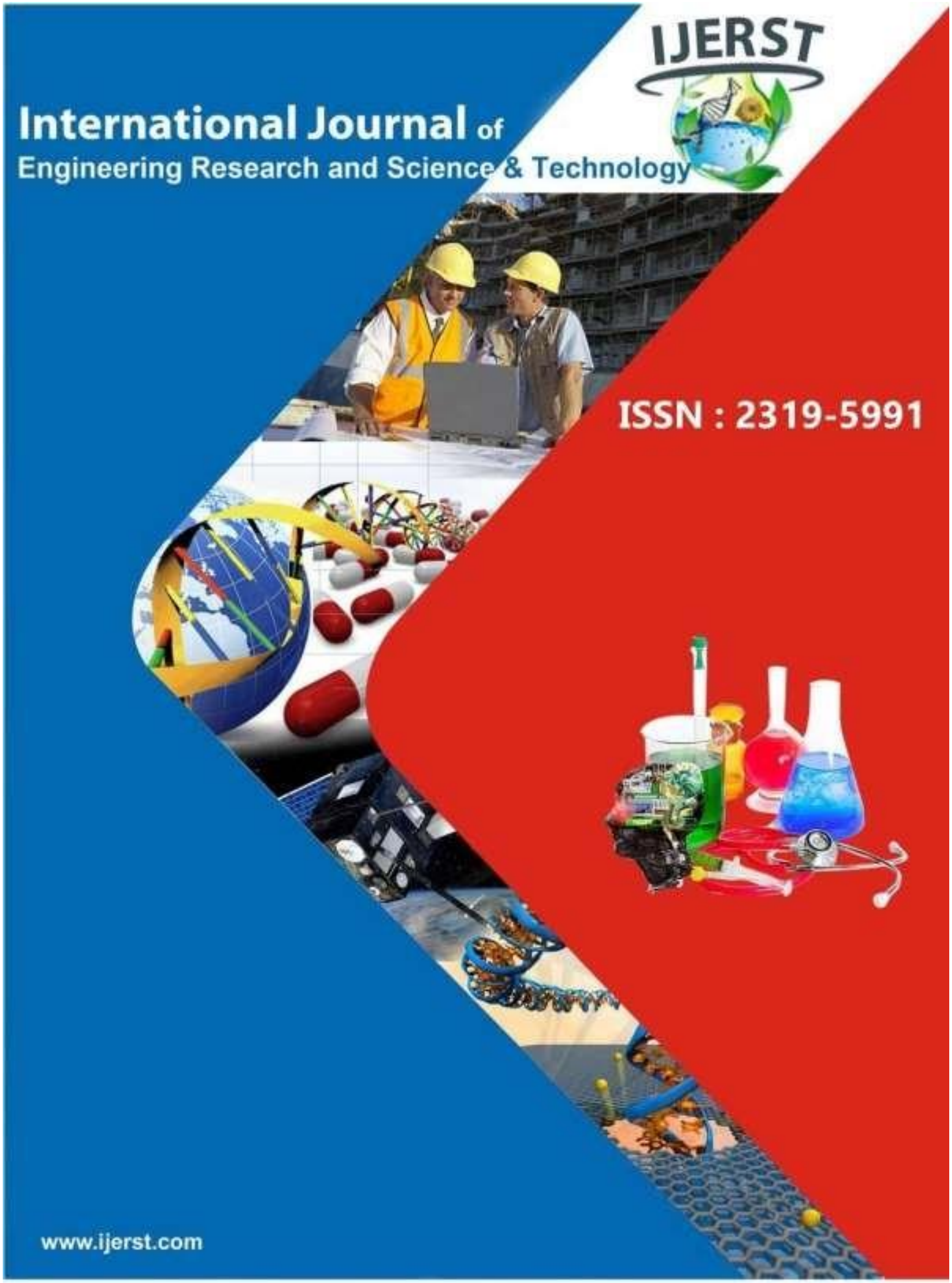

Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com
Vol. 19, Issuse.2,April 2023

Classification of Online Toxic Comments using Machine Learning Algorithms

Mrs. NVS Sowjanya , [1],M Devi Sri Surya Maneesh [2],B Sri Satya Kesava Tanuj [3],Chekuri Vineeth Krishna [4],Chalagalla Vinitha [5],GVVenkata Siva Srava [6]

Article Info

Received: 04-01-2023 *Revised: 06-02-2023* *Accepted: 18-03-2023*

Abstract:

Conversational toxicity is a problem that might drive people to cease truly expressing themselves and seeking out other people's opinions out of fear of being attacked or harassed. The purpose of this research is to employ natural language processing (NLP) techniques to detect toxicity in writing, which might be used to alert people before transmitting potentially toxic informational messages. Natural language processing (NLP) is a part of machine learning that enables computers to comprehend natural language. Understanding, analyzing, manipulating, and maybe producing human data language with the help of a machine are all possibilities. Natural Language Processing (NLP) is a type of artificial intelligence that allows machines to understand and interpret human language instead of simply reading it. Machines can understand written or spoken text and execute tasks such as speech recognition, sentiment analysis, text classification, and automatic text summarization using natural language processing (NLP).

Keywords – Convolutional neural networks, medical image analysis, machine learning, deep learning.

I. Introduction

In the early days of the Internet, people communicated only by e-mail that encountered spam. At the time, classifying emails as good or bad, whether or not spam was a challenge. Internet communications and data flows have changed significantly over time, especially since the development of social networks. With the advent of social media, classifying content as “good” or “bad” has become more important than ever to prevent social harm and prevent people from engaging in antisocial behavior.

ASSISTANT PROFESSOR
DEPT OF INFORMATION TECHNOLOGY

PRAGATI ENGINEERING COLLEGE(A),SURAMPALEM(EAST GODAVARI)A.P,INDIA

A. Purpose

Every day, vast amounts of data are released by social media networks. This massive volume of data is having a big impact on the quality of human existence However, because there is so much toxicity on the Internet, it can be harmful. Because toxic comments limit people's ability to express themselves and have different opinions, as a result of the negative, there are no positive conversations on social networks. As a result, detecting and restricting antisocial behavior in online discussion forums is a pressing requirement. These toxic comments might be offensive, menacing, or disgusting. Our goal is to identify these harmful remarks.

B. Project Overview

To classify noxious comments, this study will use different methods of machine learning. To handle the problem of text categorization, Some of the methods we use are logistic entertainment, random forest, SVM classifier, multi-navigation database, and XGBoost classifier. As a result, we maintain a data setusing six machine learning algorithms, evaluate their accuracy, and compare them. We choose the model with the highest accuracy and forecast the toxicity based on unseen data based on its accuracy level.

.

Figure 1. Examples of MRI images of the T1-CE MRI image dataset. Left: coronal view of a meningioma tumor. Center: Axial view of a glioma tumor. Right: sagittal view of a pituitary tumor. Tumor borders have been highlighted in red.

II. LITERATURE SURVEY

A. Existing Systems

1) Convolutional Neural Networks (CNN): CNN structures for categorization are Neural Network structures that can adapt to many stages. Each of these levels has a number of layers. The Embedding Layer is a CNN component that is used to solve problems with text categorization. The translation of the embedded text into CNN's description format is a function of the input layer. This approach converts each word in a text document into a standard compressedvector.

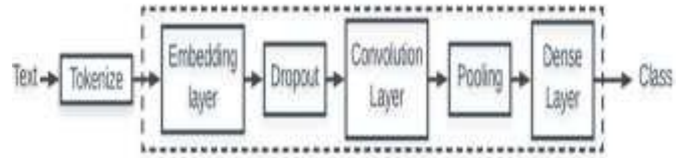

Fig: A Convolutional Neural Network

2) LSTM (Long Short-Term Memory): Long-term memory is abbreviated as LSTM. According to memory, LSDM is a kind of continuous neural network that works better than conventional continuous neural networks. LSTM is certainly better when it comes to memorizing specific patterns. Like other NNs, LSTM can have multiple hidden layers, storing the relevant data in each cell as it passes through each level, and rejecting inappropriate data. LSTM has a memory function that allows you to store data sequences. It also has another function: it works to delete unwanted data, and since we all know that text data has a lot of unnecessary data, LSTM can delete it to save time and money on calculations. As a result, LSTM's ability to memorize data sequences while deleting additional data makes it an effective text classification tool.

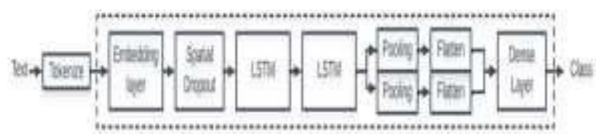

Fig: A Long Short-Term Memory Network.

III. Methodology

We will use five machine learning algorithms to classify the data the online harmful comments: logistic regression, random forest, SVM classifier, Naive Bayes, and XGBoost classifier. We'll use logistic regression to determine whether or not a comment is harmful because it either belongs to the poisonous group or does not. SVM classifiers can be used to separate data values, and can also be used in XGBoost and Random Forest systems. Because we can classify ideas into a wide range of destructive and non-toxic, we will use the concept of results in both directions. Because our data are independent of each other and have nothing in common with two different concepts, we use the Nave Bayes classifier to classify them. Because the data is marked, we can immediately use a controlled machine learning algorithm.

1) Logistic Regression: One of the few approaches to the classification of utility data is logistic regression.

Suppose you have a medical history of a patient with a tumor. It is necessary to determine if the tumor is malignant (dangerous). The concept also suggests whether it is toxic or not. 1 means positive class and 0 means negative class. Positive (1) mean that it is toxic and negative (0) means that it is non-toxic. This is calculated by the sigmoid function. Suppose we have only one piece of information. It has different features. Suppose we have the right weight mattress. Depending on the data point, we now need to provide a class tag (classify it as 1 or 0). The weight vector is applied, and the input vector is multiplied by it, which gives a scalable output. To get a value between 0 and 1, this output is placed into a Sigmoid function. We'll name thislikelihood of the projected class 1 for now. If the probability is greater than 0.5, the predicted class is 1. If the probability is less than 0.5, the predicted class is 0.

$$[x_1 \quad x_2 \quad x_3 \quad \dots \quad x_n]$$

$$sigmoid(x) = \frac{1}{1+e^{-x}}$$

Fig: Sigmoid Function

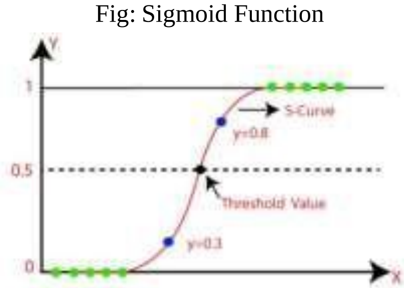

Fig : Logistic Regression curve

2) Random Forest: Random Forest is a controlled machine learning approach to classification and regression problems. Uses an explicit majority for classification and a moderate delay to create a decision tree of different data. One of the main features of the random forest algorithm is the ability to manage a set of data with classified and continuous variables, such as regression and classification. As for the rating, it performs better than its competitors.

STEPS

Step 1: Random Forest takes random entries from the Q-record database.

Step 2: A separate end tree is created for each model.

Step 3: Each end tree will have an output.

Step 4: The final decision on classification and regression is based on an appropriate or intermediate majority.

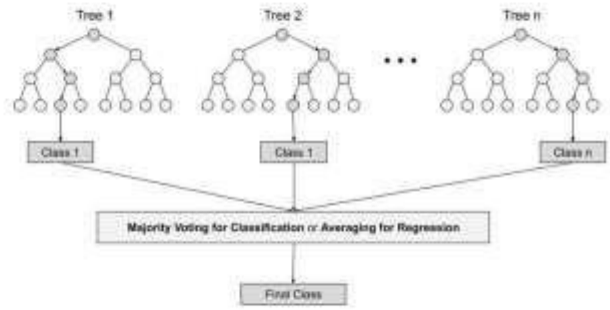

Fig : Random Forest

3) SVM Classifier: The reference vector machine (SVM) is a machine learning technique with classification and regression monitoring. SVMdefines a cloud page that classifies the boundariesbetween two data sets. SVM is often used to classify data, although it can also be used for regression. This is a fast and reliable method that works well with low data. Another advantage of SVM is that it can explorevarious input functions without increasing system problems, using different types of core functions.

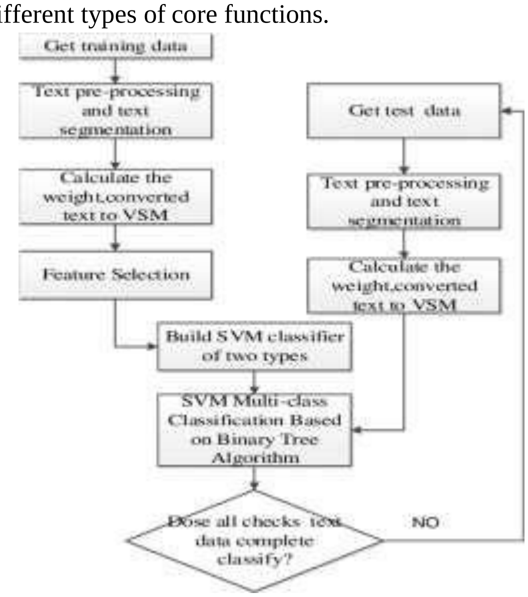

Fig : Flow chart of SVM

4) Naive Bayes: The Bayes theorem is a mathematical procedure for estimating conditional probabilities. A probability, as you may know, is the possibility of an event occurring. We call an event's probability if it has a chance of occurring.

5) XGBoost Classifier: Gradual gain, such as Random Forest (another end-of-tree algorithm), is a controlled machine learning approach for topics such as classification (men, women) and regression (expected value). The most common names to implement this method are Gradient Boosting

Machines (GBM) and XGBoost. Gradient Boosting is a training group similar to Random Forest. This means that it links the final model to a set of individual models. These individual models have a poor prognosis and are very consistent, but combining several of these weak models into one group will give the best results. Random forest-related end trees are the most popular weak sample used in gradient reinforcement machines.

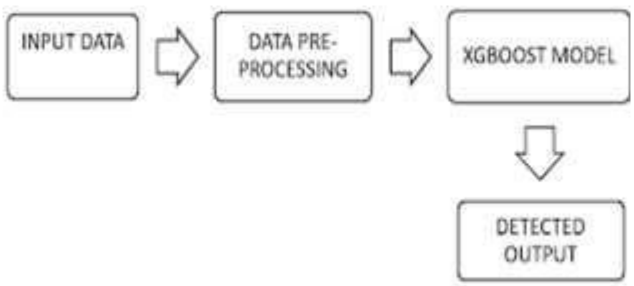

Fig: XGBoost process

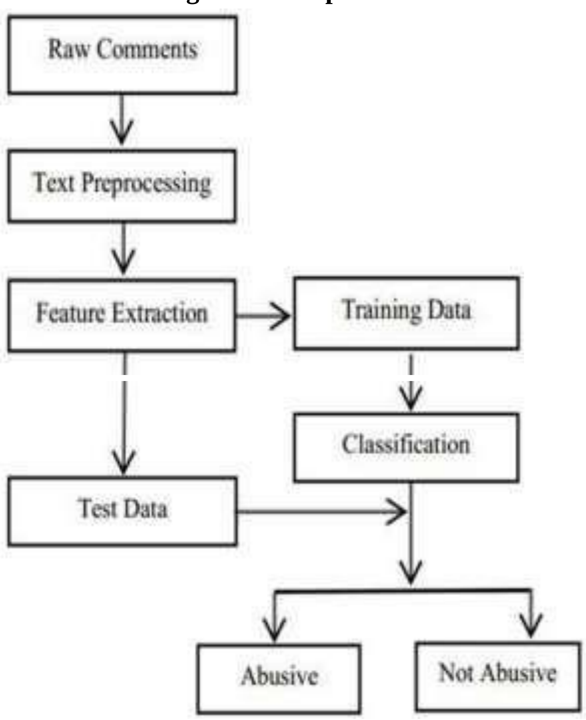

Fig: Flow chart for detecting toxic comments

IV Evaluation

4.0. Comparison with Other Methods

We upload our dataset first, then perform preprocessing such as stemming and removing stop words, before applying our machine learning models and calculating accuracy for each model. Finally, the most accurate model is chosen. The best model is then used to estimate toxicity based on unknown data.

	Algorithms	Accuracy
0	Naive-Bayes	0.791330
1	SVM	0.789219
2	Logistic Regression	0.800149
3	Random forest	0.583778
4	Xgboost	0.755186

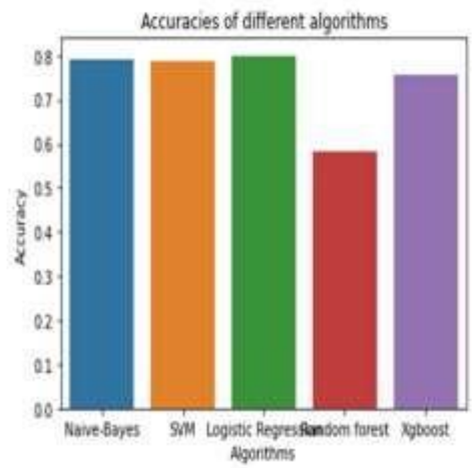

Fig: Accuracies of Algorithms

Now we predict on unseen data by taking our best algorithms Logistic Regression.

```
text='im mad at you bitch' # enter your text for testing
# get the prediction for the text
text_=count_vect.transform([text])
pred=lr.predict(text_) # predicting
prob=np.amax(lr.predict_proba(text_))#  getting probability
print(pred,prob)
```

```
['Offensive'] 0.9951340848415849
```

```
text='i am attracted to ur beauty' # enter your text for testing
# get the prediction for the text
text_=count_vect.transform([text])
pred=lr.predict(text_) # predicting
prob=np.amax(lr.predict_proba(text_))#  getting probability
print(pred,prob)
```

```
['Non-offensive'] 0.6150410062957921
```

V.CONCLUSION

We tested the accuracy of five machine learning methods: logistic rest, ship bay, random forest, SVM classifier, and XGboost classifier. After careful consideration, we can conclude that the logistic delay has the best performance in terms of accuracy, as it is the most accurate model among other models. As a result, we choose our final model based on its precise nature. If there is a logistic regression model, we get the maximum. We will use the logistic delay model as our newest machine learning strategy, as it works best with our data.

REFERENCES :

[1] Rahul, H. Kajla, J. Hooda and G. Saini, "Classification of Online Toxic Comments Using Machine Learning Algorithms," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 1119-1123, doi: 10.1109/ICICCS48265.2020.9120939

[2] M. Duggan, “Online harassment 2017,” Pew Res., pp. 1–85, 2017, doi: 202.419.4372.

[3] M. A. Walker, P. Anand, J. E. F. Tree, R. Abbott, and J. King, “A corpus for research on deliberation and debate,” Proc. 8th Int . Conf. Lang. Resour. Eval. Lr.2012, pp. 812–817, 2012.

[4] J. Cheng, C. Danescu-Niculescu-Mizil, and J. Leskovec, “Antisocial behavior in online discussion communities,” Proc. 9th Int. Conf. Web Soc. Media,ICWSM 2015, pp. 61–70, 2015.

[5] B. Mathew et al., “Thou shalt not hate: Countering online hate speech,” Proc. 13th Int. Conf. Web Soc. Media, ICWSM 2019, no. August, pp. 369–380, 2019.

[6] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang, “Abusive language detection in online user content,” 25th Int. World Wide Web Conf. WWW 2016, pp. 145–153, 2016, doi: 10.1145/2872427.2883062.

[7] E. K. Ikonomakis, S. Kotsiantis, and V. Tampakas, “Text Classification Using Machine Learning Techniques,” no. August, 2005.

[8] M. R. Murty, J. V. . Murthy, and P. Reddy P.V.G.D, “ Text Document Classification basedon Least Square Support Vector Machines with Singular ValueDecomposition,” Int. J. Comput. Appl., vol. 27, no. 7, pp. 21–26, 2011, doi: 10.5120/3312-4540.

[9] E. Wulczyn, N. Thain, and L. Dixon, “Ex machina: Personal attacks seen at scale,” 26th Int. World Wide Web Conf. WWW 2017, pp. 1391– 1399, 2017, doi: 10.1145/3038912.3052591.

[10] H. Hosseini, S. Kannan, B. Zhang, and R. Poovendran, “Deceiving Google’s Perspective API Built for Detecting Toxic Comments,” 2017, [Online]. Available:http://arxiv.org/abs/1702.08138.