A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

Machine Learning Project- Interim Presentation

Cyberbullying Tweet Classification and Sentiment
Analysis



Problem Statement

Given a tweet, we are classifying the tweet as not_bullying or bullying (religion, age, ethnicity, gender) using machine learning and natural language processing.

For this problem we are using Kaggle dataset.

Dataset- <https://www.kaggle.com/andrewmvd/cyberbullying-classification>

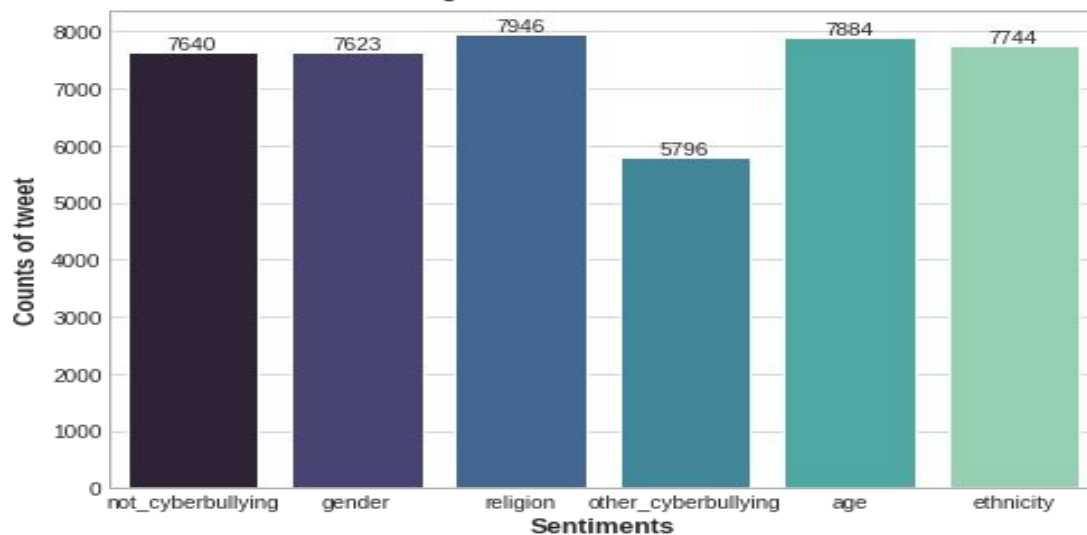


EDA and Pre-Processing

Following steps were performed for EDA & pre-processing:

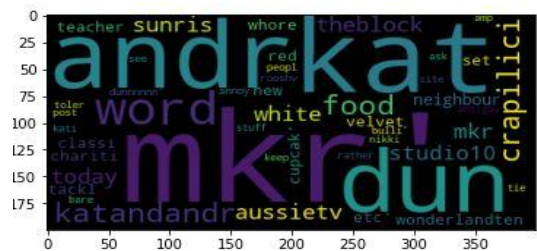
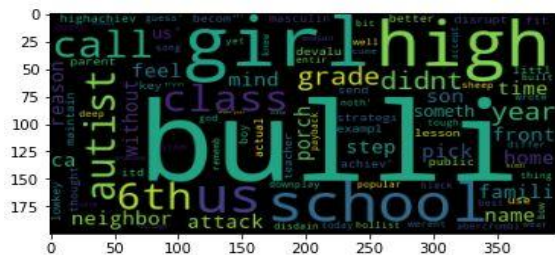
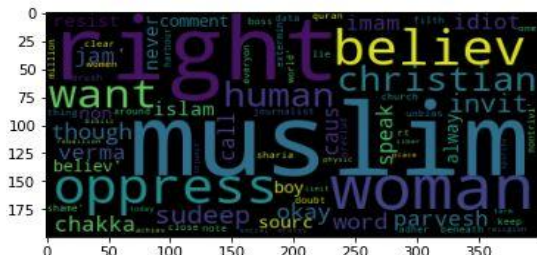
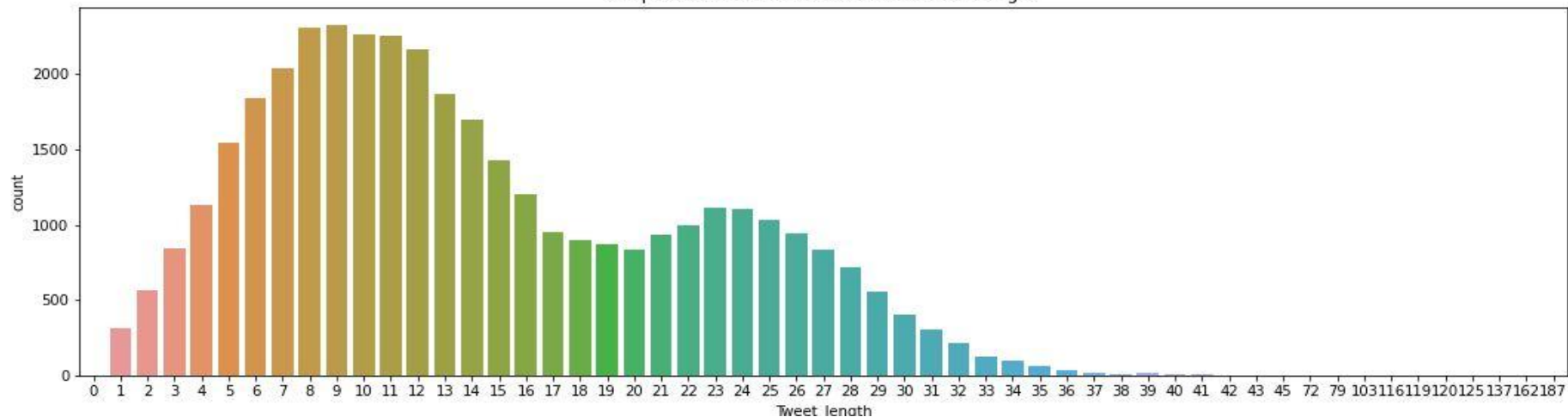
- Histogram plots were plotted to check for class imbalance
- Data cleaning was performed using various libraries like nltk, re, emoji, string
- In data cleaning we removed special characters (hashtags, URLs, etc), emojis, word contractions and multiple spaces between words.
- After cleaning we removed duplicate tweets
- After removing duplicate tweets, other_cyberbullying class was creating imbalance, so this class was dropped.
- Stemming, Lemmatization was performed and their performance was compared .
- Feature extraction was done using various word embedding algorithms like TF-IDF, Word2vec, Bag of words and GloVe and they were used to create word embeddings to train Naive Bayes model.

Showing class balance/imbalance result



	Basic clean	sentiment	tweet_length	Stemming	Lemmatization	Lemma_Stemming
0	words katandandre food crapilicious mkr	not_cyberbullying	5	word katandandr food crapilici mkr	word katandandre food crapilicious mkr	word katandandr food crapilici mkr
1	aussietv white mkr theblock today sunrise stud...	not_cyberbullying	10	aussietv white mkr theblock today sunris studi...	aussietv white mkr theblock today sunrise stud...	aussietv white mkr theblock today sunris studi...
2	classy whore red velvet cupcakes	not_cyberbullying	5	classi whore red velvet cupcak	classy whore red velvet cupcake	classi whore red velvet cupcak
3	meh p thanks heads concerned another angry dud...	not_cyberbullying	9	meh p thank head concern anoth angri dude twitter	meh p thanks head concerned another angry dude...	meh p thank head concern anoth angri dude twitter
4	isis account pretending kurdish account like i...	not_cyberbullying	8	isi account pretend kurdish account like islam...	isi account pretending kurdish account like is...	isi account pretend kurdish account like islam...

Graph of count of tweets and thier tweet length





Baseline Approaches and Advanced Techniques

Baseline Technique - Naive Bayes

Advanced Techniques -

- BERT- It generates the contextualised based word embedding using Encoders Deep learning based Architecture. The encoder is responsible for reading text input and processing. The decoder is responsible for producing a prediction for the task.
- LSTM- It is a type of recurrent neural network with extended features which helps in overcoming the vanishing gradient problem. Spacy embeddings can be used for this model.
- CNN- 1D CNN is used to extract the features from the text and using multiple convolution layers and pooling layers high level features maps are extracted which further passed to softmax layer to classify the tweet.

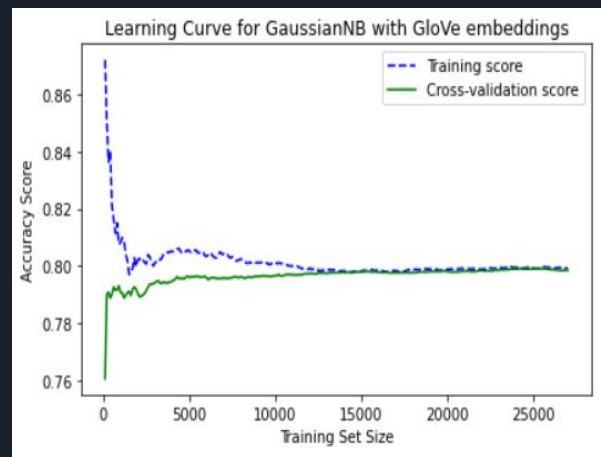
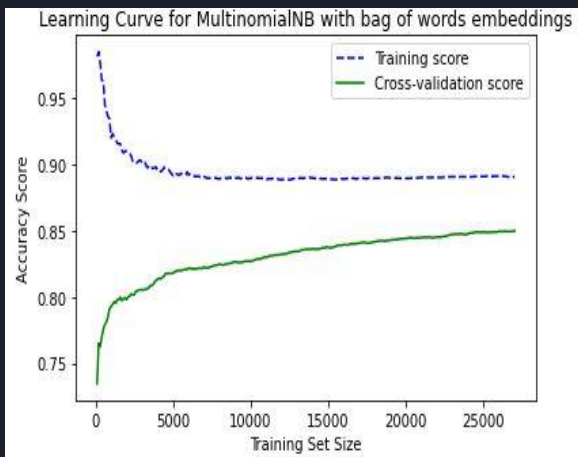
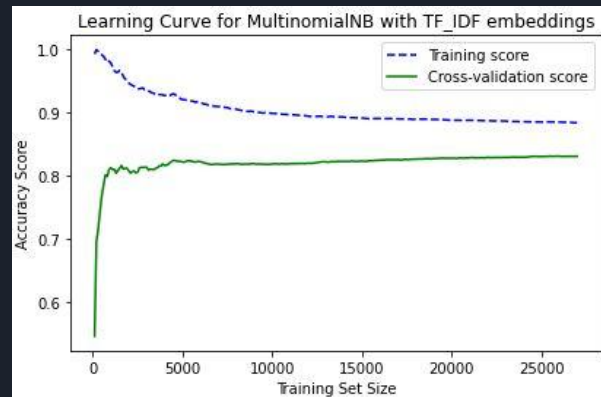
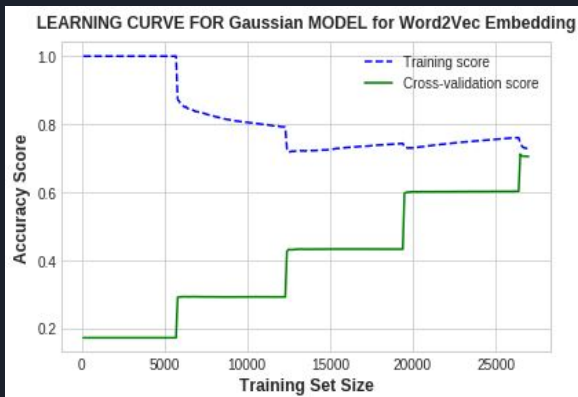


Results and Analysis

- We dropped *other_cyberbullying* class after data cleaning as it was creating class imbalance and we found that the F1 score of *other_cyberbullying* class was ~ 0.52 and accuracy was $\sim 75\%$ but after dropping, our model accuracy improved to 83%.
- We chose Stemming over Lemmatization since stemming was performing slightly better than Lemmatization
- We compared the results of different word embeddings on Naive Bayes model. We got following accuracy results -
 - Bag of words- 85%
 - Tf-idf- 83%
 - GloVe- 79%
 - Word2vec- 70%

Best performance was obtained from **Bag of words** embedding with an accuracy of **85%** on Naive Bayes.

Training Vs Testing Error Analysis





Future plans

- Use another word embedding algorithm SpaCy for training LSTM model
- Use BERT for tweet classification
- Use CNN for tweet classification
- Compare the results of above advanced techniques using various evaluation metrics like precision, recall, f1-score etc.



Individual Contribution

Every individual contributed equally to the project and it was a complete team effort.