Research Paper on Radial Basis Function Neural Network for Content Extraction

Komal Patil[1] ,Vidita Patel[2], Maulika Patel[3]
[123]Information Technology Department, Parul Polytechnic Institute, India

Abstract- Online social networks play an important role in personnel and business life of many individuals. The rapid usage of OSN leads to many issues like privacy, misleading and unrelated information/messages on user's walls. In this paper we are addressing an issue of unwanted messages. We are using rule based system to customize the user's messages on the walls and a support vector machine classifier to filter and remove unwanted messages automatically using content-based filtering. Online Social Networking in these days is most powerful way to share the Information, thought, event and many more.

Keywords- online social networks, message filtering, support vector machine classifier, rule-based system

I. INTRODUCTION

Online Social Networks plays an important in the usage of internet applications. Millions of people are using Online Social networks like Facebook, Twitter, Linked in etc. These applications improve the social and public relationships. Each user can maintain his personnel information like date birth, interests, hobbies, current location, job and so on. This information is completely private to the user and some part of information can be visible to his friends, and relations. User can share some data like status, images and any content on his friends and public profile. By using this public information and information of users friends of friends the attackers can be get the personnel information of the user. The gathered information can be misused by attackers to publish adds etc. Along with privacy controlling information on user's walls is another major issue in Online Social Networks. There is a very high chance of posting unwanted content on particular public/private areas, called in general walls. So, to control this type of activity and prevent the unwanted messages which are written on user's wall. we can implement filtering rules (FR) in our system. Also, Black List (BL) will maintain in this system. OSNs provide support to prevent unwanted messages on user walls. For example, Facebook allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends). However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them[1][2][3].
Most normal intuitive medium to convey is online social network. A few sorts of data or substance will be shared between the clients, the kind of substance are sound, video, pictures and so on. As the Amount of substance will be extremely limitless data sifting is utilized. OSN give less measure of security in posting undesirable messages. Data sifting is utilized for disconnected reason. Capacity of a client to naturally control the messages composed on the client all, by sifting extra correspondence will be termed as data separating[4].

Social network holds an architecture that must have a profile, friends, groups, Discussions, Widgets. However all these are optional. A worthy point to note is that Social networks may also have a few detriments [1].It becomes an open ground for hackers to commit faults and launch virus attacks. It may also result in scams and false use of data or information [1]. Sharing of information may sometimes be vulgar and it is inevitable. In this paper, we propose an idea of "Filtered wall" on the basis of personalized filtering recommendation algorithm [2]. This notion is fundamentally developed to preclude the appearance abusive words on the user wall [5]. An author adopt the concept of expert analysis wherein a third part is given the utmost importance. The third party serves as a critic and he checks with the available documents for the presence of those abusive words. If present, the word is blocked as per the third party's decision[2].

Machine Learning techniques are used to categorize the messages based on their contents. The content filtering is designed to control what contents may or may not be displayed. This paper focus on filtered wall architecture, it's three tier architecture. Using this text from messages and images are filtered. Web content mining strategies are designed for OSN to automatically discover the useful or unwanted information hidden within the data. Filtered wall intercepts the message the user tries to post. Machine Learning based classifier extracts meta data. Enforce the filtering and Blacklist rules and Message will be published or filtered by Filtered wall. Radial Basis Function networks classifier is used to categorize the message because its handle the noise data effectively. Filtering rules can be used to state what contents should be accepted and rejected. And also there is a list of users that are temporarily or permanently prevented from any kind of posts in a authors wall and this list is known as Black list. The unwanted messages are filtered from OSN walls on the basis of both message content and the message creator relationship and character[6].

II.   RELATED WORK

The main goal of this paper is to design a system to provide customizable content based message filtering for online social networks, based on machine learning techniques. Information Filtering Systems are designed to categorize the information which are generated dynamically and offer the information to the user fulfill their requirement [2]. In the content – Based Filtering system, each user is assumed to operate separately. So the filtering system selects the information based on the correlation between the content of the items and user preferences.

It is contradictory with the collaborative filtering system which selects information based on the correlation between the people with similar preferences [3], [4]. Content-based filtering mostly

processes the textual document in nature and this builds content-based filtering close to text classification. Content-based filtering is mainly based on the use of the ML paradigm. In that, a classifier is automatically induced by learning from a set of preclassified examples. The feature extraction procedure maps text into a compact representation of its content, which is uniformly applied to training and generalization phases[3].

There are varieties of key approaches in content-based filtering and text classification. Based on the application, each approach may having mutual advantages and disadvantages. In depth comparison analysis, has been conducted to verify the superiority of classifiers such as Boosting-based classifiers[7], Neural Networks [8], [9], and Support Vector Machines [10] over other popular methods, such as Rocchio [11] and Naive Bayesian [12]. However, most of the work related to text categorization by ML has been applied for long form text and the evaluated performance of the text classification methods strictly depends on the nature of textual documents. In content based filtering, text mining technique are employed to categorize incoming messages

N.Thilagavathi and R. Taariaka presented an inference algorithm to remove unwanted content or messages from online social network user wall. This paper presented an architecture of filtered wall which is used for information filtering. Information filtering is the process of providing appropriate information to the people who need it. The main goal of this paper is to design a system to provide customizable content based message filtering for online social network based on machine learning technique[1]. This paper mainly focuses on content based filtering. Inference algorithm determine the entailment relationship between the sentences and generate the new rule. By using this algorithm the incoming post has been filtered out based on the content[1].

Subramaniyaswami V., Logesh R. developed a paper which automatically filtered an unwanted messages from Online Social Network. An author has presented a filter wall to permeate offensive messages using rule based and text

classification technique. In this paper the used technique is naive Bayesian machine learning algorithm.

Also this paper has developed a filtering mechanism and Black list which is used for filter unwanted comments and block the user temporarily for few days who sends the abusive messages continuously. The text mining concept is also used in this paper. The short text classifier is divided in two parts i.e. stop word removal and stemming[2]. In this paper the naive Bayesian is primarily utilized for decision making and inferential statistics that handles probability inference[2].

Macro Vanetti, ElisabettaBinaghi, Elena Ferrari presented a system to filter unwanted messages from OSN user wall. A system in this paper allow the user to have direct control on the messages posted on their wall. This is achieved using flexible rule based system, and machine learning based soft classifier. The aim of this paper is to provide users a classification mechanism to avoid useless data. In this paper an author have evaluated an automated system called filter wall which is able to filter unwanted messages from OSN wall[3]. Here the technique used is Radial Basis Function Neural Network with machine learning algorithm Naive Bayesian. The main contribution of this paper is design of a system which provide customizable content based message filtering for OSN based on ML technique

III.   FILTERED WALL ARCHITECTURE

USER REGISTRATION:-
In order to be a part of the Online Social Network, an user has to first get himself/herself registered in the Online social network. The registration process would consists of the user giving his/ her details like his name, age, occupation, mail id , so on which are meant to be genuine. The details that the user has registered in the Online Social Network are acquired and stored in the Data Store Blob Store, a GUI based Data Base that can store terabytes the data store.[2]

FRIEND REQUEST AND MESSAGES POSTING ON WALL:
The authenticated user on the Online Social Network has the privilege to send friend request to another authenticated user in the OSN by issuing a friend request to that particular person. Once the other person has accepted the friend request, then, both the users can share with each other their ideas in the form of pictures, status, videos. For each person in the friend list, user will maintain some of their details like relationship strength, trust kind of relationship, etc..in a database.[2]

In Fig1.user first have to login in his/her account then and the he/she will be able to access the page. After that user can see the messages, can comments on particular image as well as can chat to the friends. If user want to add any friend then he has to send a friend request to that friend by adding him/her then there will be communication and sharing of information or messages takes place between two users. There are lots of people who sends abusive messages on social networking site

,so this paper presents a method to block such users permanently and filter such messages using machine learning text classifier Support Vector Machine. Support vector machine classifies messages into positive or negative points using hyperplane.

We are proposing new way of blocking the unwanted messages on users walls by using support vector machine classifier and filtering rules. The proposed system contains the following modules called
1.   Short text classifier
2.   Blocking the unwanted messages as well as user.
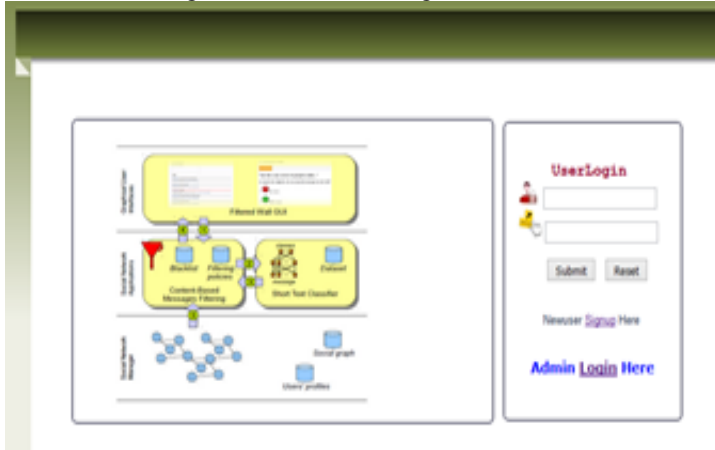
Fig.1: login page for the user

Fig.2: User wall after message filtration

Above fig. 2 shows the result about filtered messages. When user post an abusive message then it filters that message using support vector machine. It is used as classifier and shows the categories of the filtered messages.

1. Short Text Classifier:
Short text classifier aims to classify the messages according to a set of categories. Classifier is build to extract and select discriminating features of the short text messages. Aim of the short text classifier is to recognize and eradicate the neutral sentences and categorize the non neutral sentences in step by step, not in single step. This classifier will be used in

hierarchical strategy. The first level task will be classified with neutral and non neutral labels. The second level act as a non neutral, it will develop gradual membership. These grades will be used as succeeding phases for filtering process. Short text classifier includes text representation and support vector machine classification. Short text classification is comprised with two phases: Text representation and Machine Learning based classification. A support vector machine classifier is employed to classify the incoming message. It automatically categorizes the short message into the suitable category, which are neutral or non neutral messages. Non neutral messages are further to analyzed to determine the appropriateness to each category.[1][3].

2. Support Vector Machine:
A support vector machine (SVM) is an algorithm that uses a nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for the linear optimal separating hyperplane. A hyperplane is a "decision boundary" separating the tuples of one class from another. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane. The SVM finds this hyperplane using support vectors ("essential" training tuples) and margins (defined by the support vectors).
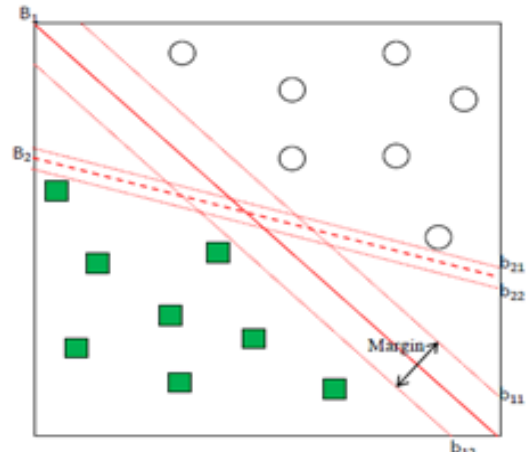

Fig.3: Support Vector Machine

SVMs are a generally applicable tool for machine learning. Support Vector Machines (SVMs) have been proven as one of the most powerful learning algorithms for text categorization. SVMs can handle with exponentially or even infinitely many features, because it does not have to represent examples in that transformed space, the only thing that needs to be computed efficiently is the similarity of two examples. Redundant features (that can be predicted from another features), and high dimension are well-handled, i.e. SVM does not need an aggressive feature selection. There are error-estimating

formulas, which can help us in predicting how good a classification of an unseen example.

3. Benefits of SVM

SVMs can handle with exponentially or even infinitely many features, because it does not have to represent examples in that transformed space, the only thing that needs to be computed efficiently is the similarity of two examples. Redundant features (that can be predicted from another features), and high dimension are well-handled, i.e. SVM does not need an aggressive feature selection. There are error-estimating formulas, which can help us in predicting how good a classification of an unseen example will be, eliminating the need for cross validation techniques [6].

4. Neural Network:

Neural network is the classifier. Neural networks are typically organized in layers. These layers are made up of a number of interconnected nodes which contain an activation function. Patterns are presented to the network via the input layer which communicate to one or more hidden layers where the actual processing is done via a system of weighted connection. The hidden layer then link to an output layer where the answer is output. Typically neural network has anything from a few dozen to hundred, thousands or even billion of artificial neurons, called units, arranged in a series of layers, each of which connects to the layer on either side.[13].
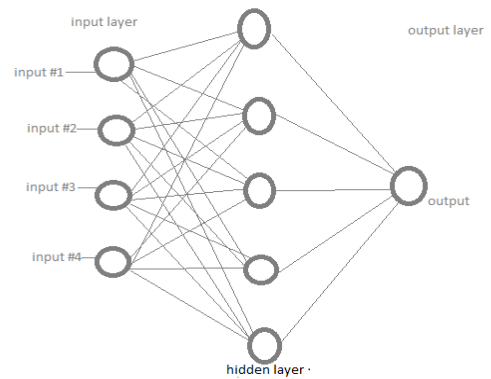

Fig.4: Architecture of Neural Network [14]

Input unit or layers receives various forms of information from the outside of the network. The connection between one unit and another are represented by a number called weight which can be either positive or negative. Information flows through a neural network in two ways. When its learning, patterns of information are fed into the networks via the input units, which triggers the layer of hidden unit and these in turn arrive at the output units. This is called feed forward network.

5. Radial Basis Function Network

Radial basis function networks are a special class of single hidden layer feed forward neural network for application to problems of supervised learning. It is a particular type of neural network. An RBFN performs classification by measuring the inputs similarity from the training set. When we want to classify a new input, each neuron computes the Euclidean Distance between the inputs and its prototype. [8]
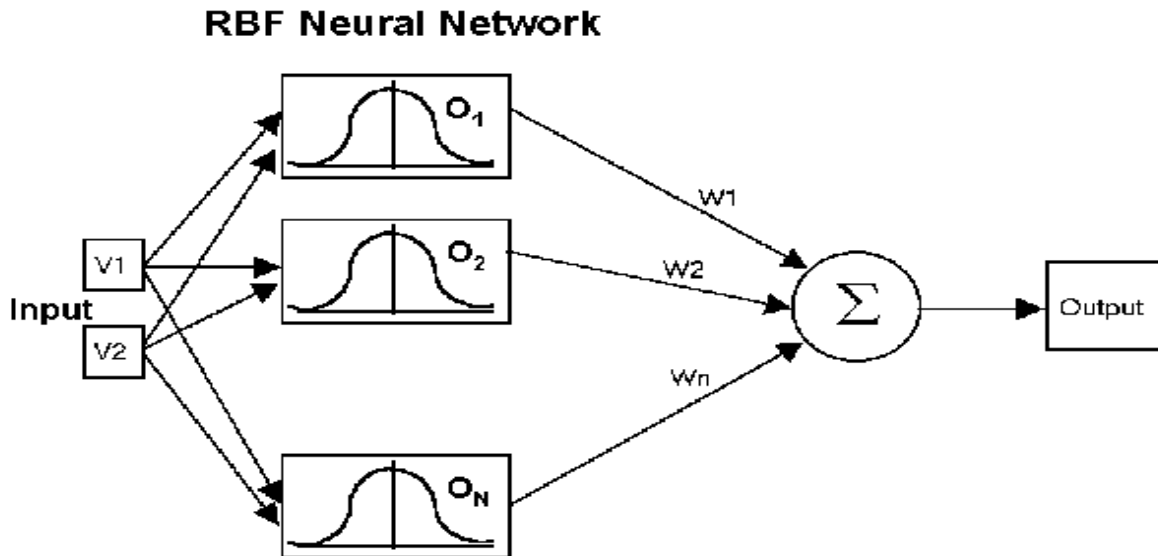

Fig.5: Radial Basis Neural Network Architecture [8]

Figure shows an architecture of RBFN. It consists of an input vector, a layer of RBF neurons and an output layer with one node per category or class of data. The input vector is the n-dimensional vector that you are trying to classify. The entire input vector is shown to each of the RBF neurons. Each RBF neurons stores a prototype vector which is just one of the vectors from the training set. Each RBF neurons compares the input vector to its prototype and output a value between 0 and 1 which is a measure of similarity. If the input is equal to the prototype , then the output of that RBF neurons will be 1.[8]

IV. CONFUSION MATRIX

In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as error matrix is a specific table layout that allow visualization of the performance of an algorithm, typically a supervised learning. A confusion matrix is a table that is often used to describe the performance of a classification model or classifier on a set of text data for which the true value are known.[15]

		predicted class	
		0	1
true class	0	True Positive (TP)	False Negative (FN)
	1	False Positive (FP)	True Negative (TN)

Fig.6: Confusion matrix [19]

Confusion matrix is a table that is often used to "describe the performance of a classifier or classification model" on a set of text data for which the true values are known. There are two possible predicted class "Yes" and "No".

Yes = there is a vulgar word in the posted message.

No = there is no vulgar word in the message.

TP = these are the cases in which we predicted "yes" there is an abusive message and it actual have.

TN = we predict No, and there is no abusive message.

FP = we predicted yes, but in actual there is no abusive word.

FN = we predicted No, but there is actually an abusive message .

In classification problems the primary source for accuracy estimation is the confusion matrix.

Accuracy = TP + TN / TP + TN + FP +FN

Precision = TP / TP + FP

Recall = TP / TP + FN. [19]

V. ALGORITHM AND FLOW OF WORK

Input: Messages, training data set
Output: Blocking messages
Algorithm:
Step1: remove unnecessary words like punctuation marks, exclamation mark, Question mark etc.
Step 2: Train the support vector machine classifier to classify the bad words and good words. And obtain the weights for testing any images.
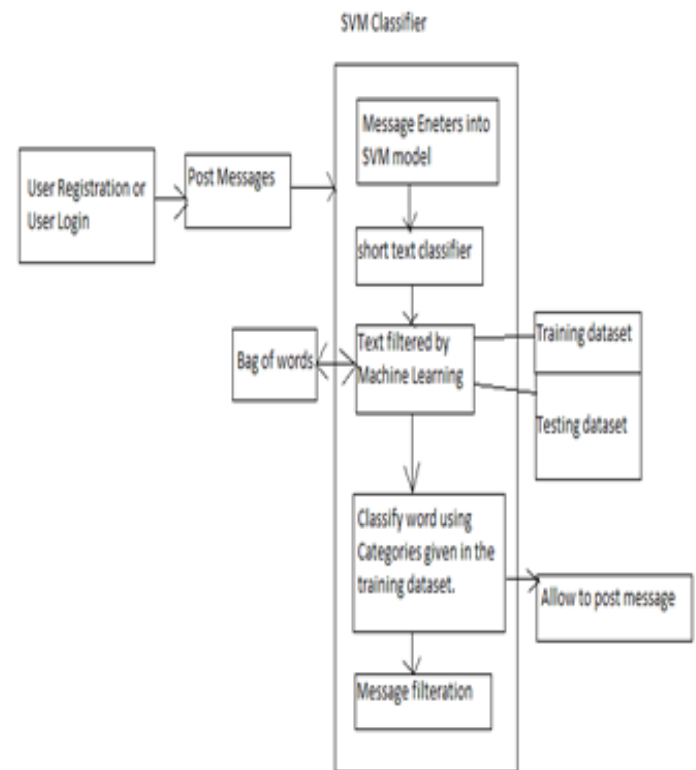**Step3:** Block the message by using support vector machine classifier.

Fig.7: flow chart of the system

As the flow chart shows, user first have to do login to the account then and then he/she can connect to the social site. If user don't have an account then he/she fir have to create it. After creating an account user can connect with his friends. Also can share information. When user post a message, it enters into machine learning module i.e. Support Vector Machine module. Then short text classification is applied to that message. After applying short text classification, stop words are removed out. Text filtration using Support Vector Machine can done. Here training and testing datasets are

presents. Also Bag of Words is presents. After text classification, words are categorised using training dataset. The categorise are mentioned in training dataset. Here the words are checked in training dataset if the word is presents in

dataset then consider that the word is non neutral and applying filtering rule it is filtered out. If the word is not present in word list then consider that it is neutral and allow to post.

IV. EXPERIMENTAL ANALYSIS

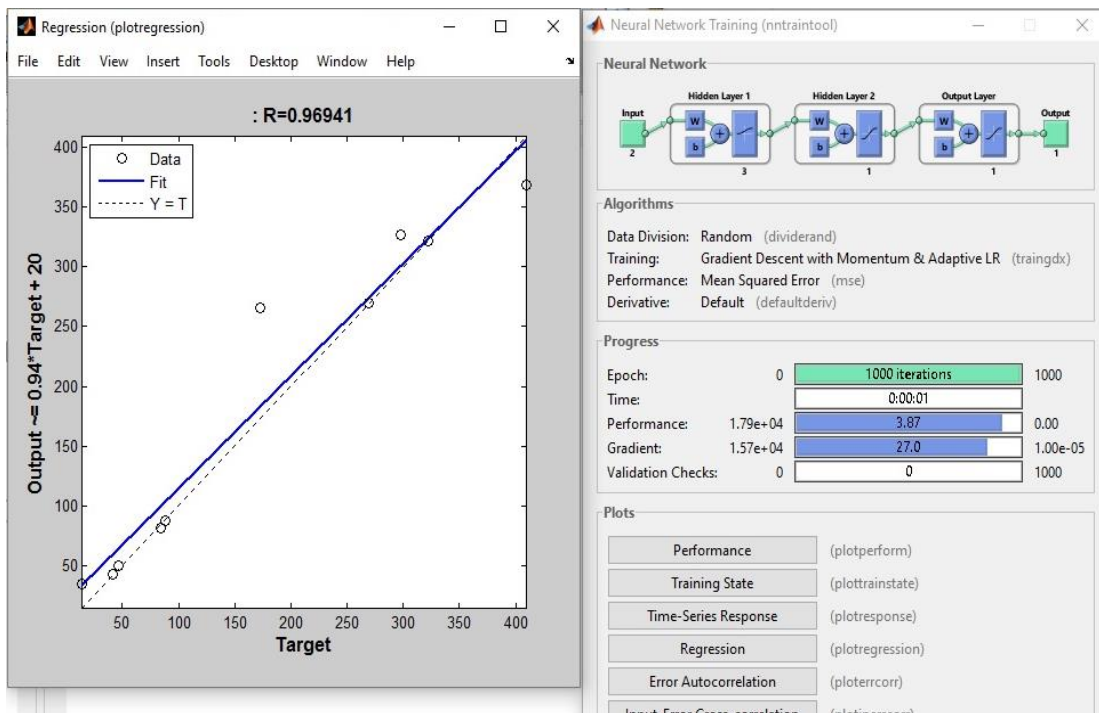Fig.8: Neural Network Algorithm

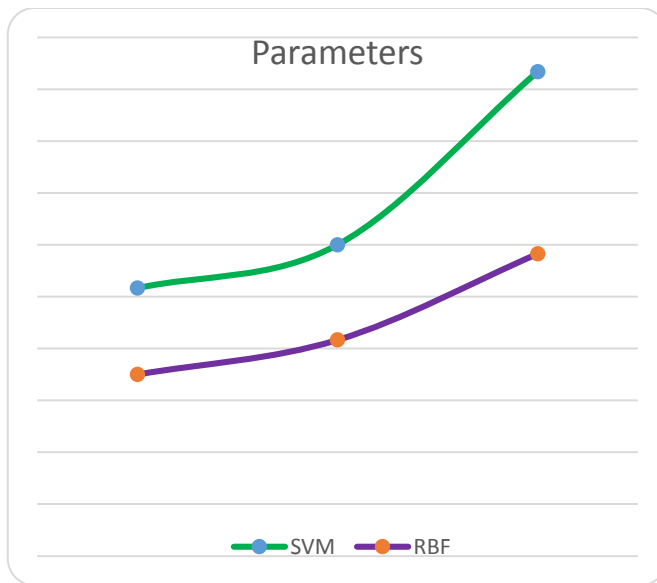Fig.9: Neural Network in Matlab
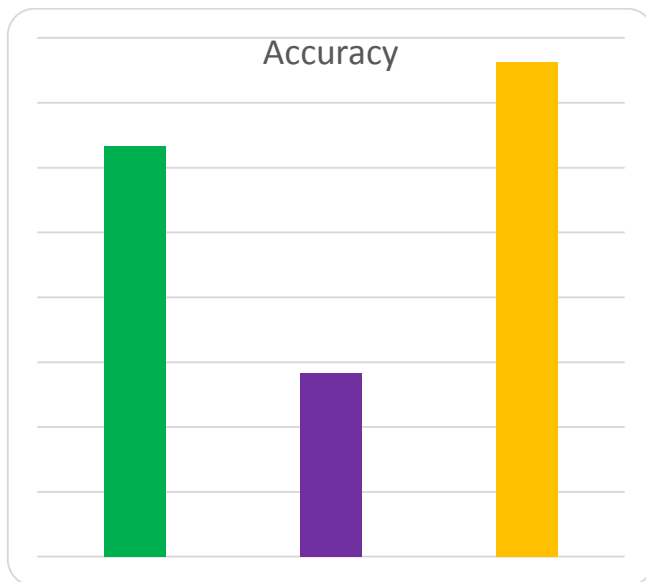
Fig.10: Graph using parameters

Fig.11: Accuracy Graph of all Algorithms

V. CONCLUEESION AND FUTURE ENHANCEMENT

Use of Internet and Online Social Network is increasing day by day. But lots of unwanted messages are also occurs on the net. These comments can cause harm or can create misunderstanding between peoples. Thus in this proposed work, the filtration of text or comments are done using short text classifier. It will first remove stop words, then perform stemming and then it will calculate the measures. The user is blocked in black list for temporary. Due to this unwanted messages are filtered out and it will increase the efficiency of OSN user wall. In Future Work the abusive pictures or video can be deleted or blocked.

VI. REFERENCES PAPERS

[1]. SubramaniyaswamyVa, LogeshRb,VijayakumarVc, and IndragandhiVd. "*Automated Message Filtering System in Online Social Network",* Elsevier , 2015.

[2]. Macro Vanetti, ElisabettaBinaghi, Elena Ferrari, Barbara Carminati and Moreno Carullo "*A System to Filter Unwanted Messages from OSN User Wall*", Transaction on Knowledge and data Engineering, IEEE February. 2013

[3]. N.Thilagavathi, R.Taarika, "Content Based Filtering in Online Social Network using Inference Algorithm" , IEEE 2014.

[4]. A. D. Swami, B.S Khade, "*Text Based Filtering System for OSN User wall*", International Jornal of Advanced Research in Computer Science and Software Engineering, IEEE February 2014.

[5]. GediminasAdomavicious, Alexander Tuzhilin, "*Towards the next generation of recommender system: A survey of the state of the art and possible Extension*", IEEE, 2005

[6]. K.Babu, P. Charles, "*A system to filter unwanted words using Blacklists in Social Networks* ", IJCSIT 2014

[7]. J. Anishya Rose, A. Pravin "*Machine Learning Text Categorization in OSN to Filter Unwanted Messages*", IJCSIT , July 2014.

[8]. IstvanPilaszy, "*Text Categorization and Support Vector Machine* ", 2010.

[9]. Moreno Carullo, ElisabettaBinaghi, Iganazio Gallo, "*An Online Document Clustring Technique for Short Web Content*", Elsevier April ,2009.

[10]. Nicholas J.Belkin, W. Bruce Croft, "*Information Filtering and information Retrival: Two Sides of the same Coin*", Dec, 1992.

[11]. Thorsten Joachims, "*Text Categorization with Support Vector Machines: Learning with Many Relevant Features*" 1999.

[12]. P. Narendra, T. Subhashini, R. Pitchaiah, "*Neural Network Classifier to filter unwanted messages from OSN user walls* ", IJMTARC, 2013.

[13]. VikramadityaJakkula , "*Tutorial on Support Vector Machine (SVM)*"

[14]. Yun-Nung Chen, Che-An Lu, Chao-Yu Huang, "*Anti-Spam Filter Based on Naïve Bayes,SVM, and KNN model*", AI Term Project Group, 2014.

[15]. Qiang Wang Yi Guan Xiaolong Wang," *SVM-Based Spam Filter with Active and Online Learning* ".