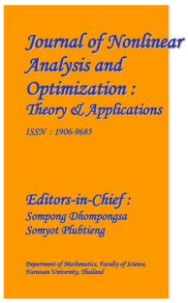
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.15 : 2024
ISSN : **1906-9685**



Paper ID: ICRTEM24_154 **ICRTEM-2024 Conference Paper**

TOXIC COMMENT CLASSIFICATION SYSTEM USING LSTM

[#1] K.Kavya Sree, [#2] J.Harshini Naik, [#3] MD Khaja Tazeem, [#4] B.K Chinna Maddileti
[1,2,3] UG Student, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana
[4] Assistant Professor, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana

***Abstract*—Over the past decade, social networking ,social media platforms have experienced exponential growth.Today, individuals have the ability to share their thoughts and opinions globally through these channels. In this context, it's expected that debates many emerge as a result of different viewpoints,these discussions can take a negative turn, escalating into conflicts on social media platforms. The identification of toxic comments presents a significant challenge for scholars in this field.Through the application of Natural Language Processing(NLP).text classification can automatically assess text and assign a set of predefined tags or categories based on its content.This particular model utilizes Long-Short-Term Memory(LSTM)Architecture to address a fore mentioned issue**

Keywords— Natural Language Processing, Long-Short-Term-Memory, Bi-directional LSTM, Python, Deep learning, Hate Speech Detection, CRNN, User-generated content.

I. INTRODUCTION

Internet negativity has always been a hottopic.The obscurity and the sense of distance of people's internet presence have encouraged people to express themselves freely.
Extreme negativity has occasionally stopped people from expressing themselves or made them give up looking for different opinions online. Issues like this be nearly all the time, across all platforms of discussion, and the modulators of these platforms have limited capabilities dealing with it. To attack the below-mentioned problem,we've developed a poisonous Comment Bracket System using Deep literacy.It helps people refrain from using negative or profane language whileinter-acting with othersand promote healthy discussion amongusers.A poisonous comment is defined as any form of textbook containing desentt,dangerous, or unhappy language that may potentially incite hostility or discomfort among compendiums.

II. RELATED WORK

Affilated exploration has looked into hate speech, online importunity,vituperative language, cyber bullying, and Obnoxious language. Generally speaking, poisonous comment discovery is a supervised bracket task and can be approached by either homemade point engineering or neural networks. A large variety of machine learning approaches have been explored to attack the discovery of poisonous language.The performances of Bi -LSTM is good and robust after data pre recycling compared to other models.Neural network approaches appear to be more effective, while point-grounded approaches save some kind of resolvable.

A. **Long-Short-Term Memory (LSTM) :** Long-Short-Term Memory (LSTM) Since we are working on a Natural Language Processing use-case, it is ideal that we use the Long Short Term Memory model (LSTM). LSTM networks are analogous to RNNs with one major difference that retired subcaste updates are replaced by memory cells. This makes them more at finding and exposing long-range dependences in data which is imperative for judgement structures. The imported "Talos*"* library since it will help us perform hyperactive parameter tuning as well as model evaluation. Using the overlook function setup the stylish parameters that would give me the loftiest delicacy.

B. **Deep Learning (DL):** To detect toxic comments, Long Short-Term Memory (LSTM) and Hybrid LSTM-CNN (Convolution Neural Network and LSTM) models are commonly used. These algorithms categorize comments based on their toxicity,including,threats,obscenity,insults and identity-based hated. Additionally recurrent neural networks(RNNs)are also employed for text classification

on multi label text data sets to identify various forms of internet toxicity2. It's essential to maintain civility in online forums, and these models play a crucial role in identifying and handling poisonous communication.

C. **Bi Directional LSTM (Bi-LSTM):** Bi Long short-term memory networks (Bi-LSTM) are a special kind of RNN's designed to be able of learning longterm dependences. Vanilla RNNs can be tough to train on long sequences due to evaporating and exploding slants caused by repeated matrix addition. LSTM break this problem by introducing a retired cell and replacing the simple update rule of the vanilla RNN with a gating mediem. Constantly, the dependences within successionall data, like rulings, are not just in one direction, thus, may be observed in both directions. Therefore, we used Bi LSTMs, in which, Bidirectional layers exploit the forward and backward depresidencies by combining the features attained going in both directions contemporaneously. the evaluation of the LSTM model. From the result, It is caused by the Naive Bayes model's ignorance of the relationship between words, which is a fatal problem. LSTM could flashback the connections and combine those meaning together to get the result, and it led to a more accurate result than our birth model.

III. METHODS AND EXPERIMENTAL DETAILS

A. No Third-Party Authorization:

When it comes to detecting toxic comments, avoiding third-party authorization ensures privacy and data security.This means the system doesn't rely on external services or platforms to analyze or moderate content, thus keeping user data within the system;s control.This approach can enhance trust and minimize risks associated with sharing sensitive information with external parties.

Toxic comment detection without third-party authorization typically involves building a machine learning model on your own using publicly available data sets or creating your own data set. This approach requires expertise in natural language processing and machine learning, but it allows you to maintain full control over the model and its deployment.

If you're not using third-party authorization for toxic comment detection with LSTM, you'll likely need to implement your own user authentication system and data privacy measures to ensure that only authorized users can access and interact with your system.

B. Natural Language Processing(NLP):

The Natural Language Processing(NLP) plays a crucial role in toxic comment detection.NLP techniques are used to preprocess text data, extract features and build models that can identify toxic or abusive language in comments or texts.Common NLP techniques used in toxic comment detection include tokenization,stemming or lemmatization, stop word removal, vectorization(e.g., using techniques like TF-IDF or word embedding), and building models such as LSTM(Long-Short-Term Memory) networks or other machine learning classifiers.These models learn to recognize patterns in text data indicative of toxicity and classify comments accordingly. Additionally, techniques like sentiment analysis and named entity recognition can also complement toxic comment detection systems by providing further context to the analysis.

C. Data Sets:

In order to have a better understanding of data distribution ,we first checked time series for toxin regarding to different identities.We use data sets from Kaggle, The data set comprises of over numerous rows. Each row contains a general poisonous target arranger from 0 to 1, a comment text,scores under colorful markers similar as severe slag,identity,attack,personality,trouble,homosexual gay or lesbian,black, intellectual or learning disability.

The Besides the common word compression mappings like "it;d: to "it would" , we added word compression mappings that were related to our data set. We collected some common misspell words and corrected them, similar as "tRump" to "Trump". We restated some special Latin words and Emoji's to English words.

The data set is resolve into 80% as training set, 10% as dev set and 10% as test set.

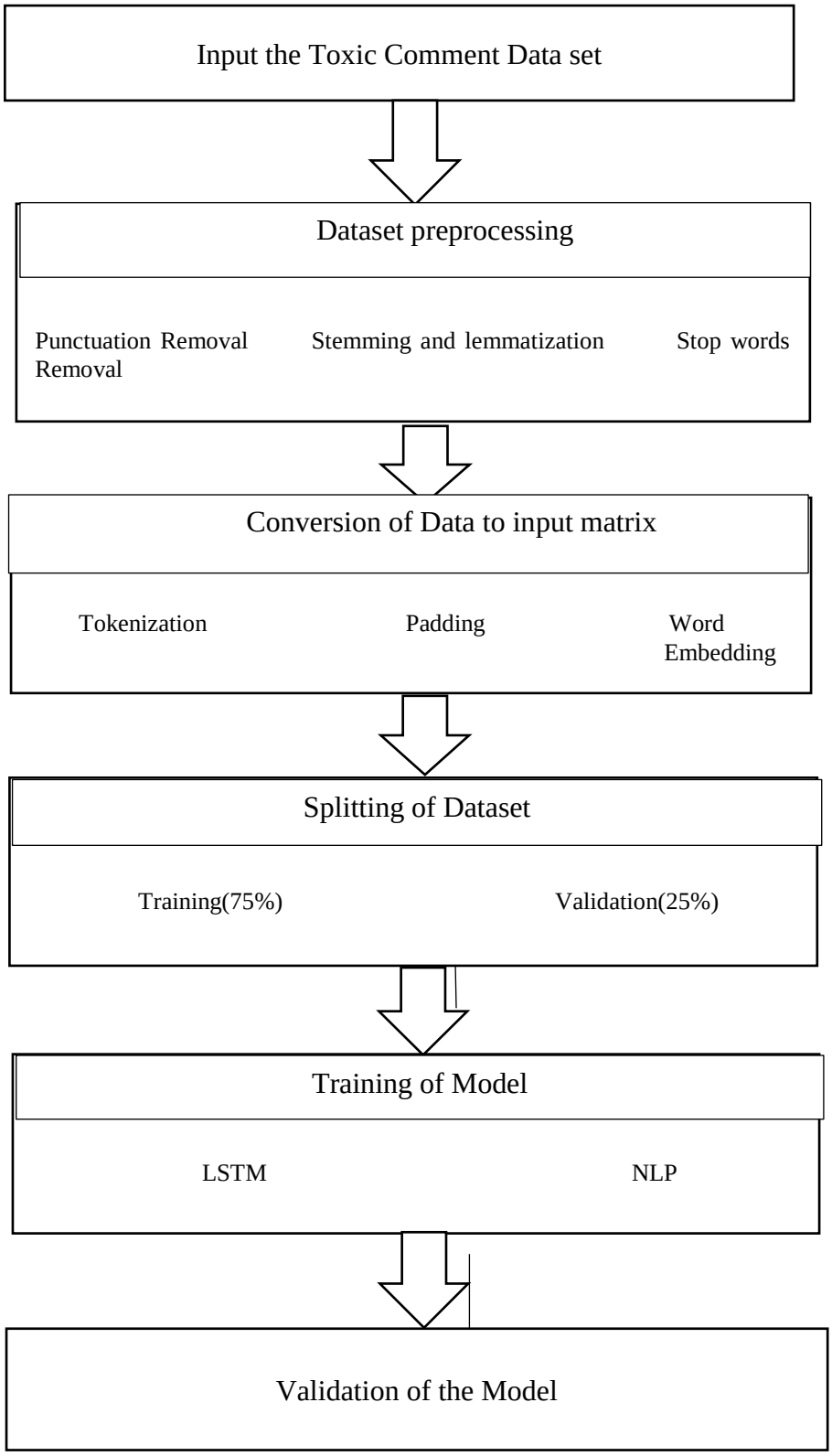

Input the Toxic Comment Data set
Dataset preprocessing
Punctuation Removal
Removal
Stemming and lemmatization
Stop words
Conversion of Data to input matrix
Tokenization
Padding
Word Embedding
Splitting of Dataset
Training(75%)
Validation(25%)
Training of Model
LSTM
NLP
Validation of the Model

Fig-1: Architecture of the model

IV. RESULTS AND DISCUSSIONS

Our Toxic Comment Classification System is developed to efficiently classify negative comments to promote healthy relationships.This technology has been extensively applied in various applications, including content moderation, online safety, and maintaining a positive online environment

No Third-Party Authorization (NTPA):

Approach: In toxic comment detection. Avoiding third party authorization ensures privacy and data security.It means the system doesn't rely on external services or platforms to analyze or moderate content, keeping user data within the systems control. This approach can enhance trust and minimize risks associated with sharing sensitive information with external parties

Applicability to Engineering: This approach there are many scenarios where this is relevant. For instance, in systems design, ensuring that authorization mechanisms are built in- house rather than relying o third-party solution can provide more control over security and access .**Privileges:** The individual or entities have the authority to access or control certain resources or actions without relying on external parties for approval or validation.

Natural Language Processing (NLP):

Approach: The NLP techniques are used to preprocess text data, extract features and build models that can identify toxic or abusive language in comments or texts. Common NLP techniques used in toxic comment detection include tokenization, stemming or lemmatization, stop words removal, vectorization and building models such as LSM networks or other machine learning classifiers.

Applicability to Engineering: Natural Language process Has numerous applications in engineering, spanning various domains. Common applications are Text classification,Information Extraction,sentiment analysis.In essence, NLP techniques can streamline various aspects of engineering work flows, from documentation and communication to decision-making and problem-solving

Privileges: That certain linguistic elements or structures possess within a given context.It allow algorithms to better interpret and generate human language in context.

Data Sets (DS):

Approach: In order to have a better understanding of the data distribution, we first checked the time series for toxicity regarding to different identities.

Applicability to Engineering: Toxic comment detection is highly relvent in engineering,particularly in fields like content moderation and social media analysis.Engineers continuously refine these models to improve their accuracy and effectiveness in real- world applications.

Privileges: Data sets manifest in various ways,such as certain comments being labeled or categorized differently based on the identity or background.

Comparison:

Each NLP models in toxic comment detection without involving third-party data sets involves evaluating their performance based solely on the data sets they were trained on.Such comparisons help identigy strengths and limitations of different approaches,aiding researchers and practitioners in selecting the most suitable models for their specific applications

Integration:

The integration of (NTPA), (NLP), and (DS) implies examining how different linguistic elements integrate or combine to convey meaning within natural language.In the context of toxic comment detection data sets without involving third-party data ,integration theory could focus on how various linguistic features within the data set,such as word choice,sentence structure, and contextual cues,combine to identify toxic comments accurately.

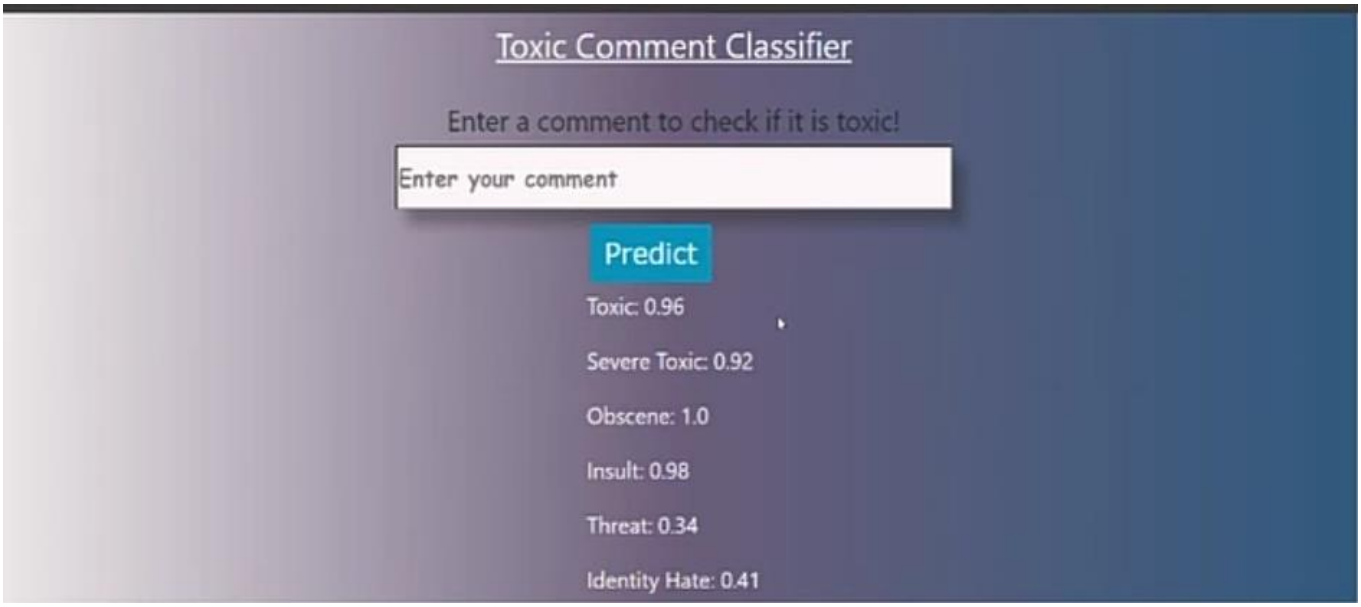


Fig-2 User Interface - 1
Based on the entered comment it predicts the Toxic,Severe Toxic,Obscene,Insult,Threat,Identity Hate of the text

Fig-3 User Interface - 2
While conversation or chatting between two users if any of the user attempts to text any toxic comment it doesn't allow to send or display.

V. CONCLUSION

In conclusion,toxic comment classification using Long Short Term Memory(LSTM) networks is a powerful and effective approach for identifying and filtering out harmful,offensive, or inappropriate content in online discussions and social media platforms.This technology has found extensive application in various domains,including content moderation,online safety,and maintaining a positive online environments.

No Third-Party Authorization (NTPA):
In conclusion,eliminating the need for third-part authorization typically entails constructing a machine learning model on your own using publicly available data sets or creating your own data set,This approach requires expertise in natural language processing and machine learning, but it allows you to have full control over the model an its deployment.

Natural Language Processing (NLP):
In conclusion,NLP technique are used to preprocess text data,extract feature,and build models that can identify toxic or abusive language in comments or texts.Common NLP techniques used in toxic comment detection include tokenization,stemming or lemmatization ,stopword removal, vectorization(e.g.,using techniques like TF-IDF or word embedding),and building models such as LSTM(Long-Short-Term Memory)networks or other machine learning classifiers.

Data Sets (DS):

In order to have a better understanding of data distribution,we first checked time series for toxicity regarding to different identities.we use data sets from Kaggle, The data set comprises of over many rows.
Each row contains a general toxic target scorer from 0 to 1, a comment text,scores under various labels such as severe toxicity,obscene,identity,attack,insult,threat,homosexual gay or lesbian,black,intellectual or learning disability

REFERENCES

[1] S. Se, R. Vinaya Kumar, M.A. Kumar and K.P Soman, "AMRITA - CEN@SAIL2015: Sentiment Analysis in Indian Languages", *MIKE*, 2015.

[2] R. Vinaya Kumar, K.P. Soman and P. Poorna Chandran, "Long short-term memory grounded operating log anomaly discovery", *2017 International Conference on Advances in Communicating Communications and Informatics (ICACCI)*, pp. 236-242, 2017.

[3]Guizhu Shen, Qingping Tan, Haoyu Zhang, Ping Zeng and Jianjun Xu, "Deep Learning with Reopened intermittent Unit Networks for Financial Sequence Prognostications", *8th International Congress of Information and Communication Technology*, 2018.

[4] Navoneel Chakrabarty, "A Machine Learning Approach to Comment Toxicity Classification", *International Conference on Computational Intelligence in Pattern Recognition (CIPR 2019)*.

[5]A. Akshith Sagar and J. Sai Kiran, "Toxic Comment Classification using Natural Language Processing", *International Research Journal of Engineering and Technology (IRJET - 2020)*.

[6] P. Vidyullatha, Satya Narayanan Padhy, Javvaji Geetha Priya, Kakarlapudi Srija and Sri Satyanjani Koppisetti, "Identification and Bracket of poisonous Commentary Using Machine Learning Methods", International Journal of Research and Innovation in Applied Science (URIAS-2021).

[7] S Viswanathan, Anand Kumar and K.P Soman, "A Sequenc-Grounded Machine Appreciation Modelling Using LSTM and GRU" in Emerging Research in Electronics Computer Science and Technology. Lecture Notes in Electrical Engineering, Singapore:Springer, vol. 545. (2022).