DETECTION OF CYBERBULLYING ON SOCIAL MEDIA PLATFORM

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF **BACHELOR OF ENGINEERING** IN **COMPUTER SCIENCE AND ENGINEERING** OF THE ANNA UNIVERSITY

MINI PROJECT WORK

2022

Submitted by

JAYANTHI T

1817119

KIRUTHIKA S

1817125

VARUNA PRIYA S

1817153

YAZHINI P

1817157

Under the Guidance of

Dr. R. BHAVANI M.E., Ph.D.,

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GOVERNMENT COLLEGE OF TECHNOLOGY

(An Autonomous Institute affiliated to Anna University)

COIMBATORE - 641 013

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GOVERNMENT COLLEGE OF TECHNOLOGY

(An Autonomous Institution Affiliated to Anna University)

COIMBATORE - 641 013

MINI PROJECT WORK

JANUARY 2022

This is to certify that this project work entitled

DETECTION OF CYBERBULLYING ON SOCIAL MEDIA PLATFORM

is the bonafide record of project work done by

JAYANTHI T
1817119

KIRUTHIKA S
1817125

VARUNA PRIYA S
1817153

YAZHINI P
1817157

of B.E.(COMPUTER SCIENCE AND ENGINEERING) during the year 2021 - 2022

Project Guide
Dr.R.BHAVANI M.E., Ph.D.,

Head of the Department
Dr.J.C.MIRACLIN JOYCE PAMILA M.E.,Ph.D.,

Submitted for the Project Viva-Voce examination held on _____

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

We express our sincere gratitude to **Dr.P.THAMARAI Ph.D.,** Principal, Government College of Technology, Coimbatore for providing us with all facilities that we needed for the completion of this project.

We wholeheartedly express our thankfulness and gratitude to **Dr.J.C.MIRACLIN JOYCE PAMILA M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering, Government College of Technology, for helping us to successfully carry out this project.

Our thankfulness and gratitude to our respectable project guide **Dr.R.BHAVANI M.E., Ph.D.,** Assistant Professor of the Department of Computer Science and Engineering and the panel members **Dr.K.KUMAR M.E., Ph.D.,** Assistant Professor (CAS) and **Mr.N. ARUMUGAM M.E.,** Assistant Professor(Contract) for their valuable suggestions throughout the various phases of the project.

We extend our sincere thanks to **Dr.J.C.MIRACLIN JOYCE PAMILA M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering, **Dr.S.RATHI M.E., Ph.D.,** Professor, **Prof.T.RAJA SENBAGAM M.E.,** Assistant Professor, **Dr.A.MEENA KOWSHALYA M.E., Ph.D.,** Assistant Professor, **Dr.R.MUTHURAM M.E., Ph.D.,** Assistant Professor, and **Prof.L.SUMATHI M.E.,** Assistant Professor for their support in the completion of the project.

We thank all the non-teaching staff and our friends for their cooperation towards the successful completion of the project.

We would like to dedicate the work to our parents for their constant encouragement throughout the project.

SYNOPSIS

National Crime Prevention Council defines cyberbullying as “The use of the Internet, cell phones or other devices to send or post text or images intended to hurt or embarrass another person”. The effects of cyberbullying also include mental health issues, increased stress and anxiety, depression, acting out violently and low self-esteem.

Detection of cyberbullying in social media is a challenging task. Across various social media platforms, cyberbullies attack victims on different topics such as race, religion and gender.

This project aims to detect bullying posts or comments on social media platform using convolutional neural network(CNN). The CNN model is trained using the posts/comments collected from twitter tweets. The dataset was split into 80:20 for training and testing respectively. The model was able to detect cyberbullying with an accuracy of 76.09%.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABBREVIATION	DEFINITIONS
CNN	CONVOLUTIONAL NEURAL NETWORKS
SMP	SOCIAL MEDIA PLATFORM
GLOVE	GLOBAL VECTORS
LSTM	LONG SHORT TERM MEMORY
SSWE	SENTIMENT SPECIFIC WORD EMBEDDING
NLP	NATURAL LANGUAGE PROCESSING
BLSTM	BIDIRECTIONAL LONG SHORT TERM MEMORY
DNN	DEEP NEURAL NETWORKS
LSA	LATENT SEMANTIC ANALYSIS
NLTK	NATURAL LANGUAGE TOOL KIT
CNN-CB	CONVOLUTIONAL NEURAL NETWORKS CONTENT BASED

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

Cyberbullying is so dangerous because it gives the ability to harass anyone in public at any time through devices and mostly, students are the victims of these kinds of harassment. Cyberbullying had the impact of amplifying symptoms of depression and post-traumatic stress disorder in young people who were inpatients at an adolescent psychiatric hospital, according to a new study published in the Journal of Clinical Psychiatry. The study addressed both the prevalence and factors related to cyberbullying in adolescent inpatients.

Ritu Kohli's[5] Case was the first cyberbullying case reported in India. A girl named Ritu Kohli filed a complaint in 2001 that someone else is using her identity in social media and she was deliberately getting calls from different numbers. She was also getting calls from abroad. A case was also filed under Section 509 of the Indian penal code. In the case of Prakhar Sharma[6] of Madhya Pradesh, the accused created a fake Facebook account of the victim and posted some vulgar messages. Cyberbullying is likely to cause destructive psychological effects, like low self-esteem, mental depression, suicide consideration and even suicide[9].

A fatal cyberbullying incident had happened on MySpace SNS[10], whereby Megan Meier, a 13-year-old teen became increasingly distressed by the online harassment being directed at her, and eventually decided to end her life by hanging herself in her bedroom in 2006. Even famous YouTubers receive negative comments publicly and privately on a daily basis. Some of these are extremely vicious to the point where they could be considered death threats. Singer Rebecca Black, who got her fame through her “Friday” music video on YouTube, was forced to quit middle school after being severely ridiculed for it.

1.2 PROBLEM STATEMENT

A system for automatic detection of cyberbullying considering the main characteristics of cyberbullying such as intention to harm an individual, repeatedly and over time, and using abusive curl language or hate speech is the need of the hour. The system should rely on the detection of cyberbullying text along with the themes/categories associated with cyberbullying such as racist, sexual, physical mean, swear and others. Most of the studies have considered calling someone stupid, ugly and idiot as cyberbullying. Things have changed, such words may or may not always be a bullying incident. If a person wants intentionally to harm an individual, they will use extreme words. So, detecting such activities is challenging as well as important to maintain people's mental stability.

1.3 PROPOSED SOLUTION

Most of the existing studies have approached this problem with machine learning models. In recent studies, deep learning based models have found their way in the detection of cyberbullying incidents, claiming that they can overcome the limitations of the machine learning models and improve the detection performance. This project proposes a CNN model to increase the performance compared to other machine learning methods. The tweets are pre-processed subjecting it to standard operations of removal of stop words, punctuation marks, stemming and lowercasing. After converting the pre-processed data into compatible form, the CNN model will be applied to this data in order to distinguish the tweets as either bully, non-bully, racism or sexism.

CHAPTER 2

LITERATURE SURVEY

[1] This paper has used convolutional neural network cyberbullying detection (CNN-CB) algorithm, which remedy the current unsolved problems. The primary goal is to develop an efficient detection approach capable of dealing with semantics and meaning and produces accurate results while keeping computational time and cost to a minimum. CNN-CB is based on deep learning which is built upon the concept of CNN which showed great success when applied to many classification tasks. The most remarkable contribution is that CNN-CB is a cyberbullying detection algorithm that has shortened the classical detection workflow. It transforms text into word embedding and feeds them to a CNN-CB.

[2] In this paper, the proposed technique has used LSTM. The model architecture is comprised of an input layer after which there is a SSWE embedding layer followed by a LSTM layer where the return sequences were made to be positive. One more LSTM layer but on this one, the return sequences are made to be negative and a dropout layer having the probability of 0.5 then followed by a dense layer containing two output classes that makes use of the softmax activation function.

[3] In the proposed system, convolution neural network is used to create a model that detect bully related tweets and predict the behavior of the new data introduced. The proposed approach use word vectors that are fed to the CNN for classification of tweets. The architecture of CNN comprises of an input, output layer and a few of hidden layers. The input layer includes succession of vectors. It is examined utilizing fixed size of filter. The filter shifts or strides only one row or one column on the matrix. Each filter distinguishes different features in the content so as to portray it into the feature map the next layer is max-pooling layer. The maxpooling layer minimizes the features in the feature map.

[4] The proposed method experimented with four DNN based models for cyberbullying detection: CNN, LSTM, BLSTM and BLSTM with attention. The embedding layer processes a fixed size sequence of words. Each word is represented as a real-valued vector, also known as word embedding. During the

training, model improves upon the initial word embedding to learn task specific word embedding. Using GloVe vectors over random vector initialization has been reported to improve performance for some NLP tasks. SSWE method overcomes this problem by incorporating the text sentiment as one of the parameters for word embedding generation.

Reference	Techniques used	Limitations	Advantages
[1]	Word embedding and CNN-CB	Generate huge number of features that require careful feature extraction and selection phases which lead to computational overhead.	This model gives better results in three metrics accuracy, precision and recall.
[2]	Word embedding and LSTM	The model was not able to correctly classify tricky sentences.	While testing our model, the test data goes through model to build hidden states which then proceed onto transitional parts for the final result.
[3]	CNN model	The number of features exceed the number of training data samples	It operates through many layers and gives accurate classification.
[4]	CNN, LSTM, BLSTM, and BLSTM with attention	The models can be further improved depending on the perceived seriousness of the post.	The models coupled with transfer learning beat state of the art results for all datasets.

Table 2.1 LITERATURE SURVEY

CHAPTER 3

PROJECT DESIGN

3.1 PROJECT WORKFLOW

CNN is an algorithm that advances current work in cyberbullying detection by adapting principles of deep learning instead of classical machine learning. CNN architecture consists of four layers: embedding, convolutional, max pooling and dense which will be described in the following sub sections. Its remarkable aspect is that it eliminates three classification phases previously employed by other detection algorithms, feature determination, extraction and selection. This is achieved through generating word and selection embedding (numerical vectors) for each word in a tweet and feeding them directly to a convolutional neural network.

In the first step, data is fetched from Twitter platform that contains some bully and non bully tweets and the dataset is created. Then the dataset is preprocessed with sequence of preprocessing steps like case folding ,stop words removal, stemming and tokenization. In the next step, the preprocessed tweets are converted to real-valued vectors using Glove which can be abbreviated as global vectors. Now, the convolutional neural network model with series of layers like convolution, max-pooling, flatten, dropout and dense layers are applied to train and test the dataset. Finally in the output, if tweets doesn't contain any offensive words, then it is classified as non-bully class. If any such offensive words is present, then the tweet is classified as either bully, sexism or racism.

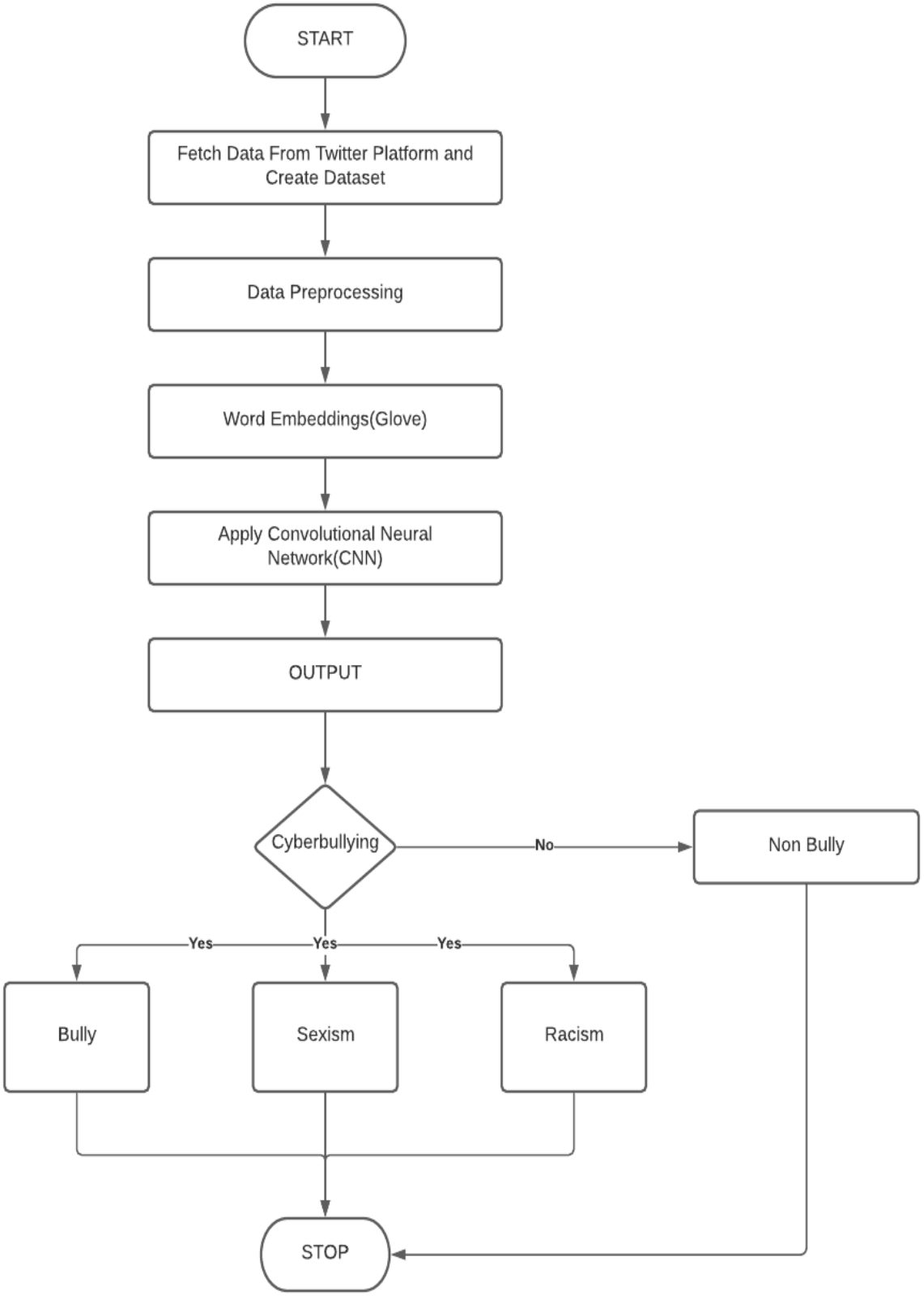


Fig 3.1 WORKFLOW DIAGRAM OF CYBERBULLYING DETECTION

3.1.1 DATASET PREPROCESSING

The noncontributory characters such as punctuation marks, blank spaces, symbols, numbers, and other various text that do not contribute towards classifying a piece of text as bullying were eliminated from the dataset. First, all text data are converted to lowercase. Then some words like “what’s” or “can’t” are converted to “what is” or “can not”. Also, all the punctuations are removed using the string library. For example, “@MalikHasanImani: @1Bcarter negro, you wear a size 5...& I'm at a 6 now still growing” will be pre-processed as “negro wear size still grow”.

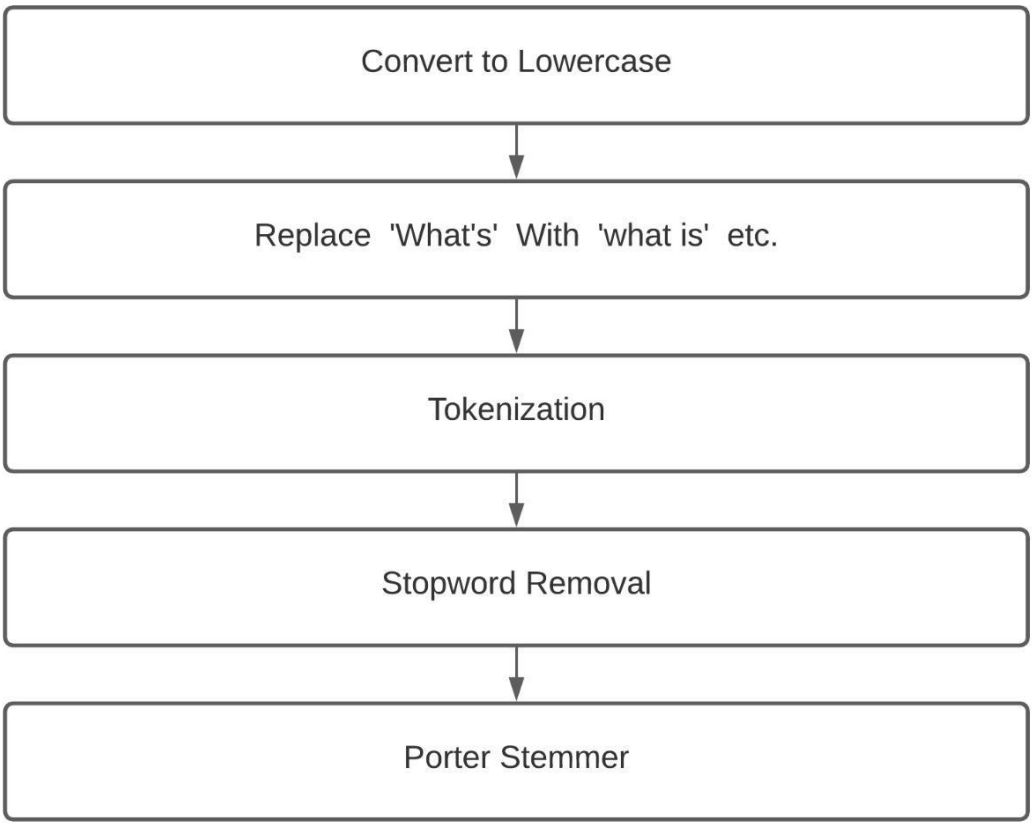


Fig 3.2 DATA PREPROCESSING PIPELINE

Tokenization:

In tokenization we split raw text into meaningful words or tokens. For example, the text “we will do it” can be tokenized into ‘we’, ‘will’, ‘do’, ‘it’. The tokenization can be performed at the sentence level or at the word level or even at the character level. Tokenization has many more variants but this project uses a regular expression tokenizer. In this tokenizer, tokens are decided based on the regular expression.

Stemming:

Stemming is the process of converting a word into a root word or stem. For the three words ‘eating’ ‘eats’ ‘eaten’, the stem is ‘eat’. Since all the three branch words of root ‘eat’ represent the same thing it should be recognized as similar. NLTK offers 4 types of stemmers: Porter Stemmer, Lancaster Stemmer, Snowball Stemmer and Regexp Stemmer. This project uses Porter Stemmer as it is known for its simplicity and speed.

Stop word removal:

Stop words are words that do not add any meaning to a sentence. Some stop words for English language are: what, is, at, a, etc., These words are irrelevant and can be removed. NLTK contains a list of English stop words that can be used to filter out all the tweets. Stop words are often removed from the text data when the deep learning models are trained since the information they provide is irrelevant to the model and helps in improving performance.

3.1.2 WORD EMBEDDING

In the proposed CNN model, the pre-trained embedding like GloVe and the embedding layer provided by Keras are used. GloVe proposes to learn word embedding directly from an aggregated global word-word co-occurrence matrix. Thus, it is easier both in time and resource to compute. The use of word embedding made CNN’s more advanced compared to the traditional detection approaches since they incorporate semantics not just features extracted from raw text. Keras embedding layer requires three parameters to be set prior to the construction of the vector space:

1) **Input dimension:** Specifies the total number of words in the vocabulary (whole corpus). This number is derived from the following. Let T be all tweets in the corpus. T= {t1,t2,t3..tn}, n=number of tweets.

Input dimension = length (Tokenized (T))

2) **Output dimension:** Specifies the size of the output vector from this layer.

3) **Input length:** The length of each vector (maximum number of words per tweet). Twitter fixed maximum tweet length was not set, since this might change over time. Input length is calculated by using the function

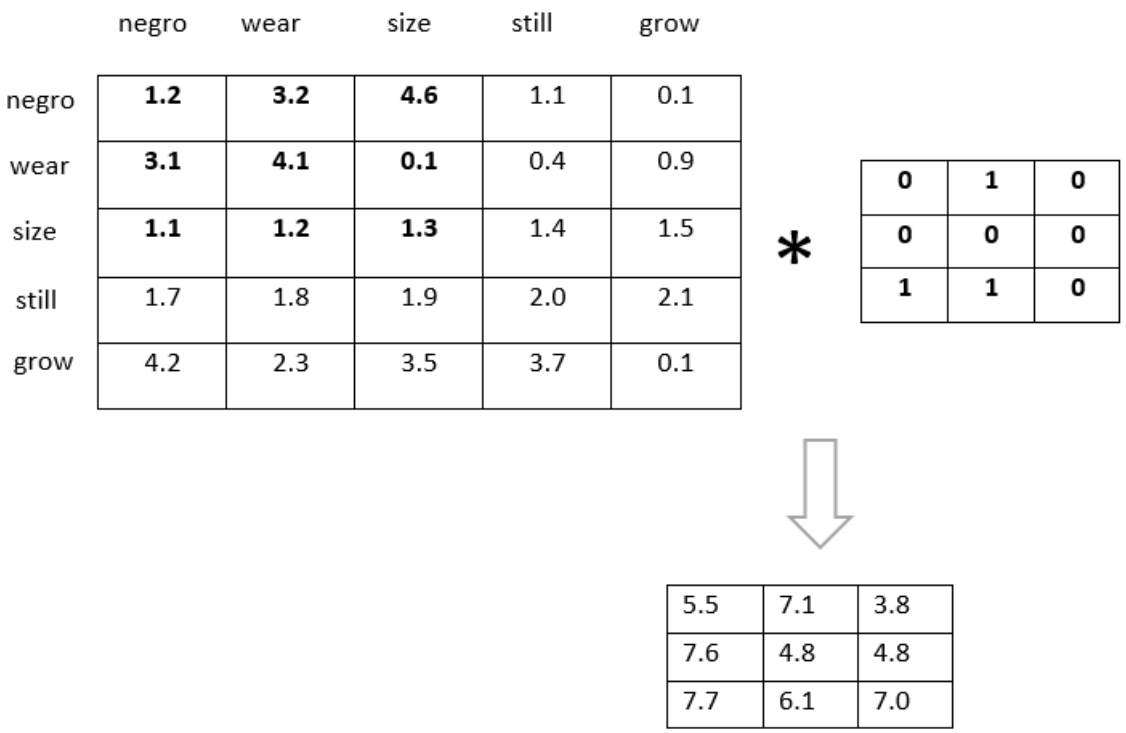
Input length = max (length for t in T)

For example, the pre-processed data from the previous example will be embedded as

	negro	wear	size	still	grow
negro	1.2	3.2	4.6	1.1	0.1
wear	3.1	4.1	0.1	0.4	0.9
size	1.1	1.2	1.3	1.4	1.5
still	1.7	1.8	1.9	2.0	2.1
grow	4.2	2.3	3.5	3.7	0.1

3.1.3 CONVOLUTIONAL LAYER

The second layer after the embedding layer is the convolutional layer. It is the heart of a convolutional neural network. Its task is to convolve around the input vector to detect features, therefore, it compresses the original input vector while preserving valuable features. This is achieved by creating a set of matrices called filters of random numbers called weights.[1] Each filter is then independently convolved around the original input vector creating many feature maps through element-wise multiplication with the part of the input. For the chosen example, the vectors from the embedding layer perform dot product with a filter having a size of 3.



3.1.4 MAX POOLING LAYER

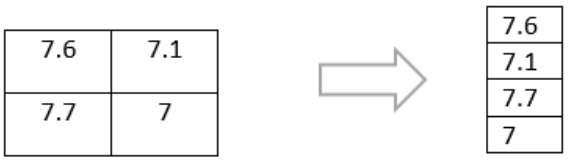
Max pooling matrix simply slides across the output of a convolutional layer and finds the maximum value of the selected area. It gives robustness and the ability to deal with complex data like images and large corpus, by compressing the input to smaller matrices.[1] This remarkable ability is achieved by both convolutional and max-pooling layers. Thus, they are used after one another. In this way, only meaningful and clear features are preserved. For the chosen example, the max-pooling layer selects the maximum element from the output of the convolutional layer.

3.1.5 FLATTEN LAYER

Flattening is converting the data into a 1-dimensional array for inputting it to

the next layer. The output of the convolutional layers is flattened to create a single long feature vector. And it is connected to the final classification model, which is called a fully-connected layer.

For the example that is chosen, the flatten layer converts the output from the max-pooling layer into one long vector.



3.1.6 DROPOUT LAYER

The Dropout layer randomly sets input units to 0 with a frequency of rate at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by 1/(1 - rate) such that the sum over all inputs is unchanged. During training, some number of layer outputs are randomly ignored or “dropped out.” This has the effect of making the layer look like a layer with a different number of nodes and connectivity to the prior layer.

3.1.7 DENSE LAYER

All layers described so far were concerned with shaping data and compressing them in a meaningful way. So far, no classification has been done. This is exactly the job of dense layers.[1] As in a neural network, dense layers are a set of fully connected layers. The number of dense layers varies, however, the last one must have 4 neurons corresponding to the number of classes in this case.

In this layer, the activation function used is softmax activation. The softmax activation function gives a probability value as output. Based on the probability value, the label of the tweet is identified and the tweet is classified as either bully, non-bully, racism and sexism. The equation for softmax function is

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Where

σ - softmax

$\vec{z}$ - input vector

z_i - standard exponential function for input vector

K - number of classes in the multi-class classifier

z_j - standard exponential function for output vector

The optimizer used is adam optimizer as it is computationally efficient, has little memory requirement, and is well suited for problems that are large in terms of data/parameters. This project uses the categorical cross-entropy as the loss function since this is a multi-class classification. The formula for categorical cross-entropy is

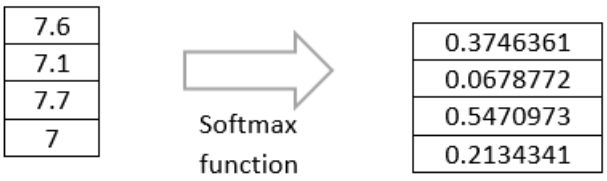
$$-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

Where M - number of classes

y - binary indicator (0 or 1) if class label c is the correct classification for observation o

p - predicted probability observation o is of class c

For the chosen example the dense layer predicts the output in probability values by using the softmax activation function.



3.2 DEEP LEARNING IN CYBERBULLYING DETECTION

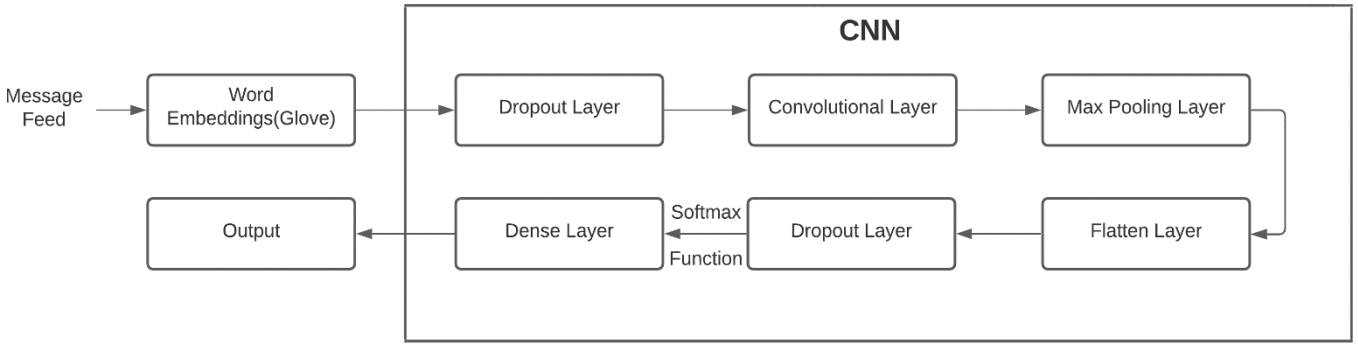


Fig 3.3 Proposed CNN Model

The tweets in the dataset which is the input are converted into meaningful data by using pre-processing techniques like removing spaces, symbols, numbers, stop words and punctuations. Word Embedding is used to change the sequence of words that are already pre-processed into a vector representation.

The output of the word embedding layer is given to the dropout layer which prevents overfitting. The convolutional layer gets the input from the dropout layer and performs the dot product with the filter. The convolutional layer gives the input to the max-pooling layer which selects the maximum element. The flatten layer converts the whole vector into one long vector.

The output dimensions were flattened using a sigmoid function. Lastly, the fully connected layers and the activation function on the outputs will give values for each class. The equation for sigmoid function is

$$\text{Sigmoid } S(x) = \frac{1}{1 + e^{-x}}$$

The CNN model can be improved by adding more layers. It is always preferred to have more(dense) layers than to have wide layers of less number. But the data should not be overfitted which can be avoided by using various regularization methods. To test the performance of our model, we took 80% of the dataset as training set and 20% of it for testing.

CHAPTER 4

IMPLEMENTATION

4.1 SOFTWARE ENVIRONMENT

4.1.1 JUPYTER NOTEBOOK

The Jupyter Notebook is an open-source web application that allows users to create and share documents that integrate live code, equations, computational output, visualizations, and other multimedia resources. A Jupyter Notebook is an easy-to-use, interactive data science environment that not only works as an integrated development environment (IDE), but also as a presentation or educational tool. Jupyter is a way of working with Python inside a virtual “notebook” and is growing in popularity with data scientists in large part due to its flexibility. It allows to combine code, images, plots, comments, etc., Further, it is a form of interactive computing, an environment in which users execute code, see what happens, modify, and repeat in a kind of iterative conversation between the user and data.

4.1.2 PROGRAMMING LANGUAGES USED

Python 3.8.6

LIBRARIES USED

- Keras
- Numpy
- Pandas
- Sci-kit learn
- nltk
- plotly

4.2 DATASET COLLECTION

The dataset is prepared with large, diverse, manually annotated and publicly available data for cyberbullying detection in social media. Variation in the number of posts across the dataset also affects vocabulary size that represents the number of distinct words encountered in the dataset. The size of a post is measured in terms of the number of words in the post. For each dataset, there are only a few posts with large sizes. Such large posts are truncated to the size of posts ranked at 95 percentage in the dataset.

Twitter:

Twitter is an online news and social networking site where people communicate in short messages called tweets. Cyberbullying is especially present on Twitter. According to data from the Pew Center, Twitter users face many forms of harassment including death threats and threats of sexual abuse and stalking and the victims of this abuse are disproportionately women. There have been several recent high-profile cases of cyber-bullying involving Twitter including #gamergate, the harassment of Robin William's daughter after his death, and Ashley Judd's decision to press charges against trolls.

This dataset includes 7000 annotated tweets. While collecting tweets and creating the dataset, we bootstrapped the corpus collection, by performing an initial manual search of common slurs and terms used pertaining to religious, sexual, gender, and ethnic minorities.

4.3 PERFORMANCE METRICS

The performance of the proposed model is measured using precision, recall, F1 score and accuracy. The equations of the metrics are,

METRIC	EQUATION
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1 score	$\frac{2(precision \times recall)}{precision+recall}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$

Where

TP – True Positive, TN – True Negative

FP – False Positive, FN – False Negative

The TP,TN,FP,FN values for all classes in this project are,

	0	**1**	**2**	**3**
TP	38	13	5	12
FN	12	37	45	38
FP	93	6	14	19
TN	57	144	136	131

Here 0,1,2 and 3 represents the classes bully, non-bully, racism and sexism respectively.

CHAPTER 5

CONCLUSION AND FUTURE ENHANCEMENTS

Cyberbullying across the internet is dangerous and leads to mishappening like suicides, depression etc., and therefore there is a need to control its spread. Therefore, cyberbullying detection is vital on social media platforms.

From suicide to lowering victims‟ self-esteem, cyberbullying control has been the focus of many psychological and technical research. Cyberbullying detection can be used on social media websites to ban users from trying to take part in such activity. In this project, a deep learning model is developed for the detection of cyberbullying to combat the situation. The Cyberbullying detection in the deep learning model had an accuracy of 62%. Often, the datasets for cyberbullying detection contain very few posts marked as bullying. The imbalance can be partly covered by oversampling the bullying posts. The results can be improved further by improving the dataset by having as much relevant data as possible and reducing the redundant data.

CHAPTER 6

SCREENSHOTS

PRE-PROCESSING

	label	tweets
0	sexism	"I want equal rights but I still want your se...
1	sexism	"Sassy - halfway between slut and classy" #MK...
2	non-bully	@GBabeuf @RJennromao @DavidJo52951945 @Novoro...
3	sexism	@GMShivers claims that only people new to the...
4	sexism	@p4ndiamond I saw him but I rarely engage mal...
...	...	...
7371	non-bully	@XaiaX I would rather it didn't. Can't we just...
7372	non-bully	Guy with the cap and Pete said.\nBye Kat and M...
7373	non-bully	And why is NBA 2K15 in my steam library? I wou...
7374	non-bully	RT @PlayHearthstone: Roses are red, \nColdligh...
7375	non-bully	@cigardubey right now? multiple browsers :P ...

7376 rows × 2 columns

Fig 6.1 Dataset before pre-processing

	label	tweets
0	sexism	want equal right still want seat bu still pay ...
1	sexism	sassi halfway slut classi mkr mkr
2	non-bully	truth understand putin crimin
3	sexism	claim peopl new industri claim sexism exist ya...
4	sexism	saw rare engag male fem zero point follow order
...	...	...
7371	non-bully	would rather becom island
7372	non-bully	guy cap pete said bye kat minion mkr
7373	non-bully	nba k steam librari would never buy game
7374	non-bully	rose red coldlight blue murloc bad poet mrrglg...
7375	non-bully	right multipl browser p set tweetdeck

7376 rows × 2 columns

Fig 6.2 Dataset after pre-processing

TRAINING

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 250, 100)          727300
_________________________________________________________________
dropout (Dropout)            (None, 250, 100)          0
_________________________________________________________________
conv1d (Conv1D)              (None, 248, 128)          38528
_________________________________________________________________
max_pooling1d (MaxPooling1D) (None, 124, 128)          0
_________________________________________________________________
flatten (Flatten)            (None, 15872)             0
_________________________________________________________________
dropout_1 (Dropout)          (None, 15872)             0
_________________________________________________________________
dense (Dense)                (None, 4)                 63492
=================================================================
Total params: 829,320
Trainable params: 829,320
Non-trainable params: 0
```

Fig 6.3 Model Summary

```
epochs = 25
batch_size = 32

history = model.fit(X_train, Y_train, epochs=epochs, batch_size=batch_size, validation_split=0.1, callbacks=[EarlyStopping(monit
```

```
Epoch 1/25
146/146 [==============================] - 5s 34ms/step - loss: 0.9476 - accuracy: 0.6074 - val_loss: 0.7681 - val_accuracy: 0.
6828
Epoch 2/25
146/146 [==============================] - 5s 35ms/step - loss: 0.7448 - accuracy: 0.6989 - val_loss: 0.7369 - val_accuracy: 0.
6847
Epoch 3/25
146/146 [==============================] - 5s 37ms/step - loss: 0.6862 - accuracy: 0.7277 - val_loss: 0.7605 - val_accuracy: 0.
6983
Epoch 4/25
146/146 [==============================] - 7s 50ms/step - loss: 0.6369 - accuracy: 0.7482 - val_loss: 0.7371 - val_accuracy: 0.
7215
Epoch 5/25
146/146 [==============================] - 7s 46ms/step - loss: 0.6082 - accuracy: 0.7551 - val_loss: 0.7355 - val_accuracy: 0.
7021
Epoch 6/25
146/146 [==============================] - 7s 50ms/step - loss: 0.5690 - accuracy: 0.7749 - val_loss: 0.7436 - val_accuracy: 0.
7331
Epoch 7/25
146/146 [==============================] - 7s 46ms/step - loss: 0.5387 - accuracy: 0.7891 - val_loss: 0.7511 - val_accuracy: 0.
7253
Epoch 8/25
146/146 [==============================] - 7s 46ms/step - loss: 0.5291 - accuracy: 0.7949 - val_loss: 0.7577 - val_accuracy: 0.
7273
```

Fig 6.4 Model Training

```
Is this retarded bitch really this stupid?
[[0.38660753 0.04153068 0.05419505 0.5176667 ]] bully


Woo can't wait to see what happens!!! #mkr
[[0.00625993 0.6444721  0.00532781 0.34394014]] non-bully


"575597554391457793     Someone get Kat a straight Jacket!! The bitch needs some time in a padded room!! #mkr    "
[[4.6976708e-02 2.0819951e-02 1.9904875e-04 9.3200427e-01]] sexism


@ArarMaher Kind of the same thing the prophet Mohammed did.
[[7.1317825e-04 4.9398992e-02 9.4877207e-01 1.1157405e-03]] racism
```

Fig 6.5 Output Prediction

```
acc = model.evaluate(X_test, Y_test)
print('Accuracy: %f' % (acc[1]))
```

```
70/70 [==============================] - 1s 8ms/step - loss: 0.6524 - accuracy: 0.7610
Accuracy: 0.760958
```

Fig 6.6 Training Accuracy

TESTING

```
Actual Class Label
['0', '1', '0', '1', '1', '0', '1', '1', '1', '1', '1', '2', '2', '2', '2', '2', '1', '3', '3', '3', '3', '3', '3', '0', '1',
'0', '1', '0', '1', '0', '1', '0', '1', '2', '2', '2', '2', '3', '3', '3', '3', '2', '2', '1', '1', '1', '1', '1', '0', '1']
```

Fig 6.7 Actual samples taken for testing

```
Predicted Class Label
['1', '3', '1', '1', '3', '1', '1', '1', '2', '2', '1', '1', '1', '3', '3', '3', '3', '3', '1', '3', '1', '0', '1', '2', '1',
'0', '1', '2', '2', '2', '1', '3', '3', '3', '3', '2', '2', '1', '1', '1', '1', '1', '2', '1', '3', '1', '3', '1', '3', '3']
```

Fig 6.8 Predicted samples

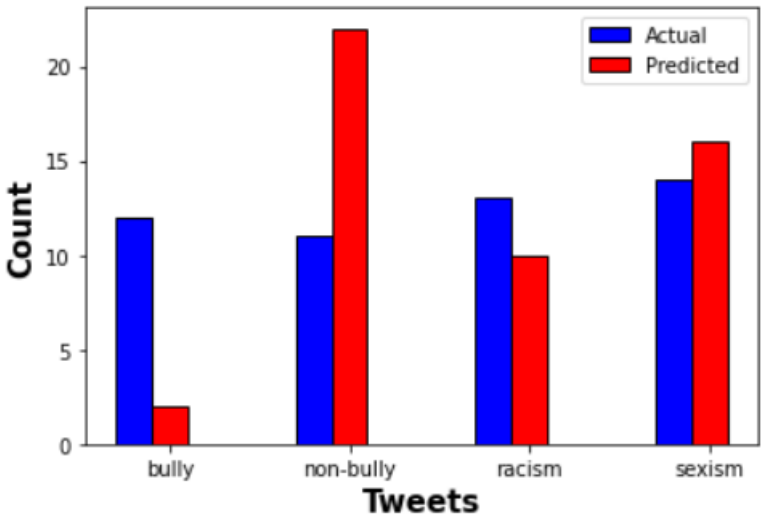


Fig 6.9 Actual vs Predicted Graph

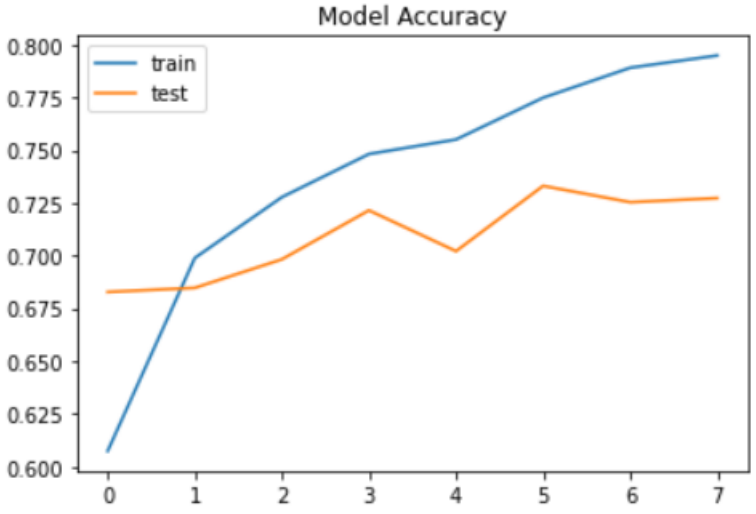


Fig 6.10 Model Accuracy Plot

```
Confusion Matrix

[[38 0 10 2]
 [26 13 2 9]
 [35  2 5 8]
 [32  4 2 12]]
```

Fig 6.11 Confusion Matrix

```
Classification Report

0:bully
1:non-bully
2:racism
3:sexism
```

	precision	recall	f1-score	support
0	1.00	0.17	0.29	12
1	0.41	0.82	0.55	11
2	0.80	0.62	0.70	13
3	0.75	0.86	0.80	14
micro avg	0.62	0.62	0.62	50
macro avg	0.74	0.61	0.58	50
weighted avg	0.75	0.62	0.59	50
samples avg	0.62	0.62	0.62	50

Fig 6.12 Classification Report

```
from sklearn.metrics import accuracy_score
accuracy=accuracy_score(y_true,y_pred)
print("Accuracy : ",accuracy)
```

```
Accuracy :  0.62
```

Fig 6.13 Testing Accuracy

CHAPTER 8

REFERENCES

[1] Al-Ajlan, M.A. and Ykhlef, M., 2018. Deep learning algorithm for cyberbullying detection. International Journal of Advanced Computer Science and Applications, 9(9), pp.199-205.

[2] Mahat, M., 2021, March. Detecting Cyberbullying Across Multiple Social Media Platforms Using Deep Learning. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 299-301). IEEE.

[3] Banerjee, V., Telavane, J., Gaikwad, P. and Vartak, P., 2019, March. Detection of cyberbullying using deep neural network. In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) (pp. 604-607). IEEE.

[4] Agrawal, S. and Awekar, A., 2018, March. Deep learning for detecting cyberbullying across multiple social media platforms. In European conference on information retrieval (pp. 141-153). Springer, Cham.

[5] Kohli, R., 1993. Political Ideas of MS Golwalkar: Hindutva, Nationalism, Secularism. Deep and Deep Publications.

[6] Jaiswal, H., 2021. Memes, Confession Pages and Revenge Porn-The Novel Forms of Cyberbullying. Indore Institute of Law-Udyam Vigyati.

[7] Dadvar, M. and Eckert, K., 2020, September. Cyberbullying detection in social networks using deep learning based models. In International Conference on Big Data Analytics and Knowledge Discovery (pp. 245-255). Springer, Cham.

[8] Alotaibi, M., Alotaibi, B. and Razaque, A., 2021. A multichannel deep learning framework for cyberbullying detection on social media. Electronics, 10(21), p.2664.

[9] Hinduja, S. and Patchin, J.W., 2010. Bullying, cyberbullying, and suicide. Archives of suicide research, 14(3), pp.206-221.

[10] Tavani, H.T., 2013. Ethics and technology: Controversies, questions, and strategies for ethical computing. Hoboken, NJ: Wiley