Message Filtering System in Online Social Networks

Nishita Chaudhary[1], Omkar Mainkar[2], Tejas Chavan[3], Madhura Vyawahare[4]

[1-3]Student, Department of Information Technology, Pillai College of Engineering, New Panvel, Maharashtra, India
[4]Professor, Department of Information Technology, Pillai College of Engineering, New Panvel, Maharashtra, India

---***---

Abstract -*In human society, Online Social Networking Sites are important as they can be used to communicate and share different types of information. Moreover, OSN are also reliable for business purposes, learning new things etc. Hence, the usage of OSN's had increased to a large extent in recent years. However, a large number of people are also subjected to cyberbullying. A 2019 study proved the associations between some types of unethical internet use and psychiatric or behavioural problems such as depression, anxiety, aggression etc. Hence, our system will try to detect and filter the abusive words, which may help to reduce the amount of cyberbullying on a significant scale.*

Key Words*:* **OSN, cyberbullying, abusive, Machine Learning.**

1. INTRODUCTION

The internet is full of every type and kind of information. To use or gain knowledge from this data there are equal numbers of human interpreters i.e. users exist on the internet. These users can share as well as communicate multiple varieties of information among them. The information consists of different types of data, such as text, audio, video, images etc. This type of data exchange takes place in the form of mails, messages and social networks posts on the internet. Social networking sites have proved to be a great platform for people to connect people from all around the world. It helps in increasing communication and sharing information with people all over the world. These sites are available anywhere on the internet. However, there are chances on Online Social Networks (OSNs) of posting unwanted content on particular wall areas, called in general walls. This unwanted content consists of vulgar, offensive, abusive words which causes cyberbullying. As people tend to spend more time on OSN, they are more subjected to cyber-bullying. Multiple studies suggest that cyberbullying can have a negative impact on mental and psychological health of users which can cause or can relate to depression, low self-esteem, and suicidal thoughts. Owing to these, we propose a model in which all the abusive, vulgar, offensive etc. words posted will be filtered.

2. LITERATURE SURVEY

The authors D.J. Mishra et.al. have discussed a new Artificial Intelligent approach in order to control bad keywords in user's walls. Filtering method is applied by using the technique of Expert System to filter unwanted keywords on the user's Wall. Their Proposed System has a total 5 components which are Preprocessing, Post-extraction techniques, Filtration of words, Filtration of words, Categorize the user with percentage level, Matching the percentage with grouped users [1]. Authors Snehal D'Mello et.al. have explained that their proposed system gives the ability to the user to control the content of messages being posted. These messages are classified into different categories and based on the output, filtering rules are applied to decide if the message is to be posted or not. They had created two classifiers for classification of messages - Radial Basis Probabilistic Neural Network and Radial Basis Neural Network. Comparison is done between these two which shows the Radial Basis Probabilistic Neural Network has proven to be more effective. Their proposed system consists of three modules:- Text classification module, content based filtering module and recommendation module [2]. Nitin Pondhe et.al. has proposed a flexible rule-based system which will give the users the ability to customize the filtering criteria. A soft classifier is used which will automatically label messages in support of content based filtering. Their proposed system architecture is a three-tier in which the first layer consists of social network manager, second layer provides the support for external social network application and third layer is graphical user interface. The main components of the proposed system are Content

Based Filtering and Short Text classifier modules [3]. Authors K Subba Reddy et.al. had proposed a model for spam detection on social networking sites. The work flow of the model is Collecting sample dataset, Preprocessing data, Feature selection from data, constructing models, evaluating it, model testing with cross validation, comparison of models. They have used classification algorithms of Support Vector Machine and Naive -Bayes in model. The measure of performance is tested using precision, recall and F-measure metrics[4]. On the same line we have developed a system to give better performance for message filtering.

3. PROPOSED SYSTEM

In the Proposed system users post the message according to their view which can be offensive or not offensive. Then preprocessing of the data is done at the beginning where the raw data is converted in efficient format. Tokenization is the act of breaking up a sentence into pieces such as words, keywords, phrases, symbols and other elements called tokens.

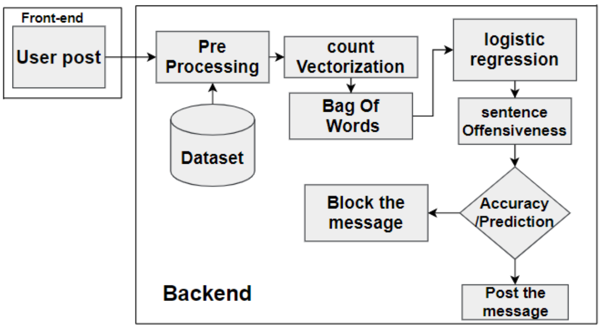


Fig.1 Proposed System.

Tokens can be individual words, root words, smaller words, phrases, or even whole sentences. In the process of tokenization, some characters like punctuation marks, white spaces and special characters are discarded. In Lemmatization, the prefix and suffix of the word are removed. We have used the WordNetLemmatizer() function on a single word.

The data users post on the wall is in the form of sentences so to convert sentences into the numerical format we have used a countvectorizer(). Count vectorizer works on frequency, it counts the occurrence of tokens. In vectorizer, we have used stop words. The stop words are the words that do not have any meaning i.e. 'is', 'are', 'at' etc. these kinds of words are removed from the sentence. After removing the stop words we have applied a vectorizer to the remaining word. Vectorizer uses a bag of words and assigns every word a special integer. Bag of words transforms documents into vectors where each word in the document is assigned a number. In a bag of words, the order of the word is lost in the bag, only the occurrence or presence of the word matters. We have applied the Naive Bayes model for training and testing of the data. We have predicted the test data.

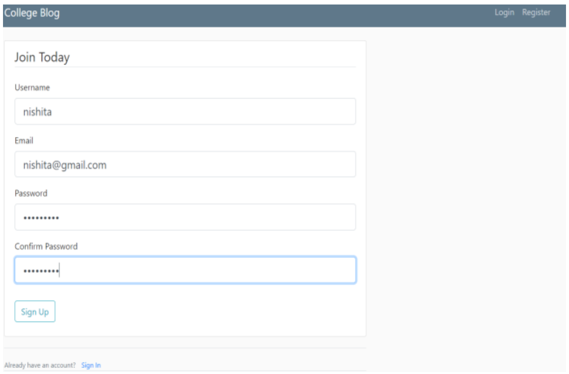
We have also applied SVC, Random forest and Logistic Regression techniques for testing and training. These models are for comparative purposes. After comparing these models we found Logistic Regression is more effective and accuracy is high in this algorithm. Accuracy of the word is being predicted, if the word is found to be offensive it will block the message that the user is posting on their wall and if the word is not offensive it will allow the user to post the message on their wall. The proposed system is divided into two parts, the frontend and backend. For frontend, we have used HTML, CSS and flask. For Backend, we have used python for filtration of messages and for storing the user's information we have used SqlAlchemy.

4. IMPLEMENTATION PROCESS AND RESULT ANALYSIS

For Implementation, we have created a social media platform called "College Blog" where users can register and then login themselves in order to be a part of the Online Social Network.

1. Registration.

Users first register themselves, the registration process involves giving their username, email address and password. The details that the user gives to the college blog is then acquired and stored in the Gui based Database that stores terabytes of data.



Join Today

Username

nishita

Email

nishita@gmail.com

Password

•••••••••

Confirm Password

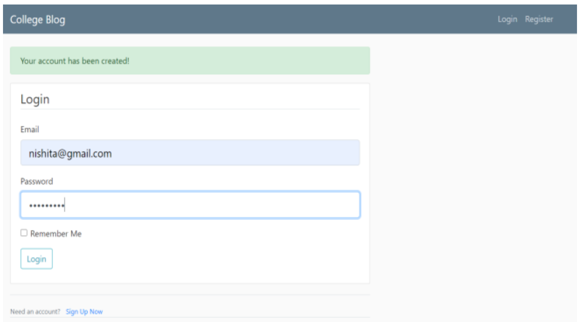
•••••••••

Already have an account? Sign In

Fig.2 Registration.

2. Login.

After registration Login is done. The authenticated user will be permitted to access his home page once user credentials are verified with the data store.



Your account has been created!

Login

Email

nishita@gmail.com

Password

•••••••••

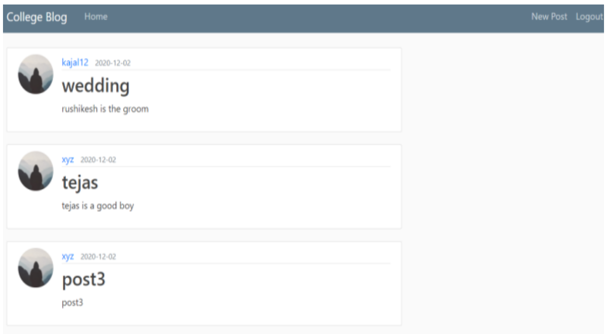
□ Remember Me

Need an account? Sign Up Now

Fig.3 Login Page.

3. Home Page.

The webpage contains the dashboard, that contains the home page, new post and logout. In the home page all the messages which are not offensive are posted. The new post contains the heading/title of the content and the main content where the users can share their ideas, views, information, or the reaction. Multiple users can post multiple new posts on the wall.



kajal12 2020-12-02

wedding

rushikesh is the groom

xyz 2020-12-02

tejas

tejas is a good boy

xyz 2020-12-02

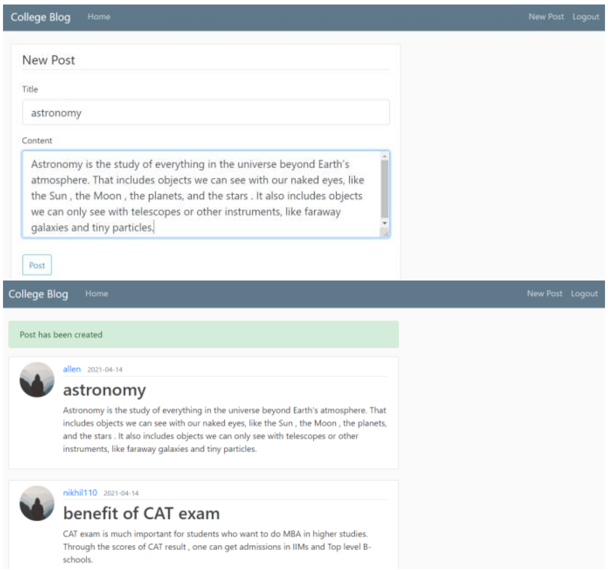
post3

post3

Fig.4 Home Page.

4. Not Offensive message.

If the user posts anything not related to offensive, then the system will allow the message to post on the wall.



New Post

Title

astronomy

Content

Astronomy is the study of everything in the universe beyond Earth's atmosphere. That includes objects we can see with our naked eyes, like the Sun , the Moon , the planets, and the stars . It also includes objects we can only see with telescopes or other instruments, like faraway galaxies and tiny particles.

Post has been created

allen 2021-04-14

astronomy

Astronomy is the study of everything in the universe beyond Earth's atmosphere. That includes objects we can see with our naked eyes, like the Sun , the Moon , the planets, and the stars . It also includes objects we can only see with telescopes or other instruments, like faraway galaxies and tiny particles.

nikhil110 2021-04-14

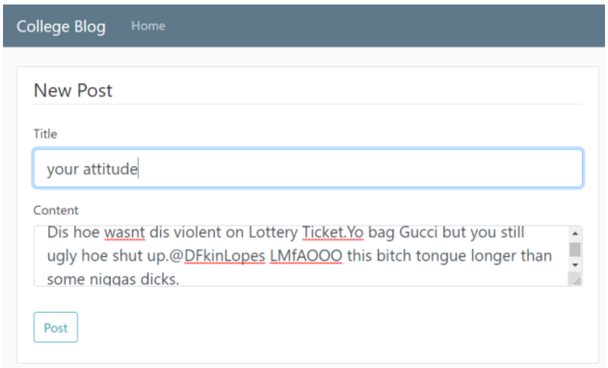
benefit of CAT exam

CAT exam is much important for students who want to do MBA in higher studies. Through the scores of CAT result , one can get admissions in IIMs and Top level B-schools.

Fig.5 User uploading new post.

5. Offensive message.

If the user shares something offensive in the sentence then the system will not allow the message to be posted on the wall, the system will block the message. Here bitch, ugly, hoe, violent etc. are offensive words.



New Post

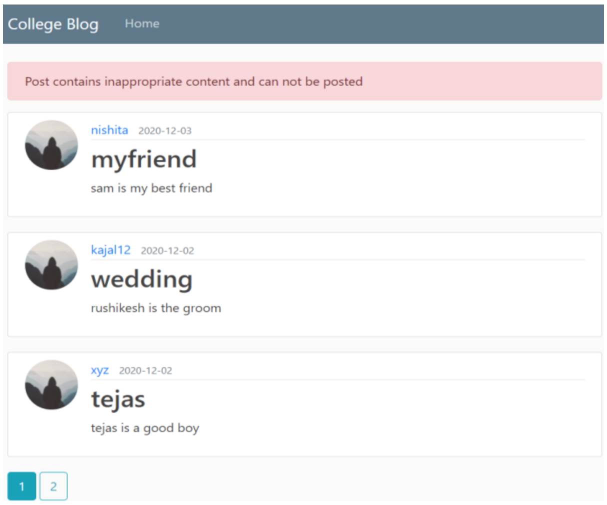
Title

your attitude

Content

Dis hoe wasnt dis violent on Lottery Ticket.Yo bag Gucci but you still ugly hoe shut up.@DFkinLopes LMfAOOO this bitch tongue longer than some niggas dicks.

Fig.6 Offensive Message.



College Blog Home

Post contains inappropriate content and can not be posted

nishita 2020-12-03
myfriend
sam is my best friend

kajal12 2020-12-02
wedding
rushikesh is the groom

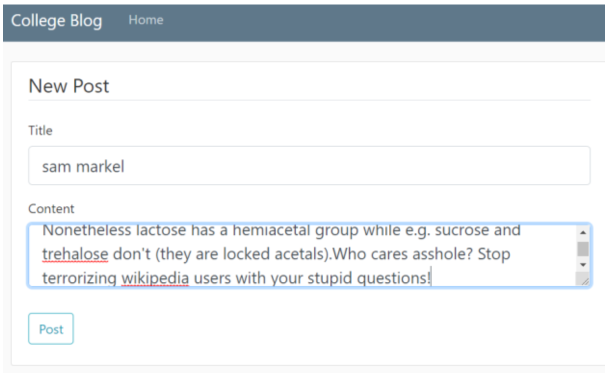
xyz 2020-12-02
tejas
tejas is a good boy

1 2

Fig.7 Blocking of Offensive message.

6. Mixture of Offensive and Non Offensive.

After if the user wants to post the message that contains a combination of both offensive and non offensive message, our system will block the message and would not allow the user to post the message on the wall.



College Blog Home

New Post

Title

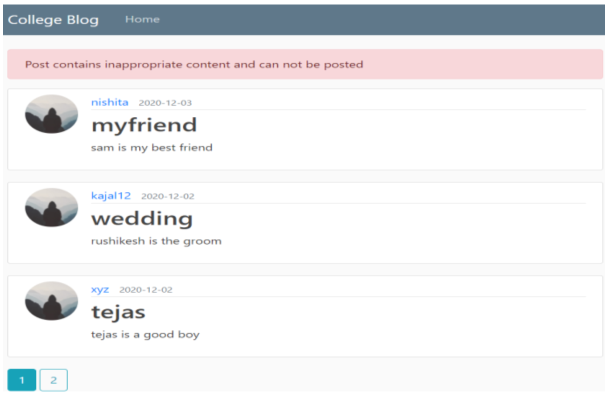
sam markel

Content

Nonetheless lactose has a hemiacetal group while e.g. sucrose and trehalose don't (they are locked acetals).Who cares asshole? Stop terrorizing wikipedia users with your stupid questions!

Post

Fig.8 Posting Offensive and not offensive messages.



College Blog Home

Post contains inappropriate content and can not be posted

nishita 2020-12-03
myfriend
sam is my best friend

kajal12 2020-12-02
wedding
rushikesh is the groom

xyz 2020-12-02
tejas
tejas is a good boy

1 2

Fig.9 Output Screen.

7. Result Analysis.

We have calculated accuracy of the system under 4 circumstances using 4 different algorithms. Accuracy can be calculated by:

Accuracy = correct prediction /total prediction*100.

Table 1. Comparative Analysis of Results

Algorithm	Accuracy	Precision	Recall	f1_score
NB	92.76	81.32	84.24	82.75
RF	94.11	95.65	73.84	83.34
SVC	95.24	90.22	85.5	87.8
LR	95.6	93.86	83.36	88.3

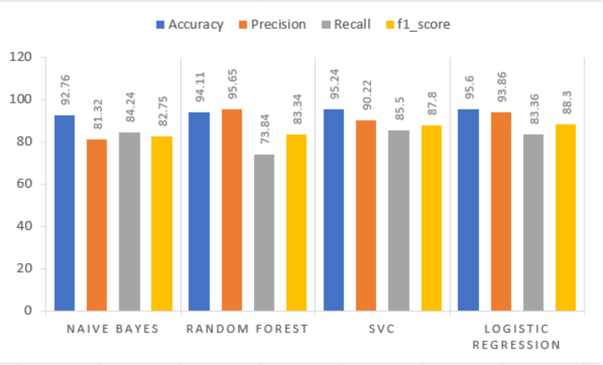


Fig.10 Comparative Analysis of Results

We can observe from the table that Logistic Regression is performing better in terms of accuracy as well as F-measure. SVC is giving approximately similar accuracy but precision is lower as compared to LR. Random forest is providing better precision but lesser accuracy than LR. Hence we chose LR to develop our final system.

8. Confusion Matrix.

A confusion matrix is a table where the performance of the classification or classifier is described. Here LR, Naive Bayes, SVC and RF confusion matrix are compared.

Confusion Matrix can be calculated by the total number of two correct predictions(TP+TN) divided by the total number of dataset (P+N). Here P=Positive, N=Negative

We are representing a confusion matrix for the two best performing algorithms.

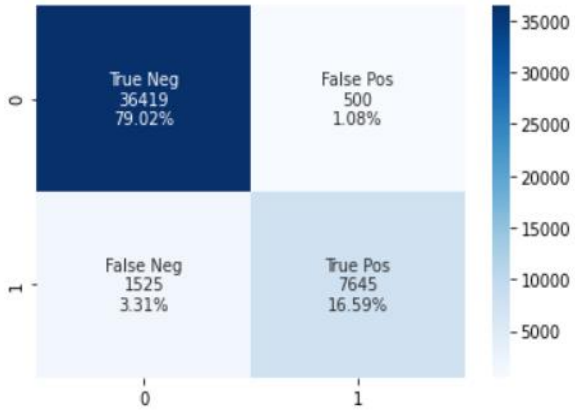


Fig.11 LR Confusion Matrix.

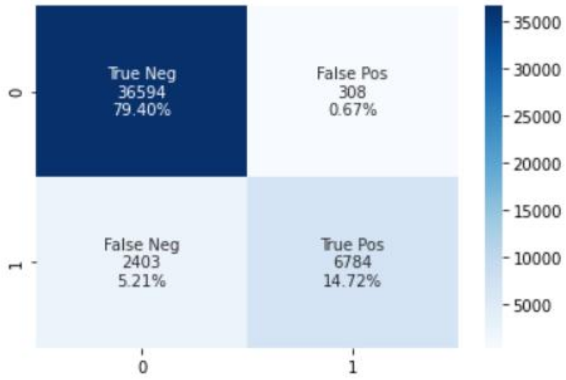


Fig.14 RF confusion matrix.

5. FUTURE SCOPE

The Application performs certain methods. There are many features that could be included in the project such as:

I. The future work can be done on image and video filtering.
II. Automatically blacklisting the user which is continuously posting offensive messages.
III. Online learning mechanisms can be used to make the machine learn and train itself, making it free from human supervision.

6. CONCLUSIONS

In this paper, we presented a Message filtering system using an online social network, our model provides filtration of unwanted messages and blocks the message to be posted on the wall.

This application is used to filter the messages which are unwanted or offensive. This can be done by using the Logistic Regression algorithm. This method is easy to implement which gives an excellent performance with less complexity. Even though it is an easy approach it provides and effective efficiency. The Future scope can be video filtering and image filtering.

REFERENCES

[1] Snehal D'mello, Dakshata Panchal, Sweedle Mascarenhas, Department of Information Technology, St. Francis Institute of Technology, "Message classification and filtration from Social networking platforms using Radial Basis Neural ", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).

[2] Dhruba Jyoti Mishra, Dr. Sangeeta Kakoty, "Message Filtering in Social Network Using Expert System", Journal of the Gujarat Research Society, Vol.21 Issue 14, Nov.2019.

[3] Nitin Pondhe, Prof. H.B.Jadhav, Asst. Prof. (Computer Engineering) VACOE, " A System to Filter Unwanted Messages on Social Networking Site" , Vol-3 Issue-1,2017.

[4] Garima Singh, Prof. Amit Yerpude , Prof. Toran Verma, Rungta College of Engineering and Technology, Dept. of Computer Science and Engineering, "Filtration of Unwanted Text Messages from Online Social Networks", 5(7):345-349, Jul 2017.

[5] KARTHIK, DR.SAVITA CHOUDHARY, Assistant Professor, VidyaVardhaka College of Engineering, "TaCbF-"Trending Architecture for Content based Filtering using Data Mining" , International Conference on Current Trends in Computer, Electrical, Electronics and Communications (ICCT CEEC-2017).

[6] M. Vanetti, E. Binaghi, B. Carminati, E. Ferrari, M. Carullo A System to Filter Unwanted Messages from OSN User Walls IEEE Transactions on Knowledge and Data Engineering, 25 (2013).