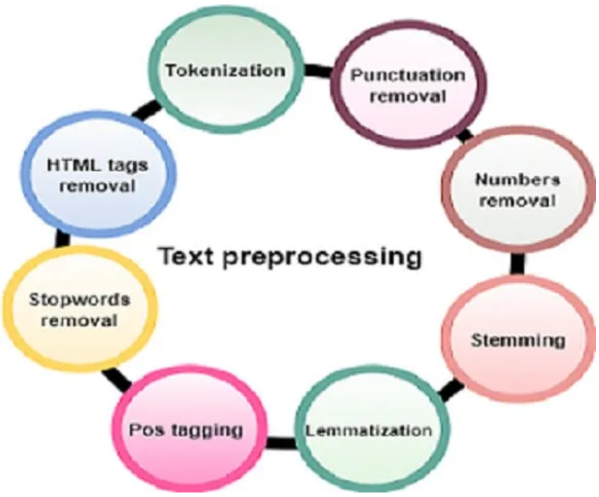
- Plot the distribution of tweets labelled on type of cyberbullying
- Plot distribution charts based on word lengths
- Plot word clouds for different label classifications
- Bar charts based on common words

Milestone 2: Week 3-4

Module 3: Data Preprocessing

The social media data (tweets in this case) consists of massive amounts of noise. Therefore, a rigorous data preprocessing will be implemented to ensure the quality and reliability of the dataset. This will involve:

- Cleaning and filtering the social media content to remove noise, irrelevant information, and duplicate posts.
- NLP techniques will be used for text normalisation, tokenisation, stemming, and removal of stop words to standardise the textual data.



Milestone 3: Week 5-6

Module 4: Building Machine Learning Techniques

The goal is to build a suite of sophisticated ML models on the transformed textual data to identify cyberbullying. These models are recursively evaluated and tuned to make them more efficiently predictive. Some of the proposed models are:

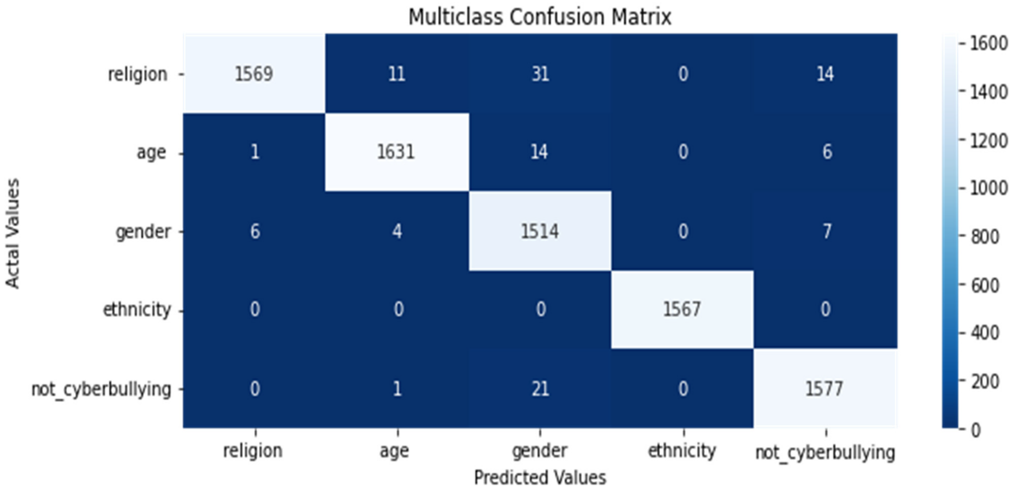
- **Convolutional Neural Networks:** CNN models are designed to process data through multiple layers of arrays. Text-based CNNs work on word embeddings in the form of matrices.
- **Random Forest:** RF combines several different classifiers to find solutions to complex tasks. A random forest is essentially an algorithm consisting of multiple decision trees, trained by bagging or bootstrap aggregating. A random forest text classification model predicts an outcome by taking the decision trees' mean output.
- **Naïve Bayes model**: A probabilistic supervised learning approach that works with a likelihood function that illustrates the probability of witnessing a specific value of a feature.
- **Support Vector Machine:** SVM is a supervised ML model that uses classification techniques to categorise new text after being given labeled training data sets for each category.

Milestone 4: Week 7

Module 5: Evaluation and Comparative Analysis of Models

The goal is to do a comparative analysis of the results obtained from the implementation of various algorithms on selected datasets.

- Utilise parameters such as accuracy, recall, precision and F1-score to carry out this analysis.
- To the best performing models, provide as series of texts to observe and record real-time predictions.



	religion	age	gender	ethnicity	not_cyberbullying
religion	1569	11	31	0	14
age	1	1631	14	0	6
gender	6	4	1514	0	7
ethnicity	0	0	0	1567	0
not_cyberbullying	0	1	21	0	1577

Milestone 5: Week 8

Module 6: Project Presentation and Documentation

- Prepare a presentation and demo with following structure:
  - Problem Statement and Objective
  - Methodology (Brief overview of models used)
  - Results & Insights (emphasise on key takeaways)
  - Visualisations
  - Q&A Session
- Clear visualisations and minimum overly technical text in presentations.
- Documentation preparation in below mentioned format:

- Project Overview: Problem statement, goals, expected outcomes
- Data Sources: Details on where data was acquired
- Data Preprocessing and Cleaning: Steps taken, techniques used, justification
- Exploratory Data Analysis: Summary of findings, key visualisations
- Model Development: Explanation of model choices, rationale for parameter selection
- Model Evaluation: Performance metrics used, comparison of different models
- Predictive Results: Examples of predictions of cyberbullying
- Appendix: Code snippets (well-commented), additional visualisations, etc.

Evaluation Criteria:

Milestone 1 Evaluation (Week 1-2):

- Successful loading of the dataset into a suitable format (e.g., Pandas Data Frame inPython).
- Identification of missing and duplicate values and handling strategy
- Approval of Initial summary statistics to understand the data distribution.
- Approval of thorough examination of data distributions (histograms, box plots, etc.).

Milestone 2 Evaluation (Week 3-4):

- Approval of steps for data preprocessing techniques and its implementation.
- Approval of outcomes of the data preprocessing through visualisations of input vs output data for each data preprocessing step.

Milestone 3 Evaluation (Week 5-6):

- Approval of the Machine Learning models and architectures to be used on the processed dataset.
- Approval of the hyperparameter tuning process and the range of parameters explored.
- Completion and approval of performance metrics for all built models.

Milestone 4 Evaluation (Week 7-8):

- Approval of the final model based on evaluation criteria
- Approval of the presentation and project documentation.
- Final code submission on GitHub.

Trigger Warning: *The cyberbullying datasets may contain strong language on sensitive topics, such as violence, abuse, discrimination, and/or mental health issues*