Mining Social Texts and Images to Filter Unwanted Content

T. Swapna, B. Sandhya Rani
Gitam Deemed To Be University, India

ABSTRACT:As internet grows quickly, socialnetworkings have become part of daily life where people can communicate with each other disseminating the multimedia information. Therefore online social networking sites are perfect for exchanging public opinions. Online social networking sites provide the security measures but they are limited. In OSN sites like face book, twitter etc there is possibility of posting any kind of data on user wall. Where such kind of data may contain unwanted messages and images which may be harmful to children and may affect the efficiency of people using OSN sites. So to prevent this problem we propose a system that provides user to control the unwanted messages posted on their wall using short text classifier and machine learning text categorization techniques in support of content-based filtering are used and also unwanted images are restricted using KNN classification.

Keywords: *Online-Social Networks, Information Filtering, Short-Text Classifier (Naïve Bayes Classification), Content Based Filtering, K-nearest-neighbor (KNN) Classification.*

I.INTRODUCTION

Millions of people share their opinion on Online Social Networking sites which are most popular interactive medium for public opinions exchanging miscellaneous type of data such as text, images, audio, video etc. The social networking sites have become a part of daily life and have brought drastic changes in communication between people. But the Online social networking sites are providing limited support to avoid unwanted messages on user walls such as face book where content based filtering is preferred for the short texts that occur in message. . An online social network like facebook, twitter, etc. , there is posting facility using which user can directly post any kind of data like images, text messages, audio, video, etc. hence there may be possibility of posting any kind of data on user wall. Such data may contain unnecessary messages or images. For example: malicious political statement, vulgar data, personal teasing statement etc., which are publicly available to friends of wall owner. Also wall owner's friendscancommentonitwhichisalsopubliclyavailable.Such post may affect user image in social networking systems and unnecessarily he will have to keep explicit watch on such own wall content which is not possible. Up to a certain extent some existing schemes like facebook allows users to define, who is allowed to put messages on their walls. However, no content-based preferences and filtering are supported and thusly, it is impossible to prevent posting of such undesired messages. To protect undesired message posting on user wall and to protect user social image is an important issue on social networking site.The motivation behind this work is to avoid overwhelmed used of unnecessary data on user's wall. As we consideredsome existing system [1] like facebook, which permitted users to define who is allowed to insert messages on their walls .But content based filtering is not provided.Our system should filtered the unwanted text and also images and enforces protection and productivity policies for business, schools and libraries to reduce legal and privacy risks while minimizing administration overhead. Filtering provide network administrator with greater control by automatically acceptable used policies.

Therefore in this paper we propose classification mechanisms in order to avoid useless data. In this paper our aim is to analyze the classification technique and to design the system to filter the undesirable messages from OSN user wall. Our present work suggests and experimentally estimates an automated system which is called Filtered Wall (FW) that should be able to filter unwanted messages from OSN user walls. Machine learning text categorization techniques are evolved to automatically assign with each short text message based on its content by using a set of categories. As internet grows quickly,pornography has become one of highly distributed information over the internet which may be harmful to the people using internet. Therefore computer must go through a series of steps in order to classify a single image. Thus image classification techniques such as K-nearest-neighbor method are used to classify the images of good and bad images.

II. RELATED WORK

Recommender systems works in three main ways, the Content-based filtering, Collaborative filtering, policy-based personalization.

Content based filtering

Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile, using information retrieval techniques such Term Frequency and Inverse Document frequency (TF-IDF) . The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user.

Text features

Profile=set of important words in item (document).

To pick the important words the usual heuristic used from text mining is TF-IDF(Term frequency * Inverse Document Frequency)

F $_{ij}$=frequency of item(feature) i in doc(item) j

N_i=number of documents that mention term i

N=total number of documents

IDF_i=log N/n_i

TF-IDF score:wij=TF_{ij}*IDF_i

Document profile=set of words with higest TF-IDF scores,together with their scores.

Several issues have to be considered when implementing a content-based filtering system. First, terms can either be assigned automatically or manually. When terms are assigned automatically a method has to be chosen that can extract these terms.

III. PROPOSED SYSTEM

The main contribution of this paper is, to propose and experimentally evaluate an automated system ,called Filtered Wall(FW),able to filter unwanted messages from OSN user walls .We exploit Machine Learning (ML)text categories to automatically assign with each short text messages a set of categories based on its content. The major efforts in building a robust short text classifier are concentrated in the extraction and selection of a set of characterizing and discriminate features.

The solutions investigated in this paper are an extension of those adopted in a previous work by us from which we inherit the learning model and the elicitation procedure for generating pre-classified data. The original set of features, derived from endogenous properties of short texts, is enlarged here including exogenous knowledge related to context from which messages originate .As far as learning model is concerned , we confirm in the current paper the use of neural learning which is today recognized as one of the most efficient solutions in text classification. Moreover, the speed in performing the learning phases for an adequate use in OSN domains, as well as facilitates the experimental tasks.

Content Diagram of the Proposed System

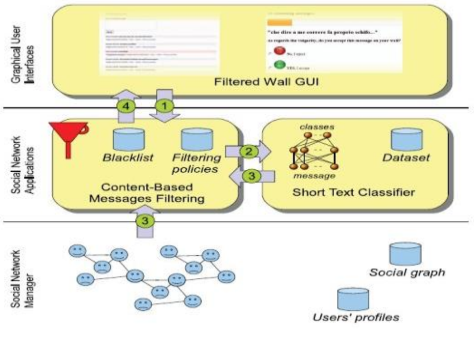


Fig 1. Architecture Diagram

The application is web based system so it requires server, browsers, scripting languages, internet which is the below architecture.

The architecture of our system is a three-tier structure. The first layer, called Social Network Manager(SNM), commonly aims to provide the basic OSN functionalities(i.e., profile and relationship management), whereas the second layer provides the support for external Social Network Applications (SNAs). The supported SNAs may in turn require an additional layer for their needed Graphical User Interfaces (GUIs). According to this reference architecture, the proposed system is placed in the second and third layers. In particular, users interact with the system by means of a GUI to set up and manage their FRs/BLs. Moreover, the GUI provides users with a FW, that is, a wall where only messages that are authorized according to their FRs/BLs are published. The core components of the proposed system are the Content-Based Messages Filtering (CBMF) and the Short Text Classifier modules. The latter component aims to classify messages according to a set of categories. In contrast, the first component exploits the message categorization provided by the STC module to enforce the FRs specified by the user. BLs can also be used to enhance the filtering process. The path followed by the message, from it's writing is summarized as follows:

1. After entering the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by FW.

2. A ML-based text classifier extracts metadata from the content of the message.

3. FW uses metadata provided by the classifier, together with data extracted from the social graph and user's profiles, to enforce the filtering and BL rules.

4. Depending on the results of the previous step the message will be published or filtered by FW.

1. Filtering rules

In defining the language for FRs specification, we consider three main issues that, in our opinion, should affect a message filtering decision. First of all, in OSNs like in everyday life, the same message may have different meanings and relevance based on who writes it. As a consequence, FRs should allow users to state constraints on message creators. Creators on which a FR applies can be selected on the basis of several different criteria; one of the most relevant is by imposing conditions on their profile's attributes. In such a way it is, for instance, possible to define rules applying only to young creators or to creators with a given religious/political view. Given the social network scenario, creators may also be identified by exploiting information on their social graph. This implies to state conditions on type, depth and trust values of the relationship(s) creators should be involved in order to apply them the specified rules. All these options are formalized by the notion of creator specification, defined as follows.

2. Online setup assistant for FRs thresholds

As mentioned in the previous section, we address the problem of setting thresholds to filter rules, by conceiving and implementing within FW, an Online Setup Assistant (OSA) procedure. OSA presents the user with a set of messages selected from the dataset discussed in Section VI-A. For each message, the user tells the system the decision to accept or reject the message. The collection and processing of user decisions on an adequate set of messages distributed over all the classes allows to compute customized thresholds representing the user attitude in accepting or rejecting certain contents. Such messages are selected according to the following process. A certain amount of non neutral messages taken from a fraction of the dataset and not belonging to the training/test sets, are classified by the ML in order to have, for each message, the second level class membership values.

3. Blacklists:

A further component of our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs is directly managed by the system, which should be able to determine who are the users to be inserted in the BL and decide when user's retention in the BL is finished. To enhance flexibility, such informationis given to the system through a set of rules, hereafter called BL rules. Such rules are not defined by the SNM, therefore they are not meant as general high level directives to be applied to the whole community. Rather, we decide to let the users themselves, i.e., the wall's owners to specify BL rules regulating who has to be banned from their walls and for how long. Therefore, a user might be banned from a wall, by, at the same time, being able to post in other walls.

Similar to FRs, our BL rules make the wall owner able to identify users to be blockedaccording to their profiles as well as their relationships in the OSN. Therefore, by means of a BL rule, wall owners are for example able to ban from their walls users they do not directly know (i.e., with which they have only indirect relationships), or users that are friend of a given person as they may have a bad opinion of this person. This banning can be adopted for an undetermined time period or for a specific time window. Moreover, banning criteria may also take into account users' behavior in the OSN. More precisely, among possible information denoting users' bad behavior we have focused on two main measures. The first is related to the principle that if within a given time interval a user has been inserted into a BL for several times, say greater than a given threshold, he/she might deserve to stay in the BL for another while, as his/her behavior is not improved. This principle works for those users that have been already inserted in the considered BL at least one time.

A. Filtered WallArchitecture

The main goal of the system is to filter the unwanted wall content posted by the other user on the particular user's wall. This post can be in text form or in an image form. The system should analyse the text / image content and allow desired content on the wall. In this system, when particular post is arrived to be published on his wall, all personal settings are considered and wall post is filtered accordingly.While filtering the text message, it is first checked that whether it is from an authentic user or not. If it is from an authentic user then its content is analysed and properly categorizes using text classification techniques. Then system checks whether user preference is matching with derived post category. If it is matched, then particular post is published as it is on hold till user permits it.

Message Filtering: For message filtering purpose, we haveto extracttextualdatafromuser'swall.Whenever any user upload textual data on general wall, with the help of Naïve Bayes Classification,thesystemcanclassifythose dataintodifferentcategorieswiththehelpofdataset.

Image Classification: K-nearest neighbor is used for classification of images.

Naïve BayesClassification Algorithm.

This algorithm is used for text classification. Text classification is the process of assigning various short texts provided by user to one or more target categories based on its content. .

The Naïve Bayes Algorithm is a Machine learning algorithm for classification problem. It is primarily used for text classification, which involves high-dimensional training data sets. A few examples are spam filtering,sentiment analysis and classifying news articles.

Bayes Theorem: Bayes theorem is stated as probability of the event B given A is equal to the probability of the event A given B multiplied by probability of A upon probability of B.

P(A/B)=P(B/A)P(A)/P(B)

P(A/B):probability(Conditional Probability) of the occurrence of event A given the event B is true.

P(A) and P(B):Probabilities of occurrence of event A and B respectively.

P(B/A):Probability of occurrence of event B given the event A is true.

Bayesian method of probability

A is called the proposition and B is called the evidence

P(A) is called prior probability of proposition and

P(B) is called prior probability of evidence.

P(A/B) is called the posterior

P(B/A) is the likelihood.

Posterior=(likelihood).(Proposition prior probability)/Evidence prior probability

Image Classification method

The problem of object classification can be specified as a problem to identify the category or class that the new observations belong to based on a training dataset containing observations whose category or class is known. Usually, classification works by first plotting training data into multidimensional space. Then each classifier plots testing data into the same multidimensional space as the training data and compares the data points between testing and training to find the correct class for each individual query point.

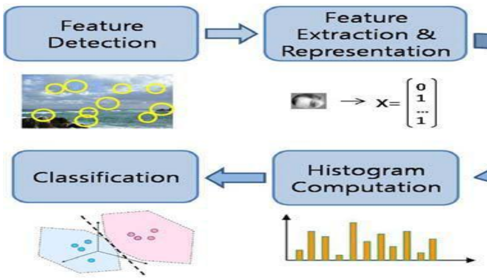


Fig. 1 A conceptual illustration of the process of image classification.

Text Classification method

Basic K-nearest neighbor classification

Training Method:save the training examples.

At prediction time: Find the K training examples(x_1, y_1)..(x_k, y_k) that are closest to test example x.

Classification: predict the most frequent class among those $Y_{i's.}$

Regression: predict the average among the $Y_{i's}$

To classify objects based on training examples in the feature space KNN is used. K-nearest neighbor is one of the simplest classification algorithms. Training process for this algorithm only consists of storing feature vectors and labels of training images. In the classification process, the unlabelled query point is simply assigned to the label of its *k* nearest neighbors. The object is classified based on the labels of its *k* nearest neighbors by majority vote. If *k*=1, the object is simply classified as the class ofthe object nearest to it. When there are only two classes, *k* must be a odd integer. However, there can still be ties when *k* is an odd integer when performing multiclass classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean distance:

$$d(x,y)=|x-y|$$

$$=\sqrt{(x-y)(x-y)}$$

$$=\left(\sum_{i=1}^{m}\ ((Xi-yi)^2)\right)^{1/2} \qquad (1)$$

Where x and y are histograms in $x=R^m$

Fig 1.shows visualizes the process of KNN classification.

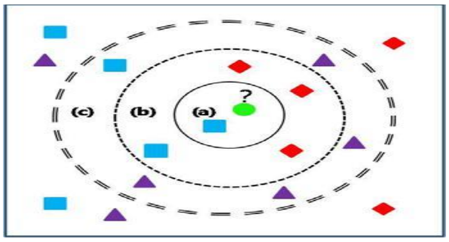


Fig. 2KNN Classification. At the query point of the circle depending on the *k* value of 1, 5, or 10, the query point can be a rectangle at (a), a diamond at (b), and a triangle at (c).

A main advantage of the KNN algorithm is that it performs well with multi -modal classes because the basis of its decision is based on a small neighborhood of similar objects. Therefore, even if the target class is multi-modal, the algorithm can still lead to good accuracy. However a major disadvantage of the KNN algorithm is that it uses all the features equally in computing for similarities. This can lead to classification errors, especially when there is only a small subset of features that are useful for classification.

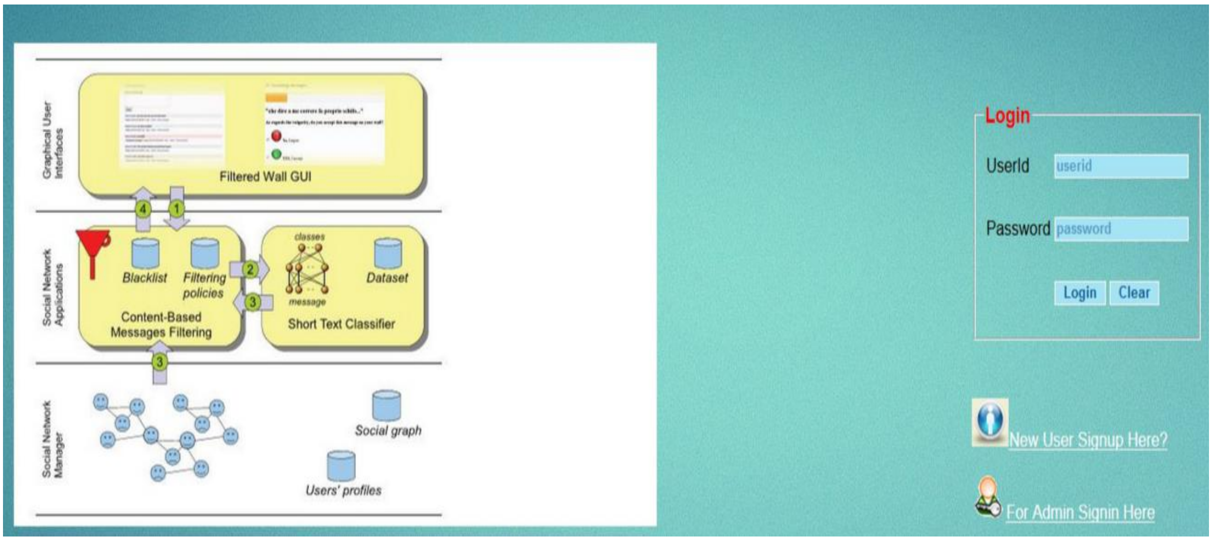
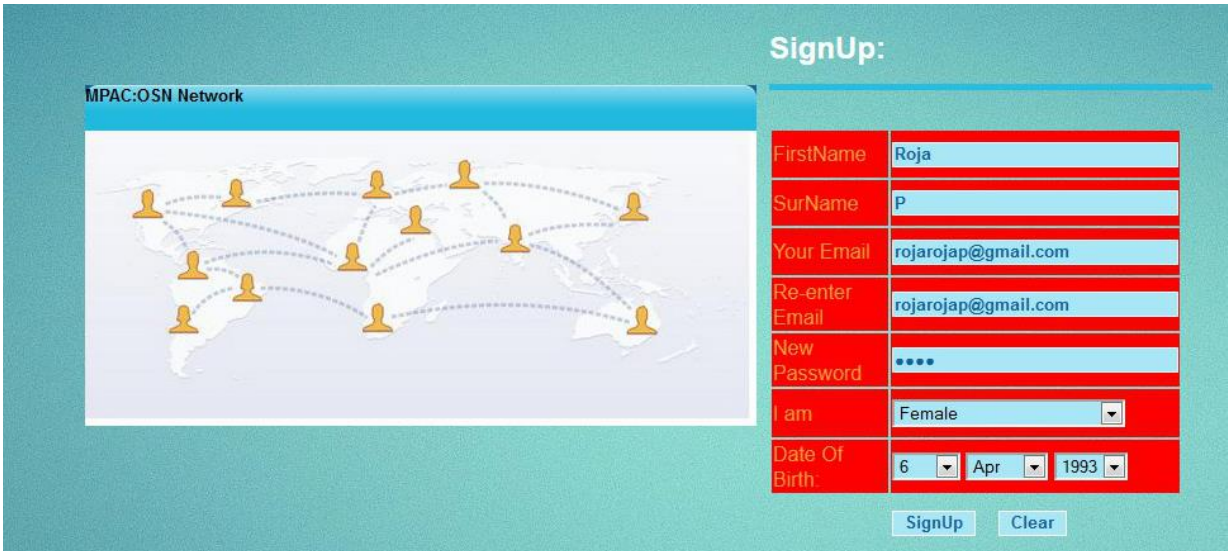


Fig 6.1: User and Admin Login Module

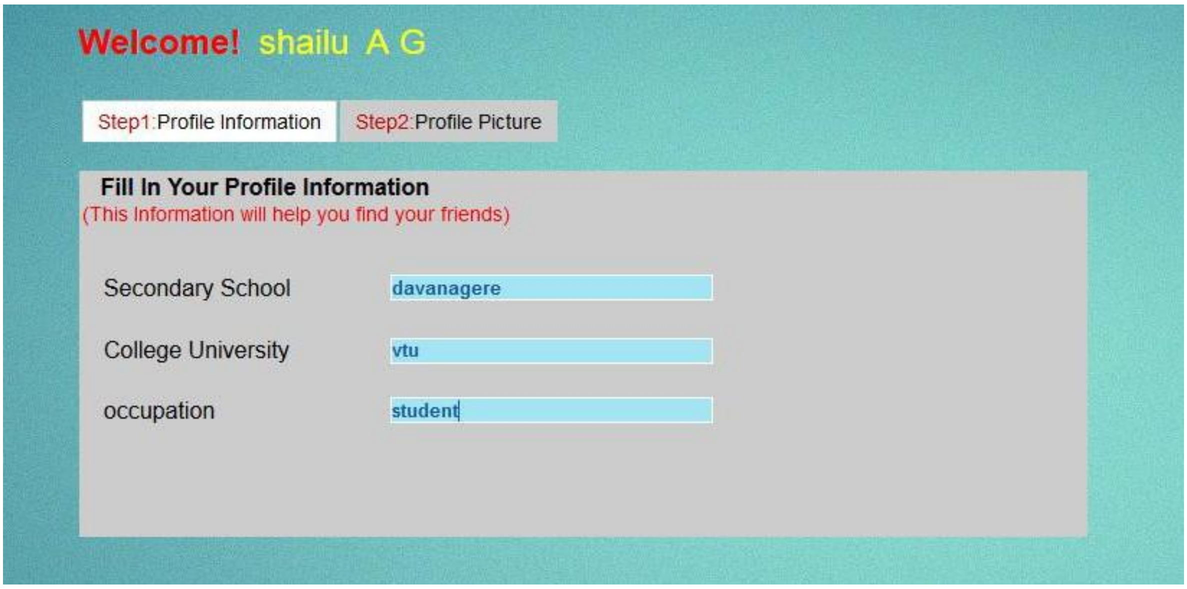


MPAC:OSN Network

SignUp:

FirstName	Roja
SurName	P
Your Email	rojarojap@gmail.com
Re-enter Email	rojarojap@gmail.com
New Password	••••
I am	Female
Date Of Birth:	6 Apr 1993

Fig 6.2: Registration Page



Welcome! shailu A G

Fill In Your Profile Information

(This Information will help you find your friends)

Secondary School	davanagere
College University	vtu
occupation	student

Fig 6.3: Registration Page Continued

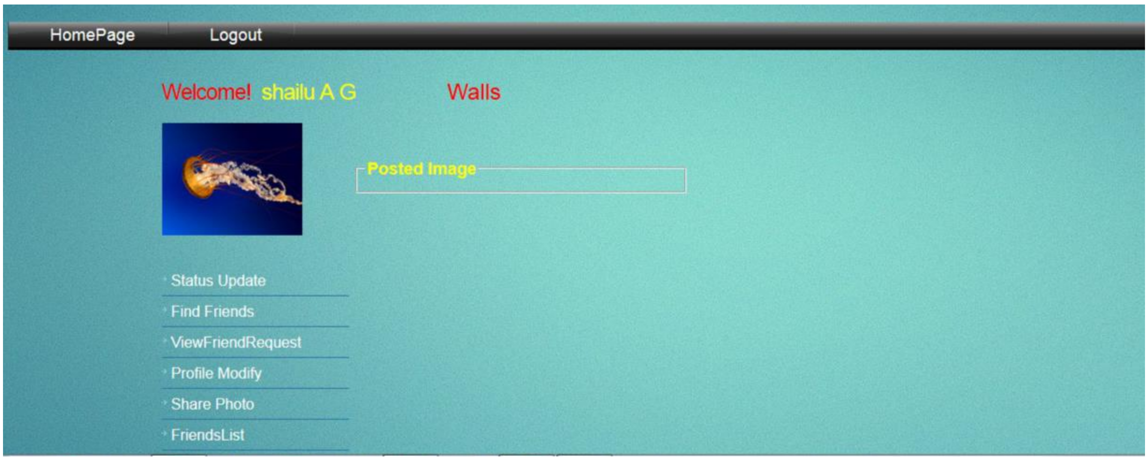
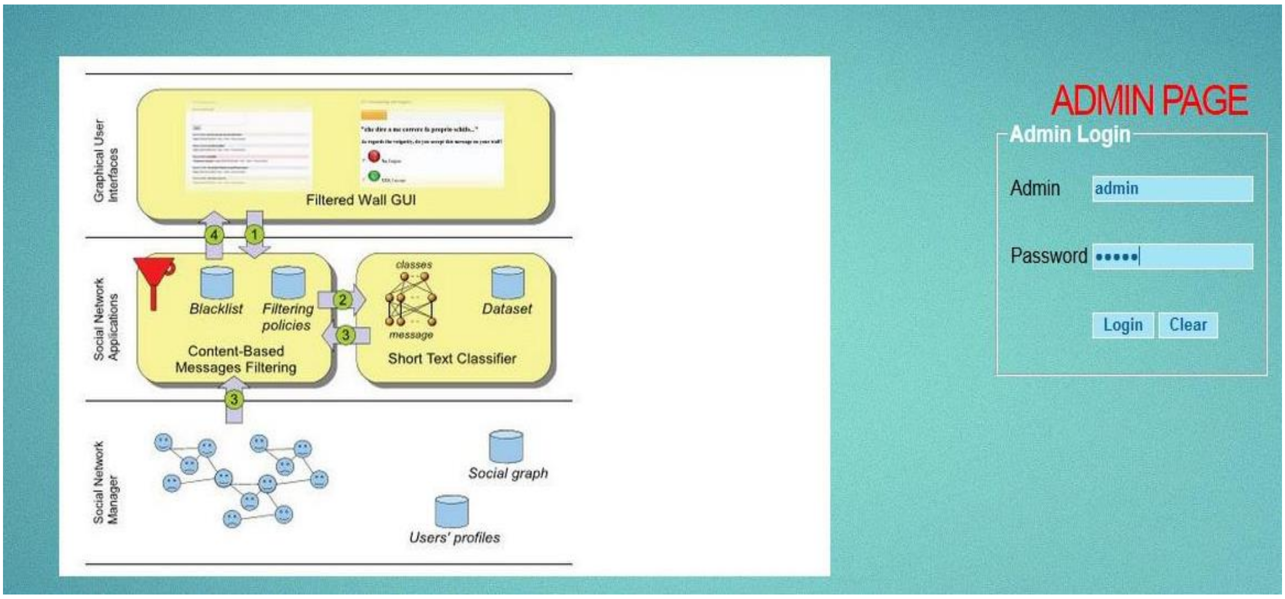


Fig 6.4: User wall



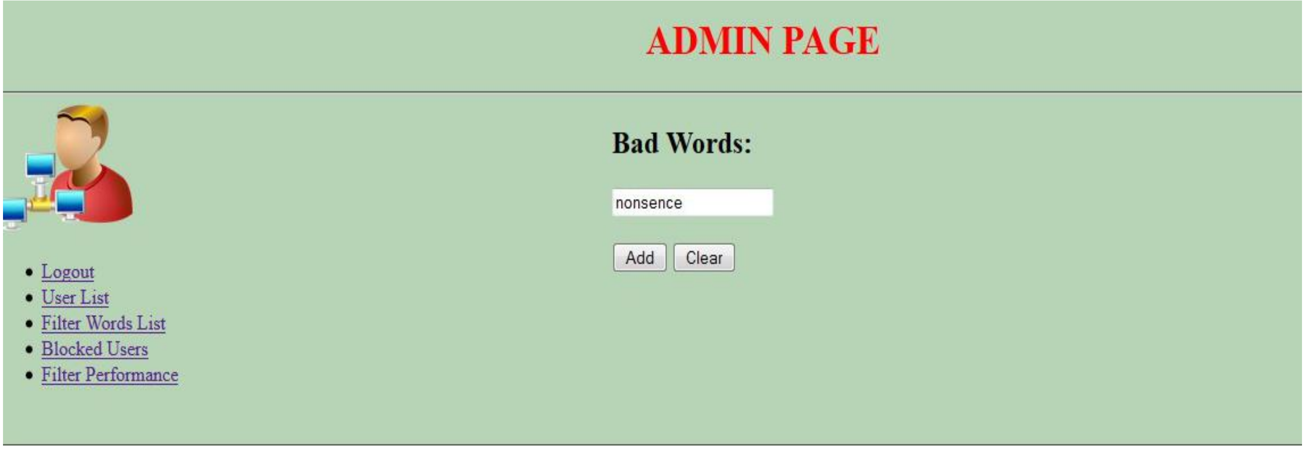
ADMIN PAGE

Admin Login

Admin: admin

Password: •••••

Fig 6.5: Admin Login Page

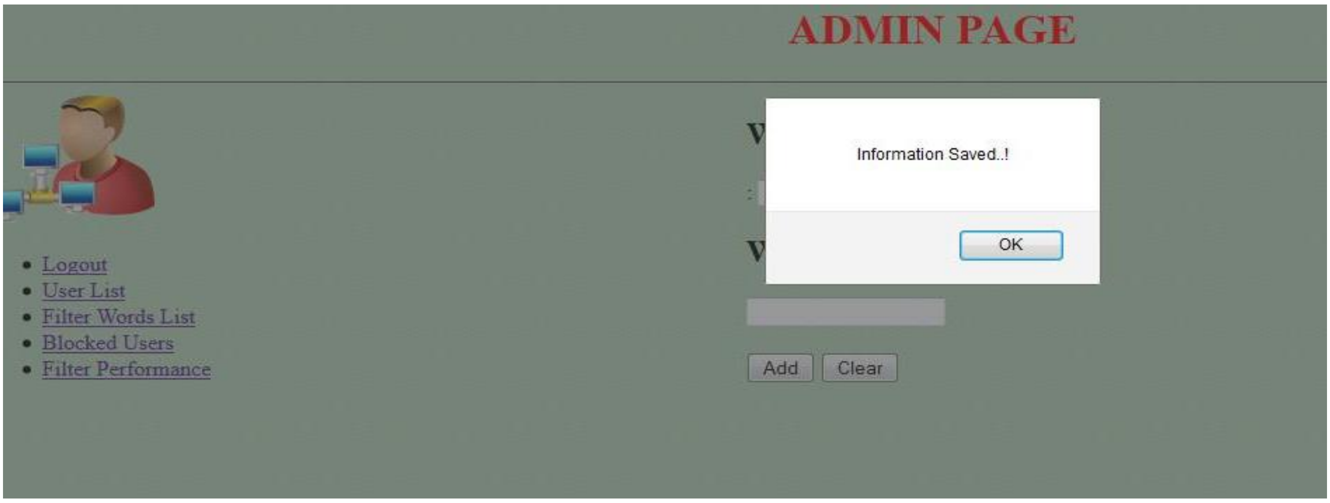

ADMIN PAGE

- Logout
- User List
- Filter Words List
- Blocked Users
- Filter Performance

Bad Words:

nonsence

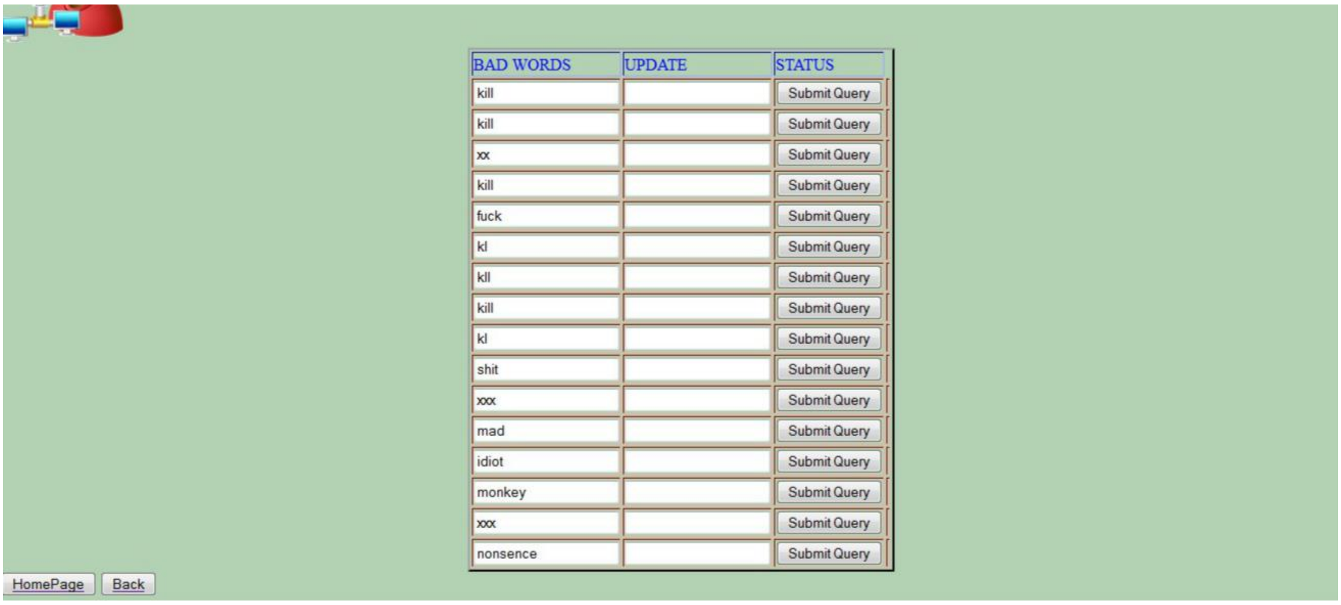
Fig 6.6: Admin Wall


ADMIN PAGE

- Logout
- User List
- Filter Words List
- Blocked Users
- Filter Performance

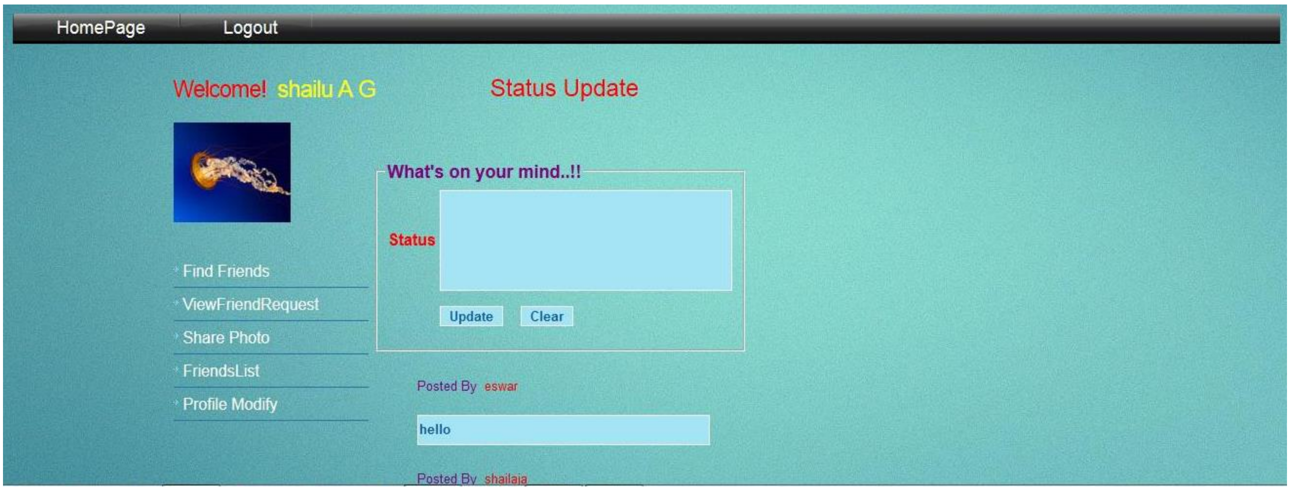
Information Saved..!

Fig 6.7: Bad Word Storing



BAD WORDS	UPDATE	STATUS
kill		Submit Query
kill		Submit Query
xx		Submit Query
kill		Submit Query
fuck		Submit Query
kl		Submit Query
kll		Submit Query
kill		Submit Query
kl		Submit Query
shit		Submit Query
xxx		Submit Query
mad		Submit Query
idiot		Submit Query
monkey		Submit Query
xxx		Submit Query
nonsence		Submit Query

Fig 6.8: Filtered Words List



Welcome! shailu A G

Status Update

What's on your mind..!!

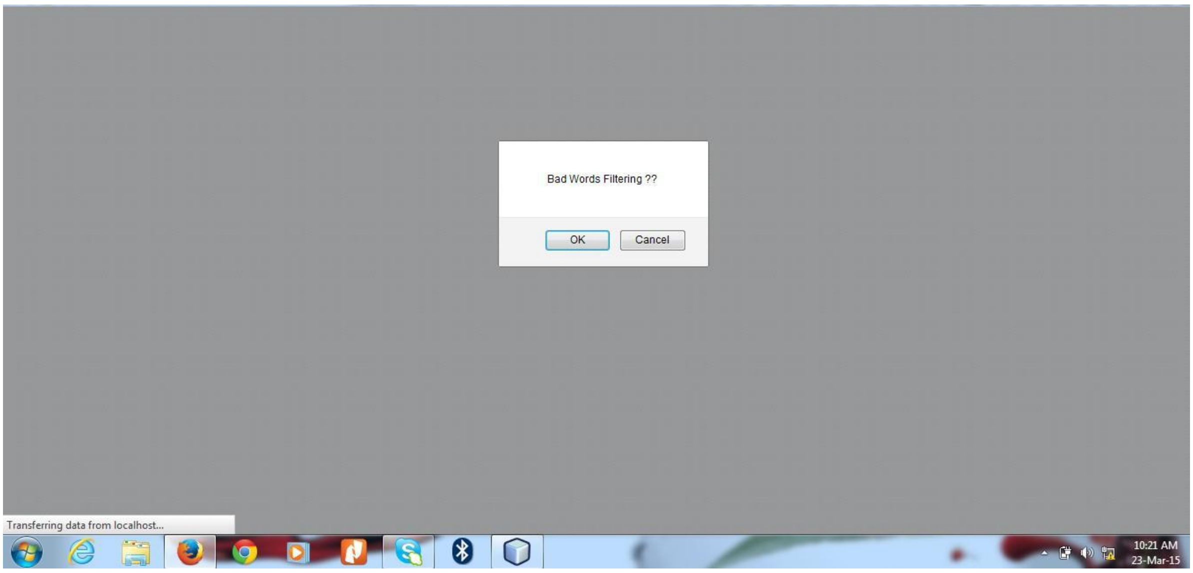
Status

Posted By eswar

hello

Posted By shailaja

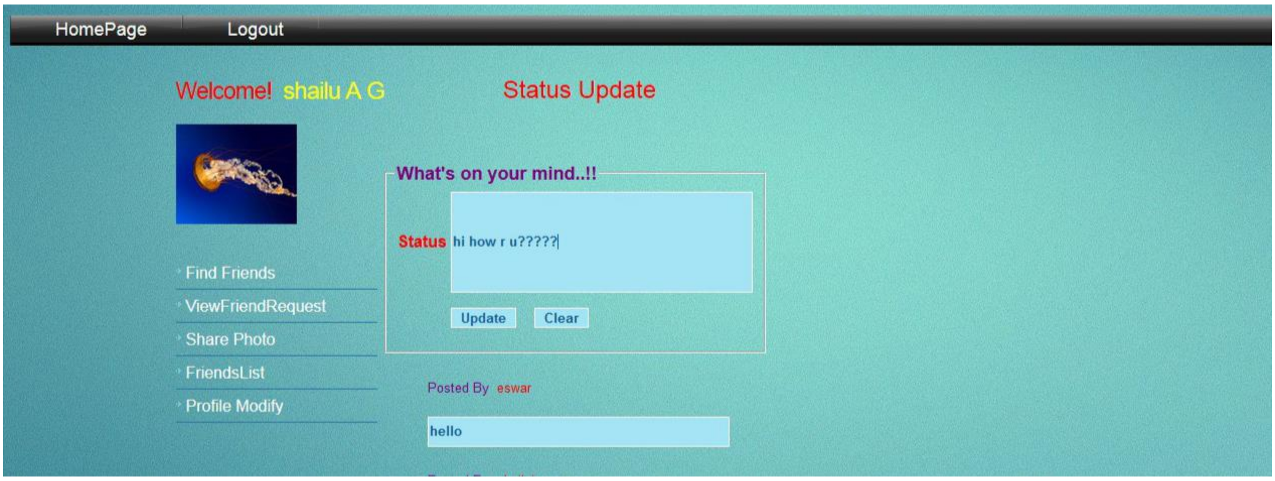
Fig 6.9: Updating Status



Bad Words Filtering ??

Transferring data from localhost...

Fig 6.11: Filtering Bad Word



Welcome! shailu A G

Status Update

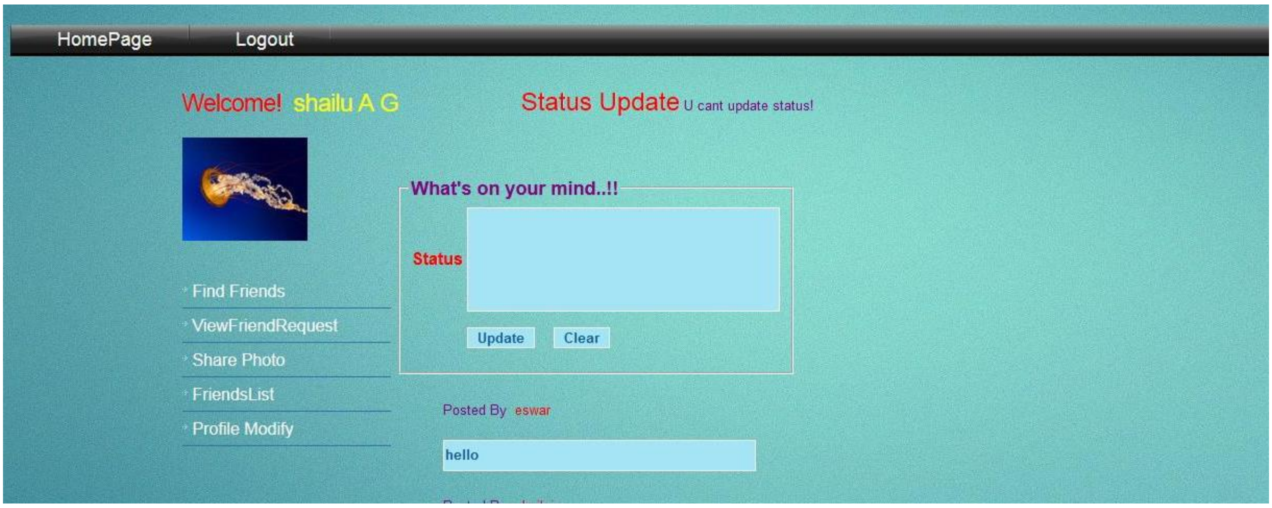
What's on your mind..!!

Status hi how r u?????

Posted By eswar

hello

Fig 6.12: Updating Normal Message after Posting a Bad Word



Welcome! shailu A G

Status Update U cant update status!

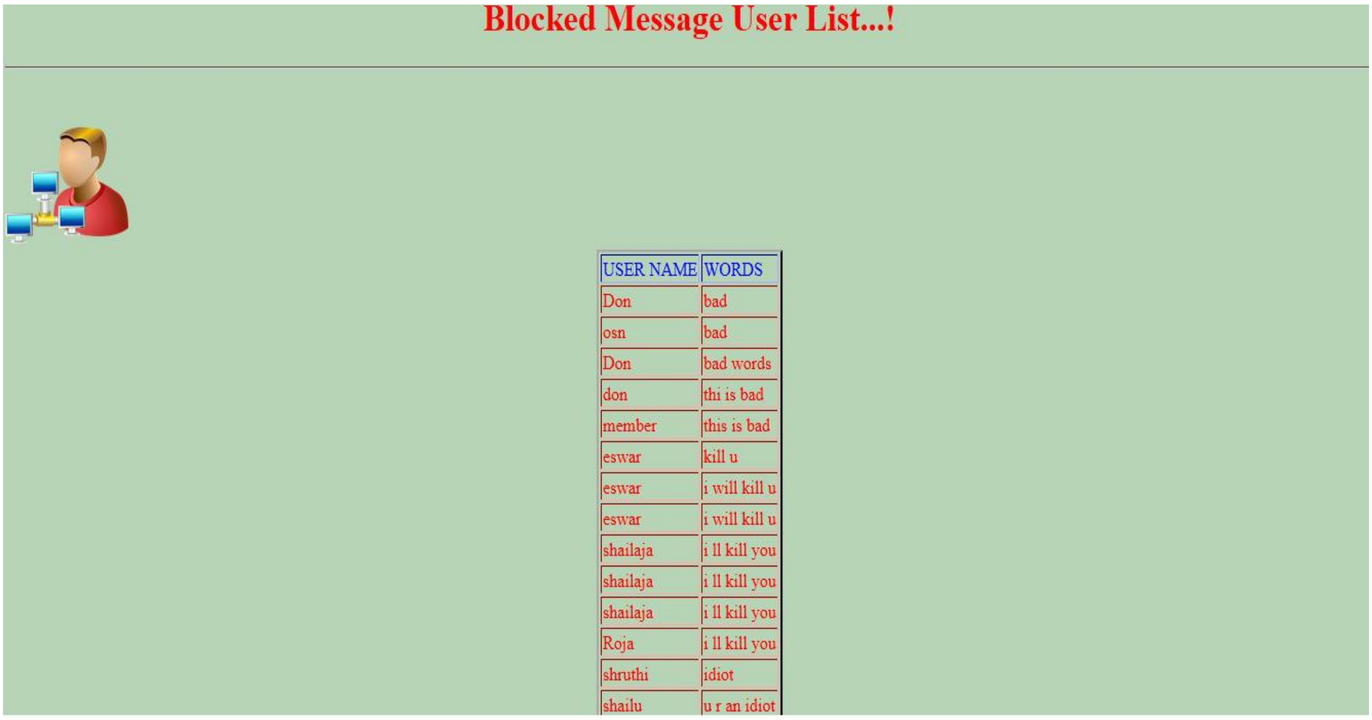
What's on your mind..!!

Status

Posted By eswar

hello

Fig 6.13: Error Message



Blocked Message User List...!

USER NAME	WORDS
Don	bad
osn	bad
Don	bad words
don	thi is bad
member	this is bad
eswar	kill u
eswar	i will kill u
eswar	i will kill u
shailaja	i ll kill you
shailaja	i ll kill you
shailaja	i ll kill you
Roja	i ll kill you
shruthi	idiot
shailu	u r an idiot

Fig 6.14: Blocked user list

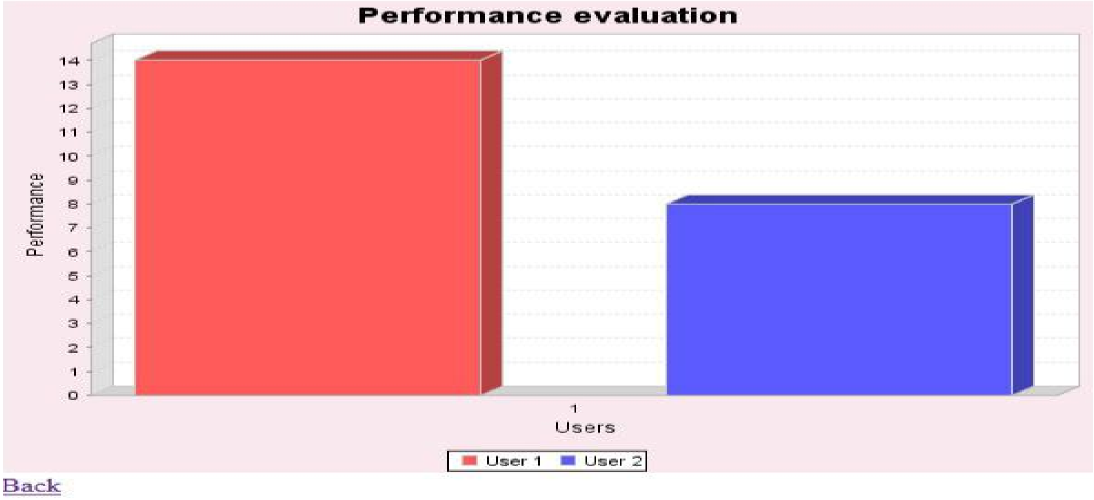


Fig 6.15: Performance graph

IV. CONCLUSION

In this paper, we propose a system for mining social texts and images to filter the unwanted content.Wealso offer a system to filter undesired messages from OSN walls using Naïve Bayes classification which is faster and highly scalable. In addition, the flexibility of the system is improved through the management of Black lists (BLs). In our proposed system we provide text as well as image filtration to filter undesired messages from OSNs wall using customizable filtering rules (FR) enhancing through Black lists (BLs). This work presents a approach that decides when user should be insert words into a black list. The system developed GUI and a set of tools which make BLs and FRs specifications more simple and easy. We have used Naive Bayes Classification algorithm for short text classification and KNN for image classification as KNN is faster and highly scalable

V. REFERENCES

[1] Vanetti, ElisabettaBinaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo Department of Computer and Communication, University of Insubria "a system to filter unwanted messages walls OSN user" IEEE Transactions

on Knowledge And Engineering Flight Data: 25 Year2013.

[2] F. Sebastiani, “Machine learning in automated text categorization,” ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002. J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, “Predicting positive and negative links in online social networks," inProc. 19th Int. Conf. World Wide Web, 2010, pp.641-650.

[3] .BharathSriram, David Fuhry, EnginDemir, Haka*Ferhatosmanoglu “Short Text Classification in Twitter to Improve*Information Filtering”, Computer Science and Engineering Department, Ohio State University, Columbus, OH 43210, USAsriram,fuhry,demir,hakan@cse.ohio-state.edu

[4] AlokChoudhary, ”Towards Online Spam Filtering in Social Networks”, Northwestern University, Evanston, IL, USA, choudhar@ eecs.northwestern.edu

[5] MarcoVanetti et. Al “A *System to Filter Unwanted Messages from OSN User Walls”* University of Insubria, Italy IEEE Transactions On Knowledge And Data Engineering Vol:25 Year 2013

[6] Adomavicius et. Al, “*Toward the next generation of recommender systems: A survey of the state-of-the-art andpossible extensions*,” IEEETransaction on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[7] F. Sebastiani, “*Machine learning in automated text categorization*,” ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.