# Toxic Comments Identification Using Deep Learning

[1]Md. Gouse,[2]K. Srikanth
[1]Bachelor of Engineering Student,[2] Bachelor of Engineering Student
[1]Information Technology,
[1]B.V. Raju Institute of Technology, Narsapur, India.

## ABSTRACT

Online forums and social media platforms have provided individuals with the means to put forward their thoughts and freely express their opinion on various issues and incidents. In some cases, these online comments contain explicit language which may hurt the readers. Comments containing explicit language can be classified into myriad categories such as Toxic, Obscene, Threat, Insult, Nationalist, Racist, Sexist, Hate speech. The threat of abuse and harassment means that many people stop expressing themselves and give up on seeking different opinions. The goal is to create a model that can predict if input text is inappropriate(toxic). Here we use Natural Language Toolkit(NLTK), LSTM, LSTM-CNN for classification of types of toxic comments.

**Keyword:** Deep Learning, Convolutional Neural Networks, Long Short-Term Memory Networks, Supervised Learning, Data Set, Data Preprocessing, Text Normalization, Tokenization, Toxicity, Lemmatization, Stop words.

## 1. INTRODUCTION

### What is Deep Learning?

Deep learning can be considered as a subset of machine learning. It is a field that is based on learning and improving on its own by examining computer algorithms. While machine learning uses simpler concepts, deep learning works with artificial neural networks, which are designed to imitate how humans think and learn.

### Types of Deep Learning

**Convolutional Neural Networks (CNNs):**

CNN's also known as ConvNets, consist of multiple layers and are mainly used for image processing and object detection. Yann LeCun developed the first CNN in 1988 when it was called LeNet. It was used for recognizing characters like ZIP codes and digits.

**Long Short-Term Memory Networks (LSTMs):**

LSTMs are type of recurrent Neural Network (RNN) that can learn and memorize long-term dependencies. Recalling past information for long periods is the default behavior.

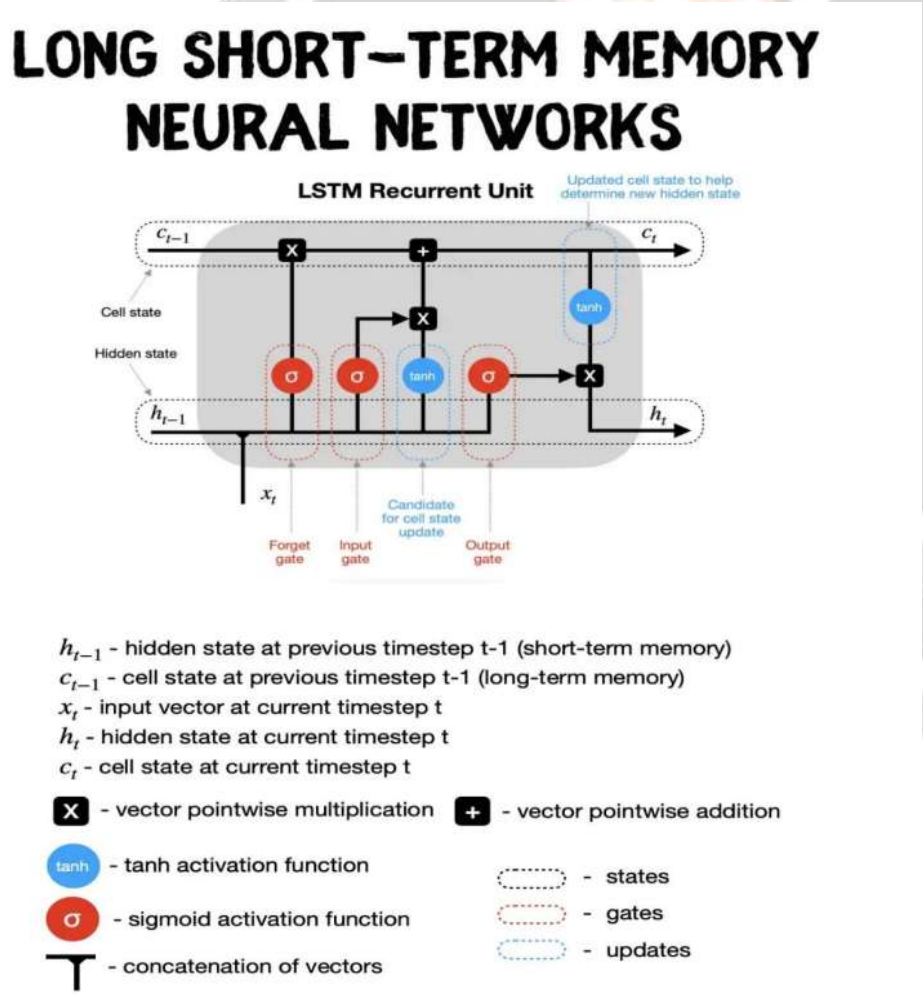
**Supervised Learning:**

LSTMs retain information over time. They are useful in time-series prediction because they remember previous inputs. LSTMs have a chain-like structures where four interacting layers communicate in a unique way. Besides time-series predictions, LSTMs are typically used for speech recognition, music composition, and pharmaceutical development. Supervised Deep Learning is similar to concepts learning in humans and animals, the difference being that the student in the former case is a computational network. Supervised deep learning frameworks are trained using well-labelled data. It teaches the learning algorithm to generalize from the training data and to implement in unseen situations. After completing the training process, the model is tested on a subset of the testing set to predict the output. Thus, datasets containing inputs and correct outputs become critical as they help the model learn faster.

.

**Supervised Learning:**

LSTMs retain information over time. They are useful in time-series prediction because they remember previous inputs. LSTMs have a chain-like structures where four interacting layers communicate in a unique way. Besides time-series predictions, LSTMs are typically used for speech recognition, music composition, and pharmaceutical development. Supervised Deep Learning is similar to concepts learning in humans and animals, the difference being that the student in the former case is a computational network. Supervised deep learning frameworks are trained using well-labelled data. It teaches the learning algorithm to generalize from the training data and to implement in unseen situations. After completing the training process, the model is tested on a subset of the testing set to predict the output. Thus, datasets containing inputs and correct outputs become critical as they help the model learn faster.



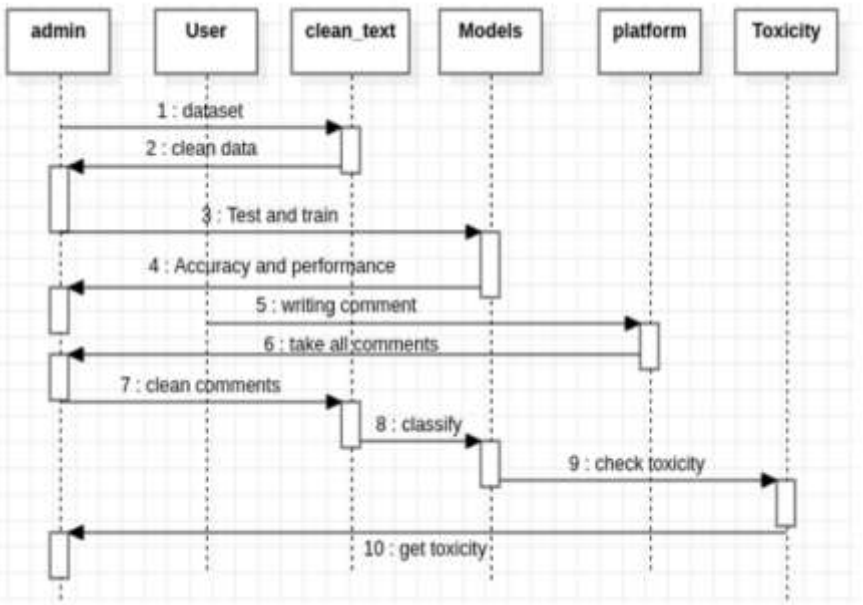

## 2. PROPOSED METHOLODOGY

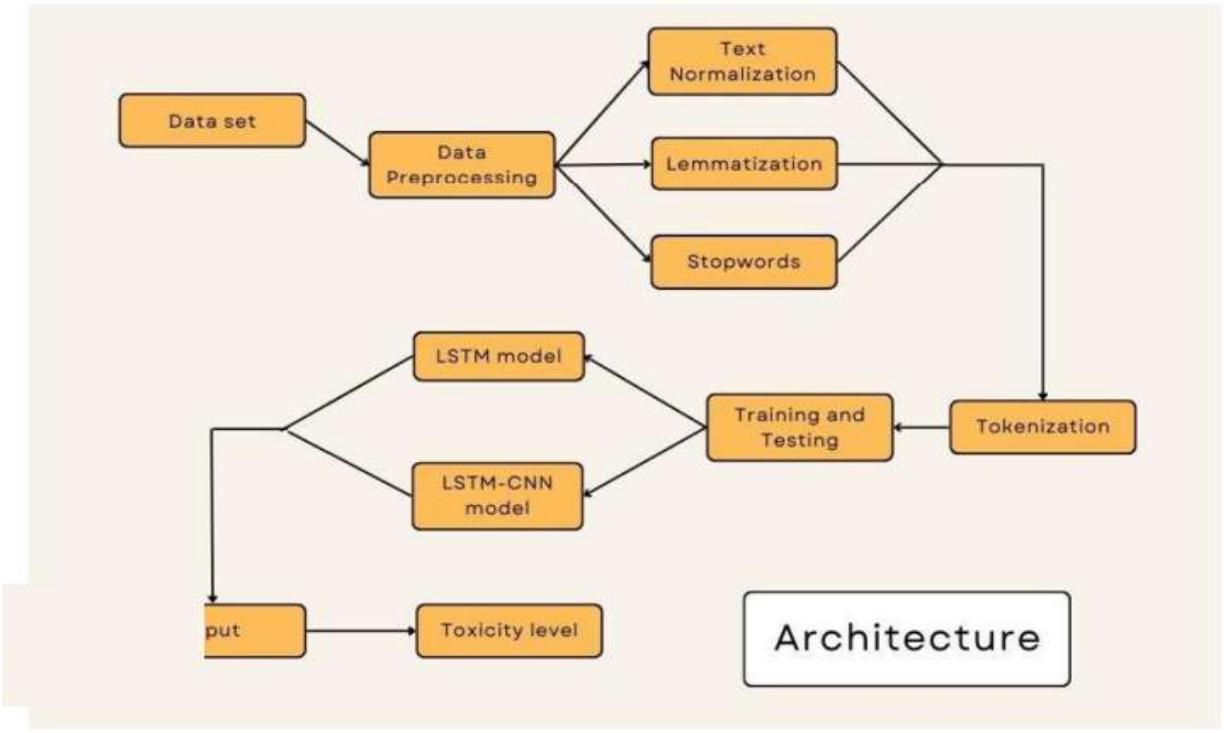
People with ASD also struggle with their limited interests and repetitive behaviors. In the list below, behavior categories are shown with examples from specific situations.

- LSTM can effectively preserve characteristics of historical information in long text sequences whereas CNN can extract the local features of the text.
- Combining the two traditional neural network architectures will help us harness their combined

capabilities.

- The goal was to compare the performance of both the deep-learning architectures and ascertain the best deep-learning model.




The proposed architecture's modules include:

1) Upload the ASD Dataset: This module is used to upload the dataset for the application

2) Pre-process the data:

The data which we take as input will get pre-processed and the stop-words are removed and the input will get cleaned like removing the commas and extra spaces.

3) Run LSTM and LSTM-CNN algorithm: now that the training data has been analyzed, the LSTM and LSTM-CNN algorithms will use to calculate the toxicity level of the comments.

4) Comparing the accuracy: now that the train data has been analyzed, the trained prediction of two algorithms are compared and most accurate will be selected as final output.

## 3. RESULTS

From taking input as text of the comment to pre-processing the text and removing the commas, extra space redundance words and sending the comment to algorithm for analyzing. Finally, the model with the most accuracy will be shown as output after comparing.

```
toxicity_level('kill each other')

1/1 [==============================] - 0s 21ms/step
Toxicity levels for 'kill each other':
Toxic:         69%
Severe Toxic:  14%
Obscene:       38%
Threat:        38%
Insult:        39%
Identity Hate: 27%
```

```
toxicity_level('wonderful movie')

1/1 [==============================] - 0s 22ms/step
Toxicity levels for 'wonderful movie':
Toxic:         0%
Severe Toxic:  0%
Obscene:       0%
Threat:        0%
Insult:        0%
Identity Hate: 0%
```

**CONCLUSION;**

The proposed project has been implemented in Anaconda for application development is developed using Deep Learning. The tasks involved in this work are divided into modules. The proposed system is efficient and has friendly user interface. The user would be able to use the app for knowing the accuracy of the toxic comment, abuse comments, threat comments percentages. Thus, successfully Comment analysis is done by considering the words that are present in the data set. LSTM and LSTM-CNN algorithm is used resulting in an accuracy of 97%.

| Deep Learning Model Accuracy Scores | | |
|---|---|---|
| Model | Score Type | |
| | Private | Public |
| LSTM | 97.70% | 97.40% |
| LSTM-CNN | 97.10% | 97.07% |

**FUTURE WORK:**

The application was implemented for knowing the toxicity level of the comments. Further we can add beautiful User Interface through which can user interact and give comment as input and also, we can show the final output of toxicity, abuse, threat in the form of graph as percentages. Also, we can take the input from the text speech through voice.

**ACKNOWLEDGMENT**

I would like to thank my guide Dr. B. Venkateshwara Rao from Information Technology department, B.V.R.I.T College, Narsapur , for his continuous guidance . Also, I would like to thank my family and friends for their continuous support.

## 4. REFERENCES

[1] J. Zhang, Y. Li, J. Tian, and T. Li, "LSTM-CNN Hybrid Model for Text Classification," 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, 2018, pp. 1675– 1680, doi: 10.1109/IAEAC.2018.8577620.

[2] P. A. Ozoh, A. A. Adigun, M. O. Olayiwola ,International Journal of Research and Innovation in Applied Science (IJRIAS) | Volume IV, Issue XI, November 2019|ISSN 2454-6194.

[3] https://www.kaggle.com/fizzbuzz/toxic-data-preprocessing.