CYBERBULLYING DETECTION USING MACHINE LEARNING

A Project Progress Report

Submitted in partial fulfillment for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by

N VEMACHALAM	(N180701)
K DHANESH	(N180049)
S NIKHIL	(N180978)
N SHANTHA KUMARI	(N180139)
G INDAMMA	(N180053)

Under the Esteem Guidance of

Mr. R Upendara Rao

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

Rajiv Gandhi University of Knowledge Technologies – Nuzvid

Nuzvid, Eluru(Dist), Andhra Pradesh – 521202.

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

Rajiv Gandhi University of Knowledge Technologies – Nuzvid

Nuzvid, Eluru(Dist), Andhra Pradesh – 521202.

CERTIFICATE OF COMPLETION

This is to certify that the work entitled, **"Cyberbullying Detection using Machine Learning"** is the bonafide work of **N VEMACHALAM (ID NO: N180701), K DHANESH (ID NO: N180049), S NIKHI (ID NO:N180978), N SHANTHA KUMARI (ID NO: N180139), G INDAMMA (ID NO: N180053)** carried out under my guidance and supervision for 4th year project of **Bachelor of Technology** in the department of Computer Science and Engineering under RGUKT IIIT Nuzvid. This work is done during the academic session August - November 2023 , under our guidance.

Mr. Upendara Rao Rayala

Assistant professor,

Department of CSE

RGUKT-Nuzvid

Mr. Chiranjeevi Sadu

Assistant Professor,

Department of CSE,

Head of the Department,

RGUKT-Nuzvid.

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

Rajiv Gandhi University of Knowledge Technologies – Nuzvid

Nuzvid, Eluru, Andhra Pradesh – 521202.

CERTIFICATE OF EXAMINATION

This is to certify that the work entitled, **"Cyberbullying Detection using Machine Learning"** is the bonafide work of **N VEMACHALAM (ID NO: N180701), K DHANESH (ID NO: N180049), S NIKHI (ID NO:N180978), N SHANTHA KUMARI (ID NO: N180139), G INDAMMA (ID NO: N180053)** and hereby accord our approval of it as a study carried out and presented in a manner required for its acceptance in the 3rd year of **Bachelor of Technology** for which it has been submitted. This is approval does not endorse or accept every statement made, opinion expressed, or conclusion drawn, as recorded in this report for the purpose for which it has been submitted.

--	-----------------------------
Mr. R upendara rao	**PROJECT EXAMINER**
Assistant Professor,	Assistant Professor,
Department of CSE,	Department of CSE,
RGUKT-Nuzvid.	RGUKT-Nuzvid.

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

Rajiv Gandhi University of Knowledge Technologies – Nuzvid

Nuzvid, Eluru, Andhra Pradesh – 521202.

DECLARATION

We **"N VEMACHALAM (ID NO: N180701), K DHANESH (ID NO: N180049), S NIKHI (ID NO:N180978), N SHANTHA KUMARI (ID NO: N180139), G INDAMMA (ID NO: N180053)''** hereby declare that the project report entitled **"Cyberbullying Detection using Machine Learning"** done by us under the guidance of Mr. R Upendara Rao, Assistant Professor is submitted for the partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering during the academic session October 2023 - at RGUKT-Nuzvid.

We also declare that this project is a result of our own effort and has not been copied or imitated from any source. Citations from any websites are mentioned in the references. The results embodied in this project report have not been submitted to any other university or institute for the award of any degree or diploma.

Date: 09-11-2023

Place: Nuzvid

N VEMACHALAM	(N180701)
K DHANESH	(N180049)
NIKHIL	(N180978)
N SHANTHA KUMARI	(N180139)
G INDAMMA	(N180053)

ACKNOWLEDGEMENT

We would like to express our profound gratitude and deep regards to our guide **Mr. R Upendara Rao** for his exemplary guidance, monitoring and constant encouragement to us throughout the B.Tech course. We shall always cherish the time spent with him during the course of this work due to the invaluable knowledge gained in the field of reliability engineering.

We are extremely grateful for the confidence bestowed in us and entrusting our project entitled **"Cyberbullying Detection using Machine Learning"**

We express gratitude to Mr. Chiranjeevi sadhu (HOD of CSE) and other faculty members for being source of inspiration and constant encouragement which helped us in completing the project successfully.

Our sincere thanks to all the batch mates of 2018 CSE, who have made our stay at RGUKT-NUZVID, a memorable one.

Finally, yet importantly, we would like to express our heartfelt thanks to our beloved God and parents for their blessings, our friends for their help and wishes for the successful completion of this project.

ABSTRACT

The usage of social media has grown rapidly over time with the growth of the internet. Social media is an influential platform where we can find both good and bad things. This rapid increase in social connectivity results in online abuse, harassment, cyberbullying, cybercrime, and online trolling. We mainly focus on Cyberbullying, because it directly affects the mental health of a person, particularly this is a serious issue for women, in some cases, some women attempt suicide. Many incidents have recently occurred worldwide due to online harassment, such as sharing private chats, rumors, and sexual remarks. Therefore, the identification of bullying texts or messages on social media has gained a growing amount of attention among researchers. The main purpose of this project is to design and develop an effective technique to detect online abusive and bullying messages by merging natural language processing and machine learning. This project takes a text as input and checks whether it's cyberbullying comment or not.

Table of Contents

List of figures

CHAPTER 1

INTRODUCTION

Generally, Bullying is done intentionally, it is not an accidental incident, where the bully hurts a targeted person on a purpose. A bully repeats the same behavior, again and again, he/she gets happiness by bullying others. A bully always tries to bully weaker people. Cyberbullying means bullying others through social media, by sending abnormal messages, posting abnormal comments, etc. Social media has certain guidelines to follow because social media has to maintain a transparent platform for the people who use it for good purposes. Cyberbullying is of various types like bullying regarding color, body-shaming, religion, sexism, etc. Cyberbullying affects people directly or indirectly. In some cases, people attempt suicide due to cyberbullying. As per the NCRB (National Crime Records Bureau) report (2014, chapter 18), the number of cases reported under cyber crime was 9622 as compared to 5693 in 2013. This indicates a significant 69.0% increase over the previous year. Further, Cybercrime is a category of crime, which seems to be affecting a wide spread of ages including teenagers, which is alarming. Children get exposed to the Internet at a young age due to educational requirements, self-interest, or due to peer/social pressure. Mostly women are facing many problems in social media. Cyberbullying or cyber-harassment refers to an electronic method of bullying or harassment. **Bullying Text:** This type belongs to bully-type comments or harassment. For example, ”go away bitch” is a bullying text or comment that we consider a negative comment. **Non-bullying Text:** These types of comments or posts are non-bullying or positive comments. For example, comment like ”This photo is very beautiful” is positive and non-bullying comments.

1.1 Machine Learning:

Machine learning is a modern innovation that has enhanced many industrial and professional processes as well as our daily lives. It’s a subset of artificial intelligence (AI), which focuses on using statistical techniques to build intelligent computer systems to learn from available data. With machine learning, computer systems can take all the customer data and utilize it. It operates on what’s been programmed while also adjusting to new conditions or changes. Algorithms adapt to data, developing behaviors that were not programmed in advance. Learning to read and recognize context means a digital assistant could scan emails and extract essential information. Inherent in this learning is the ability to make predictions about future customer behaviors. This helps you understand your customers more intimately and not just be responsive, but proactive. Deep learning is a segment of machine learning. In essence, it’s an artificial neural network with three or more layers. Neural networks with only one layer can make estimated predictions. The addition of more layers can assist with increasing optimization and accuracy. Machine learning is relevant in many fields, and industries, and has the capability to grow over time.

Real-time applications of Machine Learning:

Image recognition:

Image recognition is a well-known and widespread example of machine learning in the real world. It can identify an object as a digital image, based on the intensity of the pixels in black and white images or color images.

Speech recognition:

Machine learning can translate speech into text. Certain software applications can convert live voice and recorded speech into a text file. The speech can be segmented by intensities on time-frequency bands as well.

Real-world examples of speech recognition:

- Voice search
- Voice dialing
- Appliance control
- Some of the most common uses of speech recognition software are devices like Google Home or Amazon Alexa.

Predictive analytics:

Machine learning can classify available data into groups, which are then defined by rules set by analysts. When the classification is complete, the analysts can calculate the probability of a fault.

Real-world examples of predictive analytics:

- Predicting whether a transaction is fraudulent or legitimate
- Improve prediction systems to calculate the possibility of fault
- Predictive analytics is one of the most promising examples of machine learning. It's applicable for everything; from product development to real estate pricing.

1.2 Stemmization:

Stemming in a cyberbullying detection project involves employing natural language processing (NLP) techniques to normalize words to their base or root form, known as the "stem."

This process aids in removing suffixes and prefixes from words, ensuring that variations of the same word are treated as identical for analysis purposes. Stemming offers language independence, enabling its application across different languages without requiring language-specific rules. Its speed and efficiency make it suitable for processing large volumes of text data typically found in social media platforms where cyberbullying often occurs. By focusing on content analysis, stemming allows detection systems to prioritize the semantic meaning of messages rather than specific linguistic variations. Stemmed words serve as foundational units for integration with other NLP techniques such as tokenization and sentiment analysis. While stemming offers customization and fine-tuning options to optimize detection systems, it also poses challenges and limitations, particularly in languages with complex morphological structures or irregular word formations. Therefore, careful evaluation of stemming's impact on system performance and the implementation of additional preprocessing steps may be necessary to mitigate potential inaccuracies.

1.3 Lemmatization:

Lemmatization, similar to stemming, is a natural language processing (NLP) technique used in cyberbullying detection projects. Unlike stemming, which simply removes suffixes and prefixes to normalize words to their root form, lemmatization aims to transform words to their base or dictionary form, known as the "lemma." This process considers the context and morphological analysis of words, resulting in more accurate normalization. In cyberbullying detection, lemmatization helps reduce the complexity of text data by ensuring that different forms of the same word are treated consistently during analysis, improving the performance of machine learning models. Additionally, lemmatization supports multilingual cyberbullying detection efforts, as it can adapt to various languages and their respective linguistic nuances. Integrating lemmatization with other NLP techniques enhances the system's ability to identify harassing or abusive content across different online communication platforms. However, as with stemming, careful evaluation of lemmatization's impact on system performance and potential errors in complex languages is essential for effective cyberbullying detection.

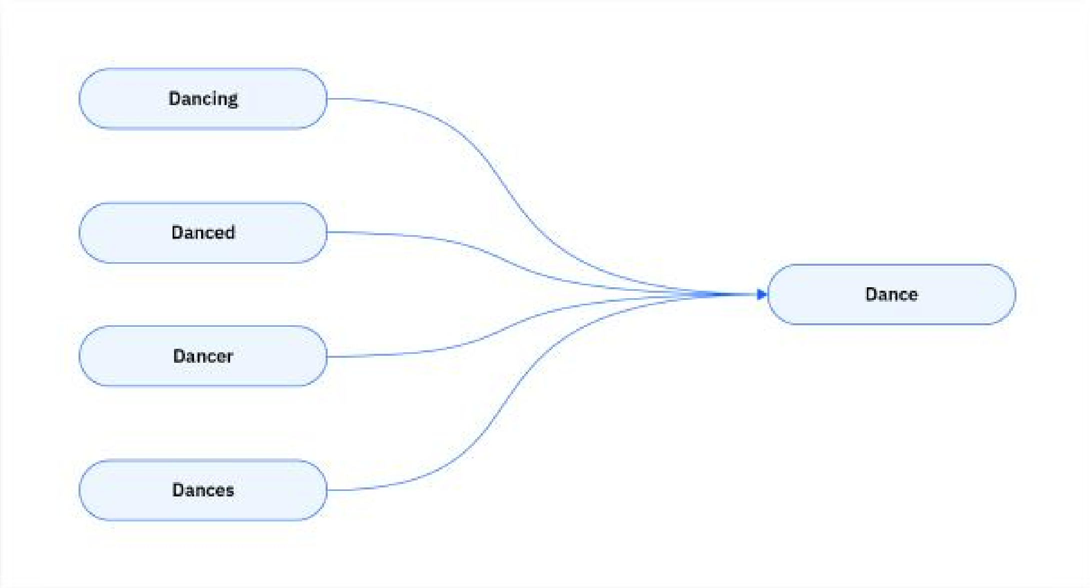

Fig-1: Stemmization and Lemmatization

1.4 Bag of Words:

The bag-of-words (BoW) method serves as a foundational technique in cyberbullying detection projects, simplifying the representation of text data for machine learning analysis. Initially, text undergoes preprocessing steps, including tokenization, lowercasing, and possibly stemming or lemmatization. Following this, BoW constructs a numerical vector for each document, where each dimension corresponds to a unique word in the corpus, and the value represents the frequency of that word in the document. These vectors serve as input features for training machine learning models, enabling them to discern patterns indicative of cyberbullying. Model training involves algorithms like logistic regression, support vector machines, or neural networks, which learn from the BoW representations to classify documents accurately. Subsequently, the system evaluates its performance using metrics such as accuracy, precision, recall, and F1-score, ensuring its effectiveness in distinguishing cyberbullying instances while minimizing false positives and false negatives. While the BoW method provides a straightforward means of text representation, it disregards word order and context, necessitating consideration of more advanced techniques for nuanced analysis in cyberbullying detection..

1.5 Tf - Idf (Term frequency - inverse document frequency) :

The term frequency-inverse document frequency (TF-IDF) method is a widely utilized technique in cyberbully detection projects for effectively representing text data and highlighting significant words in documents. TF-IDF calculates the importance of a word in a document relative to a collection of documents, considering both the frequency of occurrence within the document and the rarity of the word across the entire dataset. In the context of cyberbullying detection:

Firstly, text preprocessing tasks like tokenization, lowercasing, and potentially stemming or lemmatization are performed to prepare the text data. Following this, TF-IDF vectors are constructed for each document, where each dimension represents a unique word, and the value signifies the importance of that word in the document relative to the entire dataset. Words that appear frequently in a particular document but are rare across other documents are assigned higher TF-IDF scores, indicating their significance in characterizing the content of that document.

These TF-IDF vectors serve as input features for training machine learning models such as logistic regression, support vector machines, or neural networks. During model training, algorithms learn to identify patterns in the TF-IDF representations that distinguish cyberbullying content from non-cyberbullying content. The models utilize these learned patterns to classify new, unseen documents as either cyberbullying or non-cyberbullying.

After model training, the system evaluates its performance using metrics like accuracy, precision, recall, and F1-score to ensure its effectiveness in correctly identifying instances of cyberbullying while minimizing false positives and false negatives. The TF-IDF method enhances cyberbullying detection by highlighting important words that contribute significantly to the classification process, thereby improving the overall accuracy and reliability of the system.

While TF-IDF is a powerful technique for text representation, it may not fully capture semantic meaning or contextual information. Therefore, it is often used in conjunction with other NLP techniques or more advanced models to enhance cyberbullying detection capabilities further.

1.6 Multinomial Naive Bayes:

The Multinomial Naive Bayes (MNB) method is a prevalent approach in cyberbullying detection projects, chosen for its simplicity, efficiency, and effectiveness with text data. In this method, documents are represented as bags-of-words, and the algorithm assumes independence between features (words) within the documents. MNB calculates the probability of a document belonging to a specific class, such as cyberbullying or non-cyberbullying, based on the frequencies of words in the document and their associated probabilities in each class. This approach is particularly suitable for cyberbullying detection due to its ability to handle large datasets efficiently and its effectiveness in identifying patterns within text data. Additionally, MNB is relatively straightforward to implement and performs well even with limited training data, making it a popular choice for cyberbullying detection systems.

CHAPTER 2

REQUIREMENTS AND ANALYSIS

2.1 Hardware components:

- Processor: 64-bit, quad-core, 2.5 GHz minimum per core
- RAM: 4 GB or more.
- HDD: 20 GB of available space or more.
- Keyboard: A standard keyboard.

2.2 Software components:

- **Python:** Python offers concise and readable code. While complex algorithms and versatile workflows stand behind machine learning and AI, Python's simplicity allows developers to write reliable systems.
- **numpy:** NumPy is a Python library used for working with arrays.
- **sklearn:** Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering, and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy, and Matplotlib.
- **nltk:** NLTK(Natural Language Tool Kit) is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries
- **pandas:** Pandas is mainly used for data analysis and associated manipulation of tabular data in Dataframes. Pandas allow importing data from various file formats such as comma-separated values, JSON, Parquet, SQL database tables or queries, and Microsoft Excel.

- **re:** A regular expression (re) is a set of characters with highly specialized syntax that we can use to find or match other characters or groups of characters. In short, regular expressions, or Regex, are widely used in the UNIX world. The re-module in Python gives full support for regular expressions of Pearl style. The re module raises the re.error exception whenever an error occurs while implementing or using a regular expression.

- **sea:** Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps you explore and understand your data. Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.

- **Emoji:** Emoji Module is a Python package that allows us to use and print emojis through a Python program, and we can even use this module to use emojis inside an application we are creating using Python.

- **MatplotLib:** Matplotlib, a Python library for data visualization, aids cyberbullying detection by enabling analysts to create informative plots and charts. These visualizations help in understanding patterns such as word frequency distributions, sentiment trends, and other characteristics within text data. By visualizing these aspects, researchers can gain insights to inform the development and refinement of cyberbullying detection algorithms, facilitating effective data exploration and interpretation.

- **Windows OS 64-bit.**

2.3 Functional requirements

- A Machine learning model that detects bullying text efficiently.
- A model to pre-process the dataset.

2.4 Non-Functional requirements

- Response time.
- Maintainability.

CHAPTER 3

PROPOSED MODEL AND FLOW OF THE PROJECT

3.1 Proposed model

The proposed model aims to classify cyberbullying sentiments from cleaned tweet data using different word embedding techniques such as Bag of Words and TF-IDF. After preprocessing the text data through stemming, lemmatization, and a combination of both, the data is transformed into numerical vectors using CountVectorizer for Bag of Words and TfidfTransformer for TF-IDF representation. These representations are then used to train Multinomial Naive Bayes classifiers. The performance of the model is evaluated using classification reports and accuracy scores on both training and testing datasets. Additionally, learning curves are plotted to visualize the model's performance with varying training set sizes. The model's effectiveness in classifying cyberbullying sentiments is assessed based on these evaluations.

3.2 Flow of the project

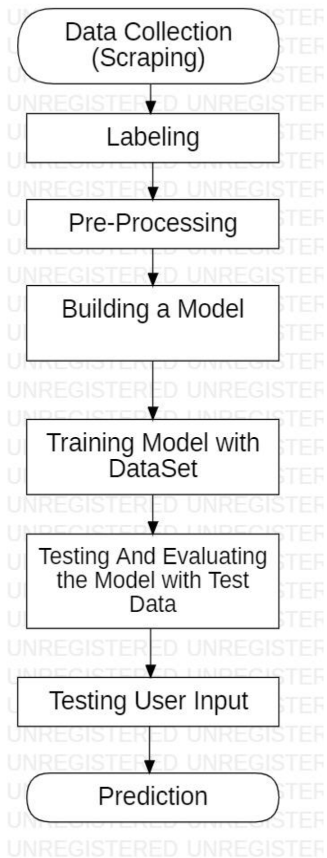

Fig-2: Flowchart for the Workflow of Project

3.3 Advantages and disadvantages

3.3.1 Advantages

- Can identify the bullying words.
- Helps to maintain social media bullying free.
- Useful to avoid negativity among people.

3.3.2 Disadvantages

- By using this model we can predict whether the given text is bullying text or not. But predictions may not be accurate.
- The text may be viewed from different perspectives, so it is difficult to predict accurately.
- For example Two friends use those bullying words in a sarcastic way, but our model predicts those texts as bullying texts. This is a major limitation.

3.4 Applications

- This can be used in the Social media Platforms for filtering bullying comments.
- This system can be used social applications to identify users who are bullying others and block them or take actions on them.
- We can use this system on kids platforms to keep them safe from cyberbullying.

CHAPTER 4

IMPLEMENTATION

4.1 Data Collection

The dataset for this project was collected from Kaggle, comprising multiple CSV files. These files were merged into a single dataset to facilitate analysis and model training. Merging the CSV files ensured a comprehensive dataset with a diverse range of cyberbullying tweets. The consolidation process involved concatenating the individual CSV files into one cohesive dataset, enabling seamless preprocessing and analysis. This merged dataset served as the foundation for exploring different preprocessing techniques, word embedding methods, and model training for the classification of cyberbullying sentiments.

	A	B	C
1	tweet_text	cyberbullying_type	
2	In other words #katandandre, your food was crapilicious! #m	not_cyberbullying	
3	Why is #aussietv so white? #MKR #theblock #ImACelebrityAU	not_cyberbullying	
4	@XochitlSuckkks a classy whore? Or more red velvet cupcake:	not_cyberbullying	
5	@Jason_Gio meh. :P thanks for the heads up, but not too coı	not_cyberbullying	
6	@RudhoeEnglish This is an ISIS account pretending to be a Ku	not_cyberbullying	
7	@Raja5aab @Quickieleaks Yes, the test of god is that good or	not_cyberbullying	
8	Itu sekolah ya bukan tempat bully! Ga jauh kaya neraka	not_cyberbullying	
9	Karma. I hope it bites Kat on the butt. She is just nasty. #mkr	not_cyberbullying	
10	@stockputout everything but mostly my priest	not_cyberbullying	
11	Rebecca Black Drops Out of School Due to Bullying:	not_cyberbullying	
12	@Jord_Is_Dead http://t.co/UsQInYW5Gn	not_cyberbullying	
13	The Bully flushes on KD http://twitvid.com/A2TNP	not_cyberbullying	
14	Ughhhh #MKR	not_cyberbullying	
15	RT @Kurdsnews: Turkish state has killed 241 children in last 1	not_cyberbullying	
16	Love that the best response to the hotcakes they managed to	not_cyberbullying	
17	@yasmimcaci @Bferrarii PAREM DE FAZER BULLYING COMIG(	not_cyberbullying	
18	@sarinhacoral @Victor_Maggi tadinhu de mim , sofrendo bul	not_cyberbullying	
19	@0xabad1dea @kelseytheodore2 twitter is basically the angr	not_cyberbullying	
20	Best pick up line? Hi, you're cute... ?: I love how people call Ja	not_cyberbullying	
21	Now I gotta walk to classss?! I officially hate the stupid bus sy	not_cyberbullying	
22	@halalcunty @biebervalue @liamxkiwi @greenlinerzjm @ha	not_cyberbullying	
23	Kids Love [illegible] @ Mohamad Bin Zayed City [illegible]	not_cyberbullying	

cyberbullying_tweets

Fig-3: Dataset CSV file

4.2 Data Preprocessing:

The data preprocessing stage involved several steps to prepare the collected cyberbullying tweet dataset for analysis and model training. Initially, the dataset was inspected for any duplicate entries, which were subsequently removed to ensure data integrity. Text cleaning procedures, including the removal of emojis, entities, contractions, hashtags, and special characters, were performed to standardize the text data. Additionally, stemming and lemmatization techniques were applied to reduce words to their root forms. Furthermore, the dataset was filtered to remove tweets exceeding 100 characters or containing less than one character. These preprocessing steps ensured the dataset was cleansed and ready for subsequent word embedding and model training processes.

	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...

Fig-4: After Stemmization and Lemmatization

4.3 Data Visualization:

The data visualization phase involved exploring and understanding the distribution of cyberbullying sentiments within the dataset. Using Python libraries such as Matplotlib and Seaborn, visual representations, including bar plots, were created to illustrate the balance or imbalance among different sentiment classes. These visualizations provided insights into the distribution of sentiments, allowing for a better understanding of the dataset's composition. Additionally, the visualization of tweet lengths using histograms provided an overview of the distribution of tweet lengths after basic cleaning, aiding in the identification of any potential outliers or patterns in tweet lengths.

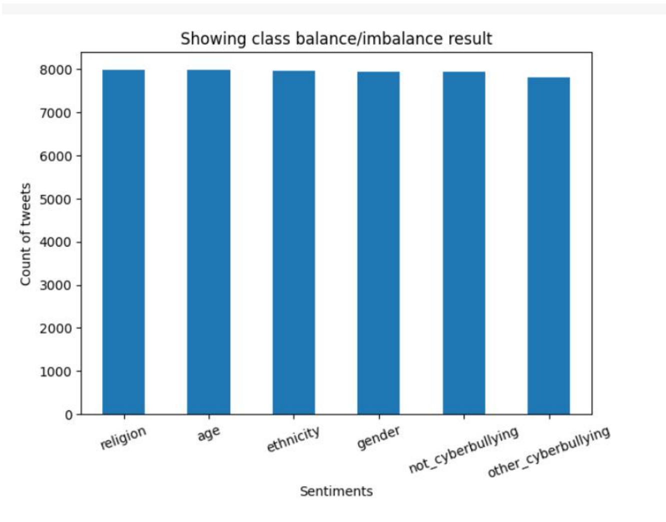

Fig-5: Balancing in Data

4.4 Multinomial Naive Bayes (MNB):

Multinomial Naive Bayes is a probabilistic classification algorithm commonly used in text classification tasks. It assumes that features are independent and follows a multinomial distribution. In the context of this project, Multinomial Naive Bayes was employed to classify cyberbullying sentiments based on preprocessed tweet data. By calculating the probability of each sentiment class given the features (word frequencies), the algorithm predicts the most likely sentiment class for a given tweet. Its simplicity, efficiency, and effectiveness in handling text data make Multinomial Naive Bayes a popular choice for text classification tasks, including sentiment analysis and spam detection.

4.4.1 Multinomial Naive Bayes using Bag of Words:

Bag of Words (BoW) is a common technique for text representation in natural language processing. It involves creating a vocabulary of unique words present in the dataset and representing each document as a vector of word counts or frequencies. Using BoW, the preprocessed tweet data was converted into numerical vectors, where each feature represents the presence or absence of a word in the document.

Multinomial Naive Bayes, a probabilistic classification algorithm, was then applied to the BoW representation of the text data. It calculates the probability of each sentiment class given the word frequencies in the document and predicts the class with the highest probability. This combination of BoW and Multinomial Naive Bayes is effective for text classification tasks, including sentiment analysis, due to its simplicity and efficiency in handling large text datasets.

For bag of words

```
[ ] y_train = np.array(y_train)
    model = 1
    accuracy_val = 0
    accuracy_train = 0
    kf = KFold(n_splits = 5)
    for train_index, val_index in kf.split(bow_cv1):
      X_KFold_train, X_KFold_val = bow_cv1[train_index], bow_cv1[val_index]
      Y_KFold_train, Y_KFold_val = y_train[train_index], y_train[val_index]
      mnb = MultinomialNB()
      mnb.fit(X_KFold_train, Y_KFold_train)
      Y_pred = mnb.predict(X_KFold_val)
      YY_pred = mnb.predict(X_KFold_train)
      print("Classification report for K-Fold model "+str(model))
      print("For validation set -")
      print(classification_report(Y_KFold_val, Y_pred))
      print("For training set -")
      accuracy_val += accuracy_score(Y_KFold_val, Y_pred)
      print(classification_report(Y_KFold_train, YY_pred))
      accuracy_train += accuracy_score(Y_KFold_train, YY_pred)
      model += 1
```

Fig-6: Naive bayes using bag of words

4.4.2 Multinomial Naive Bayes using TF - IDF:

Multinomial Naive Bayes using TF-IDF (Term Frequency-Inverse Document Frequency) is a text classification approach where the Bag of Words representation is enhanced by incorporating TF-IDF weights. TF-IDF reflects the importance of words in a document relative to a corpus, considering both the frequency of a term in a document and its inverse document frequency across the corpus.

TF-IDF vectors were created from the preprocessed tweet data, representing each document as a vector of TF-IDF weights. Multinomial Naive Bayes was then trained using these TF-IDF vectors to classify cyberbullying sentiments. By leveraging the TF-IDF representation, the classifier considers the importance of terms in each document, leading to potentially improved classification performance compared to using raw word counts.

For Tf-idf

```
y_train = np.array(y_train)
model = 1
accuracy_val = 0
accuracy_train = 0
kf = KFold(n_splits = 5)
for train_index, val_index in kf.split(tf_cv1):
  X_KFold_train, X_KFold_val = tf_cv1[train_index], tf_cv1[val_index]
  Y_KFold_train, Y_KFold_val = y_train[train_index], y_train[val_index]
  mnb = MultinomialNB()
  mnb.fit(X_KFold_train, Y_KFold_train)
  Y_pred = mnb.predict(X_KFold_val)
  YY_pred = mnb.predict(X_KFold_train)
  print("Classification report for K-Fold model "+str(model))
  print("For validation set -")
  print(classification_report(Y_KFold_val, Y_pred))
  print("For training set -")
  accuracy_val += accuracy_score(Y_KFold_val, Y_pred)
  print(classification_report(Y_KFold_train, YY_pred))
  accuracy_train += accuracy_score(Y_KFold_train, YY_pred)
  model += 1
```

Fig-7: Naive bayes using Tf - Idf

4.4.2 Conversion:

Language translation was utilized as part of the data preprocessing pipeline to ensure uniformity and consistency in the text data. Specifically, the language translation functionality was employed to detect the language of the input text and translate any non-English tweets into English. This was achieved using the Googletrans library, which provides access to the Google Translate API. By detecting and translating non-English tweets into English, the project aimed to create a standardized dataset for further analysis and model training, ensuring that all text data were in a consistent language format for effective processing and classification of cyberbullying sentiments.

```
[ ]  from googletrans import Translator
     from langdetect import detect
     def translate_telugu_to_english(telugu_text):
         translator = Translator()
         english_text = translator.translate(telugu_text, src='te', dest='en').text
         return english_text

     def process_text(input_text):
         detected_language = detect(input_text)

         if detected_language == "te":
             return translate_telugu_to_english(input_text)
         else:
             return input_text
```

Fig-8: Translating Language

OUTPUT

```
def process_text(input_text):
    detected_language = detect(input_text)

    if detected_language == "te":
        return translate_telugu_to_english(input_text)
    else:
        return input_text
```

```
[ ] user_txt = input("")
    text = process_text(user_txt)

    print(text.lower())
```

```
hindu is waste
hindu is waste
```

```
# Print transformed input
print("Transformed input:", new_input)

# Use the trained classifier to make predictions on new input
user_prediction1 = mnb1.predict(new_input)

# Print the raw prediction
print("Raw prediction:", user_prediction1)

# Map the prediction to the class name
predicted_class_name = class_mapping[user_prediction1[0]]
print("Predicted class:", predicted_class_name)
```

```
Transformed input:
Raw prediction: [0]
Predicted class: religion
```

Fig-9: Predicting output

CHAPTER 5

CONCLUSION

Data is collected by web scrapping. Collected Data is preprocessed using python functions. Model is built, trained and tested using the preprocessed data. The model takes input from the user and classifies it into different categories.

In particular, cyberbullying has become more common and has begun to raise significantly. Also, usage of social media is rapidly increasing. So, there is a need for automatic cyberbullying detection features on social media platforms. This model predicts whether the given input is a cyberbullying text or not. But it can be improved with advanced algorithms and methods to overcome the limitations and can be used on different platforms as a filter or a mechanism to detect and take measures to reduce cyberbullying on the internet.

REFERENCES

[1] Md Manowarul Islam; Md Ashraf Uddin; Linta Islam; Arnisha Akter; Selina Sharmin; Uzzal Kumar Acharjee , *"Cyberbullying Detection on Social Networks Using Machine Learning Approaches"*; 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 16-18 December 2020.

[2] Rohith Gandhi, *"Naive Bayes Classifier"*; Towards Data Science, May 5, 2018.

[3] Joe Tran, *"Random Forest Classifier in Python"*; Towards Data Science, May 2, 2020.

[4] Nikolai Liubimov, *"Introducing Label Studio, a swiss army knife of data labeling"*; Towards Data Science, Jan 28, 2020.

[5] Priyanka Dave, *"How To Scrap YouTube Comments?"*; Analytics Vidhya, Mar 27, 2021.