Recommendation System in E-Commerce using Sentiment Analysis

R. Lydia Priyadharsini[1], M. Lovelin Ponn Felciah[2]

[1] *Research Scholar, Department of Computer Science, Bishop Heber College, Bharathidasan University, TN, India*

[2]*Assistant Professor, Department of Computer Application, Bishop Heber College, Bharathidasan University, TN, India*

Abstract: *Sentiment analysis is one of the current research topics in the field of text mining. Mining sentiment from natural language is a very difficult task. Sentiment analysis provides important information for decision making in various domains. Various sentiment detection techniques are available but may not render most accurate results in all perspective. In this paper, sentiment of users regarding the services provided by E-shopping websites is considered. The opinion or sentiment of the people is implied by reviews, ratings and emoticons. The products which have got positive feedback from the previous users are recommended for the current users. For this stochastic learning algorithm which analyzes various feedbacks related to the services is used. The opinion is classified as negative, positive and neutral. Analysis takes place based on classification. The proposed system will find out fake reviews about a product with the help of MAC address along with review posting patterns. User will login to the system by giving the username and password and will view various products and can give review about the product. To find out the review is fake or genuine, the MAC address of the system which is unique is checked by the proposed system. Hence fake reviews are not accepted.*

Keywords:
Text Mining, Sentiment Analysis, Ratings, Reviews, Emoticons, Recommend, Fake reviews

I. INTRODUCTION

Smart phones, Laptops and internet have made online shopping very easy. As e-commerce is developing rapidly these years, online shopping has become more and more popular because of its variety, cheap price (high discounts) and fast logistic systems. More people intend to do online shopping these days. Meanwhile the quality and delivery of products is uneven, thus users' comments become the important information to judge the product's quality.At the same time the, product manufacturers can obtain the opinion of the users to improve the products. Confronting to the massive data in the websites, analyzing and concluding the information manually is very tedious. So how to extract useful information and build objective products' quality test system automatically to deal with the massive textual information is

emerging in the related research field. Opinion Mining is a technology based on text mining and natural language processing. It provides the approach to cope up with the problem .So generating summary of the products has been attracting many researchers during these years. Emotional orientation of each review is focused with Document-level

sentiment analysis. It recognizes the opinion of the contents which authors express, mainly discusses the sentence-level opinion mining and treats the statements of the product features for each viewpoint as analysis objects, then the authors' opinion inclinations is determined. Therefore

sentence-level sentiment analysis is the main task on opinion mining. The approach can find the specific details of the comments and has a high confidential degree, but the operation is very complex. For example, consider a laptop, its feature can be classified into performance, price, appearance, endurance time, brand and so on. Each feature or attribute is considered for which the user can specify the comment and a comprehensive evaluation in order to avoid the overgeneralization is done. Feature-specific opinion mining attracts much attention. An object is an entity. It can be a product, person, event, organization, topic or something else. It is associated with a hierarchy or taxonomy of components or a set of attributes. Meanwhile, each component also can have its own set of subcomponents or attributes. A feature is defined to show both components and attributes and it is the subject of a review. In fact, people obey the grammatical rules to organize sentences while writing articles. But while writing reviews informally, people usually neglect it and there is more chance to have grammatical mistakes as well as usage of sarcastic words in reviews. This phenomenon is especially prominent when people give away their comments after online shopping. Sentiment analysis also known as opinion mining refers to the use of natural language processing and computational linguistics to extract subjective information from the given data and classify opinions. It is a broader concept and many tasks are involved in it. The most important are as follows:

A. Sentiment Classification- The process is also known as sentiment orientation or sentiment polarity. The opinion is classified into any one of the following categories- positive, negative or neutral. There are three levels of classification.

1) Document level- It considers the entire document as an opinion and classifies accordingly. With this classification, it is possible to predict whether the review expresses a positive or negative opinion.

2) Sentence level- If the given sentence is subjective, this level of classification classifies it as positive or negative. Hence, the process is more related to the task of subjectivity classification that distinguishes sentences that express factual information from sentences that express subjective opinions.

3) Aspect level- The sentiment is classified with respect to specific aspects of the entities. Therefore, this is the finest-grained and the most complex level as it is essential to extract many features and their relationships with high precision.

B. Subjectivity Classification- It consists of evaluating whether a given sentence is subjective or not. It can be considered as the step that precedes sentiment classification. The accuracy of sentiment classification can be improved by employing a better subjectivity classification.

C. Opinion summarization- It is focused on extracting the main features of an entity shared with one or several documents and the sentiments regarding them. The task involves single document as well as multi-document summarization. Single document summarization analyzes the changes in the sentiment orientation throughout the document. In multi-document summarization, once the features have been detected, the system must group the different sentences which express sentiments related to those features.

D. Opinion retrieval- The documents that express opinions or views are retrieved based on the given query. In this task, the documents are ranked as per the relevance score and the opinion score with respect to the query. A basic e-commerce reviews analysis is shown in fig 1.

Fig 1:Reviews in E-Commerce Website

II. RELATED WORK

Michael Jahrer, eta. al,… [1] proposed recommender systems which help users to discover items within large web shops, to navigate through portals or to find friends with similar interests. The most interesting applications for recommender systems have thousands of users which generate huge amounts of data. For example, online shops collect purchase data and provide each user with a personalized shopping page on the login. The sources of information used for the recommender system can be widespread. Users generate events like the purchase of a product, rating a product, creating a bookmark or clicking on a specific item. Independently of the area of application or the type of information used, it is a major goal to increase the accuracy while retaining the capability of being able to use big datasets. Generating more accurate predictions is of general interest. For a subscription service like Netix, good recommendations are a key to customer loyalty. In the case of online stores better recommendations directly increase the revenue. In this article, a systematic empirical analysis of different blending methods on the Netix dataset is provided. The Netix dataset is one of the largest available benchmark datasets for collaborative filtering algorithms today. It contains about 108 ratings, collected in a time period of 7 years.

Yu Zhang,et.al,…[2] implemented Collaborative filtering (CF) which is an effective recommendation approach based on the intuitive idea that the preference of a user can be predicted by exploiting the information about other users which share similar interests. In particular, CF techniques exploit past activities of the users, such as their transaction history or product satisfaction expressed in ratings, to predict the future activities of the users. In recent years, CF-based recommendation systems have become increasingly popular because it is generally much easier to collect the past activities of users than their profiles, partially due to privacy considerations. Collaborative filtering is an effective recommendation approach in which the preference of a user on an item is predicted based on the preferences of other users with similar interests. A big challenge in using collaborative filtering methods is the data sparsity problem which often arises because each user typically only rates very few items and hence the rating matrix is extremely sparse. In this paper, the problem is addressed by considering multiple collaborative filtering tasks in different domains simultaneously and exploiting the relationships between domains. It is referred to as a multi-domain collaborative filtering (MCF) problem. To solve the MCF problem, a probabilistic framework is proposed which uses probabilistic matrix factorization to model the rating problem in each domain and allows the knowledge to be adaptively transferred across different domains by automatically learning the correlation between domains. Also the link function is

introduced for different domains to correct their biases.

Raghunandan H. Keshavan,et.al,…[3]studied a low complexity algorithm, based on a combination of spectral techniques and manifold optimization, called as OPTSPACE. Performance guarantees that are order-optimal in a number of circumstances is proved. Collaborative filtering was studied from a graphical models perspective which introduced an approach to prediction based on Restricted Boltzmann Machines (RBM). Exact learning of the model parameters is intractable for such models, but the authors studied the performances of a contrastive divergence, which computes an approximate gradient of the likelihood function, and uses it to optimize the likelihood locally. Based on empirical evidence, it was argued that RBM's have several advantages over spectral methods for collaborative filtering. An objective function analogous to the one used in the present paper was considered early on in Srebro and Jaakkola, which uses gradient descent in the factors to minimize a weighted sum of square residuals. Salakhutdinov and Mnih justified the use of such an objective function by deriving it as the (negative) log-posterior of an appropriate probabilistic model. This approach naturally leads to the use of quadratic regularization in the factors. Again, gradient descent in the factors was used to perform the optimization. Also, the paper introduced a logistic mapping between the low-rank matrix and the recorded ratings.

Morgan Harvey,et.al,…[4] implemented content filtering systems based on techniques from information retrieval, and designed to assist in the process by narrowing down the number of items a user has to look through in order to fulfill a particular information need. These systems rely on textual descriptions of items and seek to match these descriptions with a user's profile in order to suggest useful items. One significant issue with this content-based filtering is that for some types of items it can be extremely difficult to choose suitable descriptive terms to search for. Another, more accurate, approach to discovering items of interest is provided by ratings-based collaborative filtering systems, which use past ratings to predict items the user may like. Such systems predict which items a given user will be interested in based on the information provided in their user profile. These profiles consist of votes or ratings for items in the system that the user has already viewed and evaluated. The profiles of other users are frequently also exploited to improve predictions for the target user. Profiles are generally constructed explicitly from user ratings, however they may also be compiled implicitly by considering a user's purchase or bookmark history. Explicit ratings systems are commonly found on movie and music recommendation sites such as Movie Lens or imdb where users can give each item a rating from 0 to 5 stars. Zero indicates that the user strongly dislikes the item and five indicates that they really like the item,

however any discrete set of values could be used. Implicit systems can also be used, for example in online retail stores such as Amazon where users purchase items or add them to a "wish list", indicating that they are interested in that kind of item.

HaoMa,et.al,…[5] provided the process of trust generation is a unilateral action that does not require user to confirm the relationship. This also indicates that user does not need to even know user in the real life. "Social friendships" refer to the cooperative and mutual relationships that surround us, such as classmates, colleagues, or relatives, etc. Lots of social networking Web sites, like Facebook and Orkut, are designed for online users to interact and connect with their friends in the real life. From the definition, it is clear that see that trust-aware recommender systems cannot represent the concept of "social recommendation", since the idea of "social recommendation" anticipates to improve recommender systems by incorporating a social friend network. Secondly, trust-aware recommender systems are based on the assumption that users have similar tastes with other users they trust. This hypothesis may not always be true in social recommender systems since the tastes of one user's friends may vary significantly. Some friends may share similar favors with this user while other friends may have totally different tastes. Hence, trust-aware recommendation algorithms cannot be directly applied to generate recommendations in social recommender systems. Thirdly, due to the rapid growth of Web 2.0 applications, online users spend more and more time on social network related applications since interacting with real friends is the most attractive activity on the Web. On the contrary, only few online systems, like Epinions, have implementations of trust mechanism. Thus, in order to provide more proactive and personalized recommendation results to online users, more attention has to be given to the research of social recommendation, in addition to the existing research of trust aware recommendation.

III. EXISTING METHODOLOGIES

Recommender Systems are indispensable to provide personalized services on the Web. Recommending items which match a user's preference has been researched for a long time, and there exist a lot of useful approaches.

A. *Collaborative filtering with explicit feedbacks:*

First, the existing Collaborative Filtering method with explicit feedbacks is discussed. Collaborative Filtering with explicit feedbacks that both positive and negative feedbacks are observed in the dataset. The Collaborative Filtering methods can be divided into the memory-based method, the model based method and the combination of the two. The memory-based method includes the Neighborhood method, which calculates the similarity of the users or items. The model-based method includes the Matrix

Factorization model, the Probabilistic model and Cluster based model. This approach is based on the idea that there are latent factors which represent the user-item preference relationships between users and items, and unknown preferences can be predicted using latent factors and the relationship between users-latent factors and items-latent factors. The biggest problem in Collaborative Filtering is the sparseness of observed values. It means feedbacks are observed in very small portion of all possible user-item pairs. However the Matrix Factorization model is known to work better than other models even if the data is sparse.

B. *Collaborative filtering with implicit feedbacks:*

Here the existing Collaborative Filtering method with implicit feedbacks is discussed. Basically, a dataset with implicit feedbacks consists of user-item pairs where the user provided feedbacks to the item. Often timestamps are also provided. Existing works for Collaborative Filtering with implicit feedbacks assume that implicit feedbacks are observed as one-class positive feedbacks and missing values do not indicate the negative feedbacks. Thus, existing Collaborative Filtering methods with explicit feedbacks cannot be directly applied to the dataset with implicit feedbacks because they require both positive and negative feedbacks in the dataset. To address this problem, many existing approaches try to find possible negative values hidden in missing values. And employed a weighted Matrix Factorization model. They initially filled all missing values with negative values, and assign weights to discount the relative contribution of each value to prediction. The weights are determined based on the number of items to which a user provided feedback, or the number of users who gave feedbacks for an item, or given uniformly. This approach has the problem of running time. Basically, the Matrix Factorization model can save computation for missing values, but the weighted Matrix Factorization model must cope with all elements even when the original dataset is given very sparse and also proposed a sampling based method. This approach samples only a part of the missing values and replaces them with negative values. Three kinds of sampling methods are proposed: User-oriented sampling, Item-oriented sampling and Uniform sampling. User-oriented sampling assumes that the number of negative values hidden in missing values and the amount of past feedbacks given by a user are related. Item-oriented sampling assumes that the number of negative values hidden in missing values and the amount of past feedbacks given for an item are related. In this work, this method is combined with our proposals. Also, proposed an extension of the weighted Matrix Factorization model. They incorporate similarity matrices over items and users to the weighted Matrix Factorization model. Finally proposed the joint model of the Matrix Factorization model and non-negative

Matrix Factorization model which classify missing values into positive and negative feedbacks. Packet and Koenigstein proposed a bayesian generative model which predicts the probability with which missing values are converted to negative feedbacks. There exist other approaches that use the auxiliary data to treat missing values. Limitations of existing approaches as follows:

- Only analyzes ratings from user reviews
- Fake reviews can't be analyzed by existing work
- User can't be identify genuine reviews
- Handle only limited number of product reviews

IV. PROPOSED FRAMEWORK

A recommendation system has been implemented based on hybrid approach of stochastic learning and context based engine. We have tried to combine the existing algorithms for recommendation to come up with a hybrid one. It improves the performance by overcoming the drawbacks of traditional recommendation systems. Recommender systems being a part of information filtering system are used to forecast the bias or ratings the user tends to give for an item. Among different kinds of recommendation approaches, collaborative filtering technique has a very high popularity because of their effectiveness. These traditional collaborative filtering systems can even work very effectively and can produce standard recommendations, even for wide ranging problems. For item based on their neighbor's preferences entropy based technique creates better suggestions than others. Whereas other techniques like content based suffers from poor accuracy, scalability, data sparsity and big-error prediction. To find these possibilities user-based collaborative filtering approach is used. In this Item based collaborative filtering technique , first the User item rating matrix is examined and the relationships among various items is identified. Then, these relationships is used in order to compute the recommendations for the user. Then using cosine similarity which is a similarity weight is going to play an important role in the collaborative item based filtering approach and hence in order to maintain or select the trustable users from the given set of user. Hence they give us a method to increase or decrease the significance of a particular user or item. In the present methodology adjusted similarity for computation of similar weights of items is used.

Sentiment analysis of natural language texts is a large and developing field. Sentiment analysis or Opinion Mining is the computational treatment of opinions and subjectivity of text. Sentiment analysis is an Information Extraction task that intends to acquire writer's feelings expressed in positive or negative comments, after analyzing his documents. The term 'Presence' is more important to sentiment analysis then term 'Frequency' which was earlier used for

traditional information retrieval. It has also been reported that unigrams surpass bigrams for classifying movie reviews by sentiment polarity. In the proposed framework stochastic learning algorithm is implemented to analyze reviews, ratings, and emoticons. The proposed work is described as follows:

A. *Framework construction:*

E-commerce framework is used to buy the products in online to easy retrieval the mobile products. This module is used to create android and web site for recommending best mobiles in specific area. Admin is the responsibility for maintaining the all details in server and server can be design in server. There are two accounts such as admin and user account. Admin can login to the system and post item details with expiry dates. User can login to the mobile to choose the language and area. Then view the products with specified filter. This module is used to create web site buy or post products for users. Admin can login to the system and post products with features. User can login to the system to view product details.

B. *Reviews collection:*

Admin collect reviews and have various types of reviews. Reviews may be rating reviews, text reviews and smiley reviews. All reviews are stored in database for future evaluation. Ratings, reviews and emoticons are stored in database. Rating, Reviews and Emoticons are the evaluation or assessment of something, in terms of quality (as with a critic rating a novel), quantity or some combination of both.

C. *Sentiment analysis:*

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and ratings for applications that range from marketing to customer service to buy the products efficiently. Admin can analyze whether the product is positive or negative. In star rating, star count values is calculated. In text reviews, extract keywords and matched with database. Then smiley reviews are calculated based positive and negative symbols.

D. *Recommendation system:*

Recommender systems are a subclass of information filtering system that seek to predict the "rating" or "preference" that a user would give to an item. User can search the product in search bar. And view the list of products based on price and review details. Implement the stochastic learning algorithm to classify the products such as positive or negative. Positive products are display in recommendation panel based on ratings and reviews. If the product has

negative review means, automatically the positive products in recommendation panel.

E. *Fake reviews monitoring:*

In this module, fake reviews are analyzed by admin. A media access control address (MAC address) of a computer is a unique identifier assigned to network interfaces for communications at the data link layer of a network segment. Admin can get user account details, Mobile address and Order id details. So one user can post a review that will be a genuine one. The proposed framework is shown in fig 2.
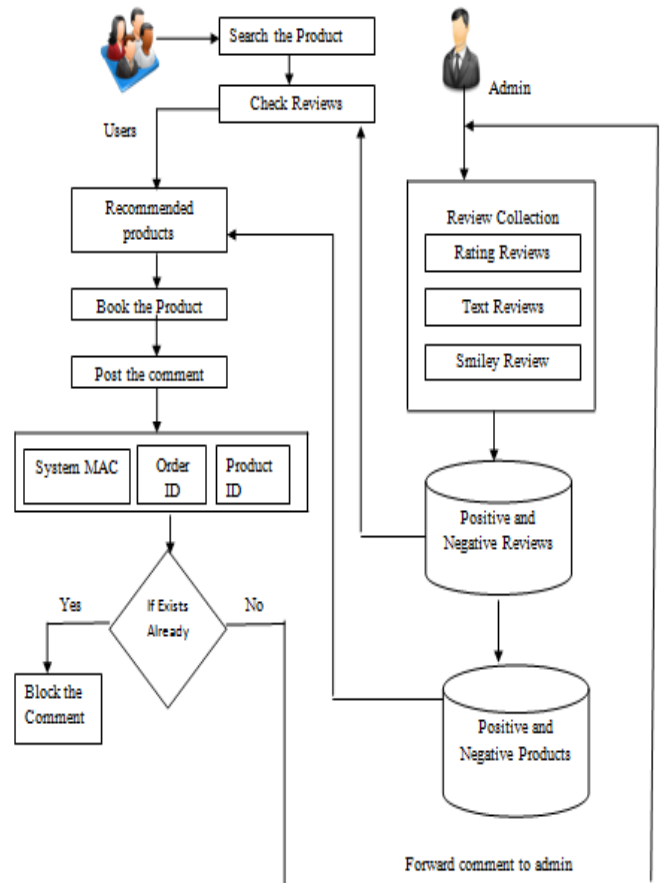

Fig 2 Proposed Framework

V. EXPERIMENTAL RESULTS

A website for booking mobile phones is created using C#.NET framework as front end design and SQL SERVER as back end.
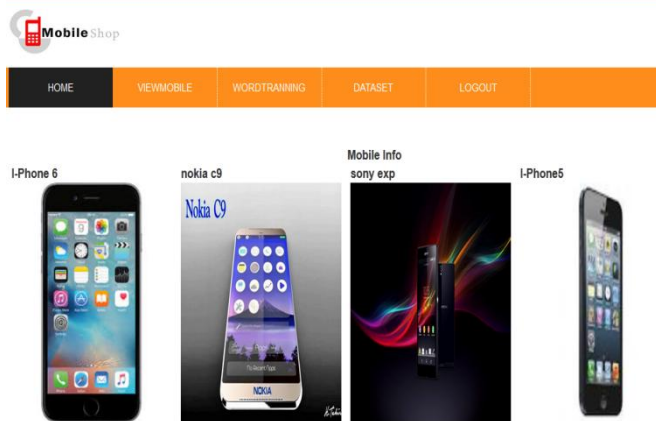
Fig 3. Home Page

There are a number of evaluation matrices used to evaluate the retrieval performance. For matric evaluation, precision and recall is used. Where, Precision measures the availability of relevant products from the product database based on feedbacks and Recall measure the availability of relevant products from the overall database

$$Precision = \frac{Number\ of\ Relevant\ Products\ extracted}{Total\ Number\ of\ Products\ extracted}$$

$$Recall = \frac{Number\ of\ Relevant\ products\ extracted}{Total\ Number\ of\ Products\ in\ Database}$$

To analyse the recommendation system, various types of recommendation algorithms are used.Some products in e-commerce were taken based on sentiments. Then average on precision and average on recall is calculated for every sentiment. Result shows that proposed framework provided the better result in comparison of framework. The performance chart is shown in Fig 4.
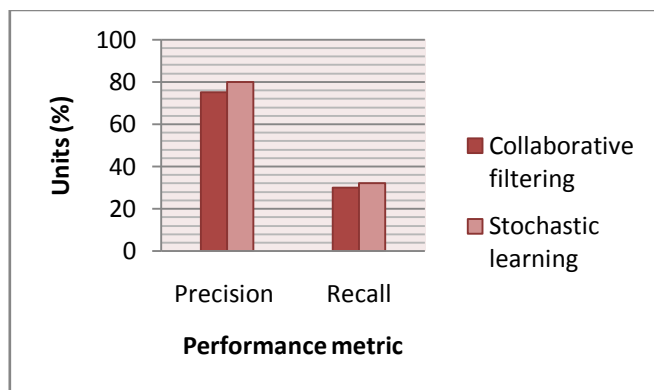

Fig 4: Performance graph

VI. CONCLUSION

In this proposed work, a novel implementation of a product recommendation system based on hybrid recommendation algorithm is presented. The main advantage of this method is the visual organization of the data based on the underlying structure, and a significant reduction in the size of the search space per result output. Also the user can easily search the products anywhere at any time. Ratings, reviews and emoticons are analyzed and categorized as positive and negative sentiments. The products can be searched with the help of review based filtering. MAC based filtering approach can be used to avoid fake reviews. This method was evaluated against real-time user data collected through an online website, by using a subset of the products liked by each user as input to the system. The current results are notably better than random approach. However, it is felt that with a better dataset and a number of improvements this method may achieve better results. Hybrid Recommendation is one of the main modules of the system which helps to overcome the drawbacks of the traditional Collaborative and Content Based Recommendations. Thus promising results are obtained using this model.

REFERENCES

[1] M. Jahrer, A. Toscher, and R. Legenstein, "Combining predictions for accurate recommender systems," in KDD'10, pp. 693-702, 2010.
[2] Y. Zhang, B. Cao, and D. Y. Yeung, "Multi-domain collaborative filtering," in Proc.UAI, pp. 725-732, 2010.
[3] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," Journal of Machine Learning Research, vol. 11, pp. 2057-2078, 2010.
[4] M. Harvey, M. Carman, I. Ruthven, and F. Crestani, "Bayesian latent variable models for collaborative item rating prediction," in CIKM'11, pp. 699-708, 2011.
[5] H. Ma, D. Zhou, C. Liu, M. Lyu, and I. King, "Recommender Systems with Social Regularization," in ACM WSDM, pp. 287-296, 2011.
[6] P. Bedi, H. Kaur, and S. Marwaha, "Trust based recommender system for semantic web," in IJCAI'07, pp. 2677-2682, 2007.
[7] S. Scellato, A. Noulas, and C. Mascolo, "Exploiting Place Features in Link Prediction on Location-based Social Networks," in KDD'11, pp. 1046-1054, 2011.
[8] Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, "Joint Social and Content Recommendation for User-Generated Videos in Online Social Network," IEEE Trans. Multimedia, vol. 15, no. 3, pp. 698-709, 2013.
[9] Y. Chen, A. Cheng, and W. H. Hsu, "Travel Recommendation by Mining People Attributes and Travel Group Types from Community-Contributed Photos," IEEE Trans. Multimedia, vol. 15, no. 6, pp. 1283-1295, 2013.
[10] M. Lovelin Ponn Felciah and R. Anbuselvi , "Smartphone Product Review Sentiment Analysis using Logistic Regression" in I J C T A, 9(26) 2016, pp. 343-349
[11] M. Lovelin Ponn Felciah and R. Anbuselvi , "A comparative study of machine learning algorithms on opinion classification" International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.85 ,2015
[12] Vaibhavi.N , Patodkar,Shaikh I.R. " Sentimental Analysis on Twitter Data using Naïve Bayes", Sixth Post Graduate Conference for Computer Engineering(cPGCON 2017) Procedia International Journal on Emerging Trends in Technology(IJETT).
[13] Payal S Gadhave ,Mahesh D Nirmal,"Twitter Based Sentiment Analysis using Hadoop", Sixth Post Graduate Conference for Computer Engineering (cPGCON 2017) Procedia International Journal on Emerging Trends in Technology(IJETT).