Cyberbullying Classification

- **Group Member name:**
    1) Shrinivas Patil
    2) Nishant Asodariya

- **Problem Statement:**

As social media became an integral part of human life, the vast majority use it as a medium of day to day life communication. But, most of the time the fringe elements in the society use it as a medium to bully anyone. Cyberbullying can impact anyone's life at any time, anywhere and related anonymity. We can fight this by creating NLP models to flag potential harmful tweets.

- **Programming language:**

Python 3.8

- **Dataset:**

Dataset is available on Kaggle Datasets. The dataset has 6 categories like religious, age, gender, ethnicity, other and not cyberbullying. The link for the dataset is as follows:
https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification

- **Existed Solutions:**

There are several solutions for this problem. These solutions are developed by using bags of words and different machine learning algorithms. I am going to work with TFIDF vectorizer or Word2Vec with Naive Bayes Classifier to get good accuracy.

- **Algorithm:**

For this problem, we will use naive bayes classifier with TfIdf vectorizer or word2vec as preprocessing techniques. We will remove all the special characters from the text and convert it into lowercase. Then will remove all the stopwords from the text. Now, the given text will be encoded into machine understandable language by vectorizers. Finally, this data will be given to the naive bayes classifier.

- **Timeline:**

  The project will take 2 weeks of time. The final algorithm will be ready up to 2tnd of dec.

- **Computing Platform:**

  Google Colab notebooks, GPU- P80

- **Roles & Responsibilities:**

  1) **Shrinivas Patil:**

     I have prior experience in machine learning and NLP projects. I have worked on several ML and NLP projects like Twitter sentiment analysis, text emotion classification, restaurant review classification. I will create different machine learning models to fit clean data to get better accuracy.

  2) **Nishant Asodariya:**

     I will analyse the dataset which is taken from kaggle. I will check whether there are any null records present in the dataset. If they are then I will remove those null records. I will make sure that there will not be any duplicate records in the dataset.