

FAKE REVIEW AVOIDANCE IN ONLINE REVIEW SHARING USING SENTIMENTAL ANALYSIS

Sabitha.V¹, Preethi.S.P², Revathi.G³, Vinothini.R⁴

¹Assistant Professor, Dept of IT, Jeppiaar SRR Engineering College, Tamilnadu, India.

^{2,3,4} Dept of IT, Jeppiaar SRR Engineering College, Tamilnadu, India.

Abstract - The reviews shared on the comment section are used to a fullest extent to buy any product on the web shopper but one can hardly read all the reviews to evaluate the overall opinion on that particular item. Thus a summarization of all the reviews would be attractive which gives overall picture of the item that the customer is searching for. The current strategy allows the fake reviews to enter into the comment section which affects both customer and web shopper equally. The suggested unsupervised strategy performs comprehensively by avoiding the entry of fake reviews into the comment section.

Key Words: Online shoppers, Fake reviews avoidance, summarization, unsupervised technique, graph, admin, user interface, bank transaction, key generation, review sharing, graph generation, rating.

1. INTRODUCTION

Decision making is one of the essential quality humanity needs. We ask for suggestion from family and friends before any purchase. Such recommendations really have strong influence on the purchase decisions taken. To select the best among the numerous products is a critical task. In online shoppers the consumers are always greatly influenced by the Electronic Word of Mouth (eWOM), the opinion given by the other users. Therefore the recommendation and opinion given by the consumers are very valuable for buyer to judge a product and for a well informed selection. All the online shoppers Amazon, Flipkart and many others strongly depend on this eWOM to sell their products and acquire a wealth of opinion from the customer side. These shared reviews act as mine of information that helps the vendor to develop their business strategy. As these feedback influences the business greatly it is very important to avoid fake reviews getting into the review section. The challenging part is to effectively handle the thousands of reviews given in comment section. Another task is to understand what the consumer express about the product in the comment section.

2. RELATED WORK

In existing, we are able to give the fake reviews on a particular product which we have not even brought. It will affect the product rating as it is not about sharing a true experience. Here sentimental Analysis doesn't give the

correct information to the next user who intent to buy that item. It confuses the formation of the information and changes the meaning of the comments given. This setback gives entirely wrong opinion to the customers and also affects product rating. The next big setback is that the buyer feels monotone to read the long list of reviews given by the previous consumers. Therefore there is a need of a condensed content that gives accurate overall feedback about the product. In existing, they use Breakdown algorithm in which it is impossible to limit the reviews given by a user as it does not identify a user uniquely and so anyone can comment on a product without buying it.

3. DISADVANTAGES OF EXISTING WORK

- ✓ The fake reviews enter into the comment section easily. Thus the consumer gets misled.
- ✓ Any third party can comment on any products without experimenting it and thus it does not ensure any security.
- ✓ Therefore this affects both customer and sales person.
- ✓ Does not help to improve sales.

4. PROPOSED WORK

In proposed, we are not allowed to give fake reviews on a particular product. A key is generated each time a user buys a product which is unique for each user. This unique key is used by the user to share his view on the product bought thus avoiding fake reviews. This unique key generation is done using random-number generator that generates a sequence of number that cannot be logically predicted rather than by random chances. These product reviews are considered by the online shoppers which enables them to make use of these opinions and improve sales. Here we propose Clustering Sentiment algorithm (CSA) is a dual prediction (DP) strategy in which likelihoods of the shared reviews and stored training data set are checked and predicted as positive, negative or neutral. The one-to-one correspondence of reviews shared and trained data are checked which is known as dual training (DT). The sentimental analysis gives the view whether the product is positive, negative or neutral. Overall performance is evaluated in the form of a graphical representation.

4.1 ADMIN MAINTAIN PRODUCTS

It is used to authenticate the authorized admin. This module will check whether the username and password is correct or not (valid admin id and valid password). So we are preventing unauthorized product owner entering into the login window to admin window. Here product owner updates, deletes and modulate their products.

➤ Admin Maintain the Products

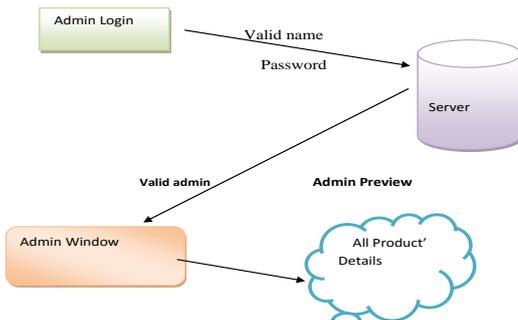


Fig.4.1.1.Admin login

4.2 USER INTERFACE

This interface will check whether the username and password is correct or not (valid user id and valid password). Here we validate the login user and server authentication. This prevents the unauthorized user entering the user window. If we enter any invalid username or password we can't enter into the user window and it shows an error message.

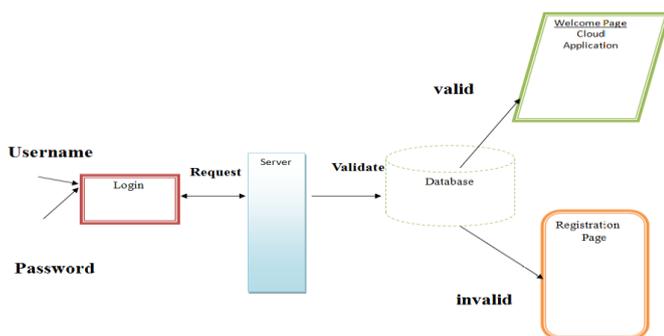


Fig.4.2.1. User login

4.3 USER TRANSACTION

Here user will transfer the amount to provider (web shopper). The user buys the product by using online transaction. The amount for that particular product will be transferred from user account to product admin account.

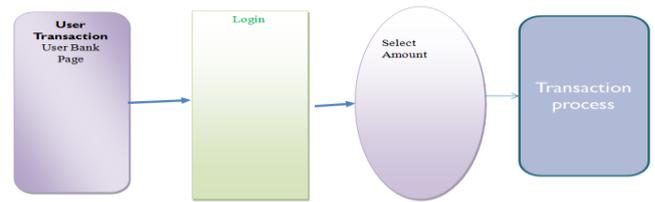


Fig.4.3.1.Bank transaction

4.4 KEY GENERATION

This helps the buyer to share their Opinions and provides security. A random-number generator (RNG) is used to compute a sequence of numbers that cannot be reasonably predicted better than by a random chance. Key Generation is the process for generating keys to our files. That key will have to be a unique for every member at the time of each receives. This unique key is used to share reviews in the web shoppers.

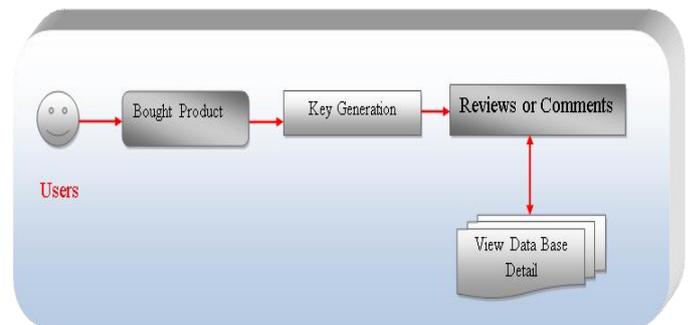


Fig.4.4.1.Key generation

4.5 DUAL SENTIMENTAL ANALYSIS

Here all the comments given are split into individual elements. Then the one-to-one correspondence between the individual elements and database elements are checked. This process is called dual training (DT). In DT, the likelihoods of the individual elements and stored training data set are checked.

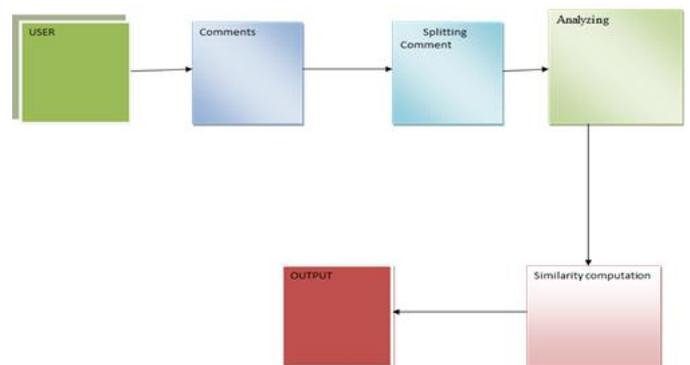


Fig.4.5.1.Dual sentimental analysis

4.6 PERFORMANCE EVALUATION

The evaluation of the entire reviews given by the user is done here. One can scarcely read all surveys thus overall performance evaluation is done by generating a graph. The overall rating is also specified for the benefit of user and producer. It gives the view whether the product is positive, negative or neutral.

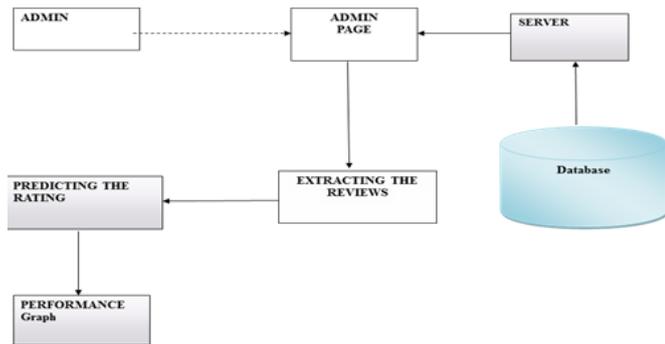


Fig.4.6.1.Performance evaluation

5. ADVANTAGES OF PROPOSED WORK

- ✓ Avoiding fake reviews.
- ✓ Maintain the security for each review shared.
- ✓ Showing survey through graph of particular product
- ✓ Improve The sales of the web shoppers by providing right information.

6. PROPOSED DIAGRAM

The proposed model block diagram is represented as follows:

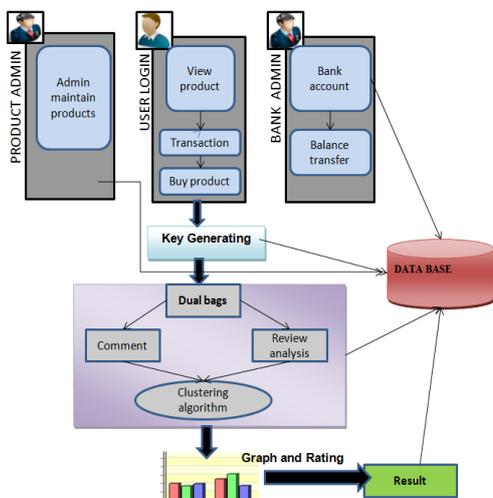


Figure6.1: Architecture Diagram

7.ALGORITHM

7.1. CLUSTERING ALGORITHM

clustering is the classification of objects into different groups according to their similarities and dissimilarities. The partitioning of a dataset into subsets, so that the data in each subset share some common characteristics often according to some defined distance measured.

7.2. K- MEANS ALGORITHM

In the original k-means algorithm, the resulting set of clusters strongly depends on the selection of initial centroids which is random. Thus in our paper, we will propose a method for calculating the initial centroids, which will make the k-means algorithm more efficient.

Algorithm is divided into two phases. In phase one we define the initial cluster, while in phase two, data elements are moved in appropriate clusters.

7.2.1. PHASE I:To find the initial cluster

Input: Array $\{a_1, a_2, a_3, \dots, a_n\}$

Output: A set of initial clusters.

Steps:

- 1.Find the size of cluster S_i by calculating (n/k) .
where, n = number of data points (D_p) $(a_1, a_2, a_3, \dots, a_n)$
 k = number of clusters.
2. Create 'K' number of arrays A_k .
- 3.Move data points (D_p) from input array to A_k until S_i .
- 4.Continue step 3 until all D_p is removed from input array.
- 5.Exit with having 'K' initial clusters.

7.2.2. PHASE II: To find the final cluster

Input: A set of initial clusters.

Output: A set of K clusters.

Steps:

1. compute the arithmetic mean M of all initial clusters C .
2. Set $1 \leq j \leq k$
3. compute the distance D of all D_p to M of initial clusters C_j .
4. If D of D_p and M is less than or equal to two other distances of M_i $(1 \leq i \leq k)$ then D_p stays in same cluster Else D_p having less D is assigned to corresponding C_i .

5. For each cluster C_j ($1 \leq j \leq k$), recompute the M and move D_p until no change in clusters.

ADVANTAGES OF K-MEANS ALGORITHM

- ✓ Ease of implementation and high speed performance.
- ✓ Measurable and efficient in large data collection.

DISADVANTAGES OF K-MEANS ALGORITHM

- ✓ Selection of optimal number of clusters is difficult.
- ✓ Selection of the initial centroids is random.

8. CONCLUSIONS

Nowadays, the online customer reviews have become a valuable resource for any products purchase decisions. Besides helping consumers to make best choices among the products it also help vendors to get valuable feedback from their consumer to improve their product quality which in turn helps to improve their business .This strategy of data mining have been widely used to elevate sales on online. These opinion shared on the web shoppers are taken as resource raw data and the valuable statistical information are mined from the source data efficiently using CSA. There may be thousands and thousands of feedback for a same product. Thus CSA helps to summarize the review content and automatically analyze the opinion-rich feedbacks using machine learning as manual analysis is impossible.

9. RESULT

The problem of entry of fake reviews into the comment section exists and to avoid it we generate a unique key. This key is used to enter the reviews into the comment box which ensures security.

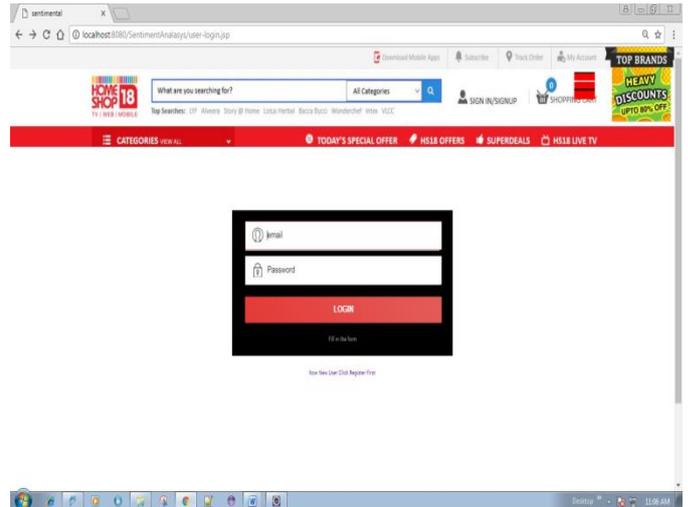


Fig.10.2.User interface

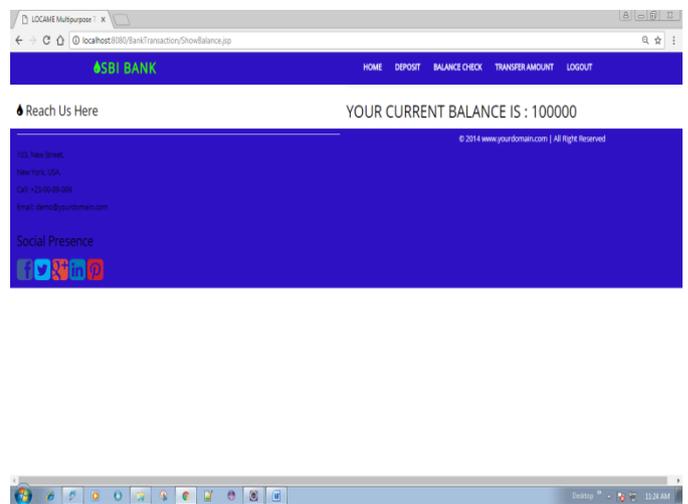


Fig.10.3.User transaction



Fig.10.4.Key generation

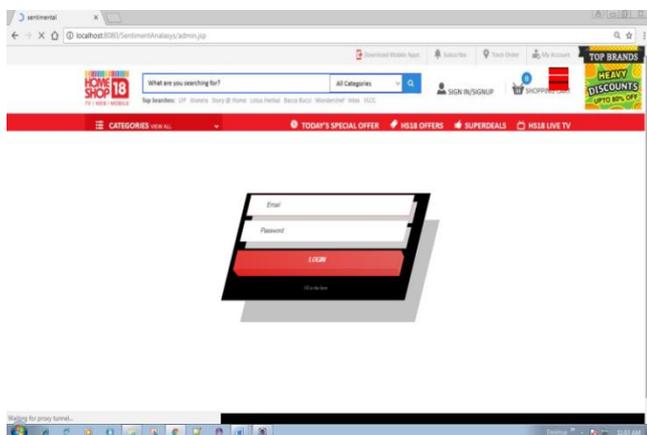


Fig.10.1.Admin maintain products

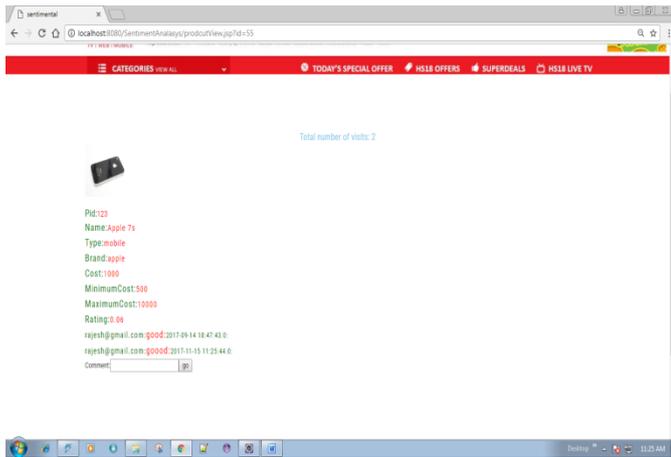


Fig.10.5.Comment sharing

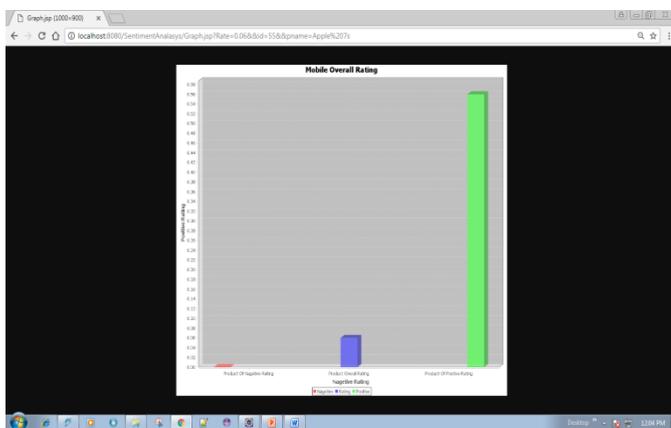


Fig.10.6.Graph generation

6. REFERENCES

[1] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, 2013.

[2] C.-L. Liu, W.-H. Hsaio, C.-H. Lee, G.-C. Lu, and E. Jou, "Movie rating and review summarization in mobile environment," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 3, pp. 397–407, May 2012.

[3] M. Pontiki et al., "SemEval-2014 Task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Ireland, 2014, pp. 27–35.

[4] T. Brychcin, M. Konkol, and J. Steinberger, "UWB: Machine learning approach to aspect-based sentiment analysis," in *Proc. 8th Int.*

[5] Z. Hai, K. Chang, and J.-J. Kim, "Implicit feature identification via co-occurrence association rule mining," in *Proc. 12th Int. Conf. Comput. Linguist. Intell. Text Process. (CICLing)*, Tokyo, Japan, 2011, pp. 393–404.