

2017

Feature level review table generation for E-Commerce websites to produce qualitative rating of the products

Kumar Raja

Dept. of CSE, St Peter's University, Chennai, India, kumarrajadr@outlook.com

S Pushpa

Dept. of CSE, St Peter's University, Chennai, India, cse@stpetersuniversity.org

Follow this and additional works at: <https://digitalcommons.aaru.edu.jo/fcij>



Part of the [Computer Engineering Commons](#)

Recommended Citation

Raja, Kumar and Pushpa, S (2017) "Feature level review table generation for E-Commerce websites to produce qualitative rating of the products," *Future Computing and Informatics Journal*: Vol. 2: Iss. 2, Article 7.

Available at: <https://digitalcommons.aaru.edu.jo/fcij/vol2/iss2/7>

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Future Computing and Informatics Journal by an authorized editor. The journal is hosted on [Digital Commons](#), an Elsevier platform. For more information, please contact rakan@aarj.edu.jo, marah@aarj.edu.jo, u.murad@aarj.edu.jo.



Feature level review table generation for E-Commerce websites to produce qualitative rating of the products

D.R. Kumar Raja*, S. Pushpa

Dept. of CSE, St Peter's University, Chennai, India

Received 23 June 2017; revised 7 August 2017; accepted 10 September 2017

Available online 20 October 2017

Abstract

It is widely acknowledged today that E-Commerce business is growing rapidly. This is happened only because of people are completely depending on the ratings and reviews given by the customers who are already purchased and using the products. Online surveys, customer reviews on shopping sites are the key sources to understand customer requirements and feedback to help upgrade the product quality and achieve greater outcomes. Now the challenge is that whether those reviews came from product level or feature level will be the million dollar question. To overcome this problem we are proposing a new algorithm to give feature level rating for the product which is called Feature Level Review Rating Analysis (FLRRA) algorithm.

© 2017 Faculty of Computers and Information Technology, Future University in Egypt. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Online surveys; E-Commerce business; FLRRA algorithm; FLRT; Reviews; Ratings

1. Introduction

Analysis of online user-generated content is receiving attention for its wide applications from both academic researchers and industry stakeholders. Due to the ease and availability of internet every customer is able to check the durability, sustainability, and availability of the product while sitting at their home [1]. Every customer will check for the product rating and reviews before purchasing the product. To speak in general not only the E-commerce websites but also every service-oriented businesses like hotels, hospitals, theatres, shopping malls, amusement parks, and almost every service providing businesses are running now online with the help of internet.

For example, before purchasing a product first do some survey on the product. Generally this is done with the help of the person who already bought earlier but now with the help of online shopping we will look for the reviews given by the customers who bought and using the product. Not only before purchasing a product but also before visiting any new place if we are supposed to book a holiday tour then, first of all, we will check for which place we can go, where we can stay and what are all the tourist spots available near to the place like that everything we can check on the internet. The only thing which every customer will look at is the feedback given by the people who have already visited that place in terms of reviews.

The biggest problem for the customer who is going to take decision depending upon the ratings given by the existing customers is “weather the ratings are genuine are not?” taking this into consideration we did a case study on one product which is available on Amazon.in and found that the existing rating system that which will be explained in detail in later sections is not genuine that means the present rating system is creating Flase_Positive rating for the products.

* Corresponding author.

E-mail addresses: kumarrajadr@outlook.com (D.R. Kumar Raja), cse@stpetersuniversity.org (S. Pushpa).

Peer review under responsibility of the Faculty of Computers and Information, Future University in Egypt

In this paper we are proposing a new method with the help of a generic algorithm which is used to generate the rating of the products depending upon the reviews written by the customers.

2. Existing work

As we have considered Amazon as the source of data, the rating process which is used to rate the products in Amazon will lead to Flase_Positive which in turn leads the customers to purchase the products by seeing the false rated products [2]. For example, consider a sample product review which is given by a user in Amazon as shown below.

Fig. 1 shows that the process of giving the ratings to products. As on when if the rating is in the form of stars we can understand that highest number of stars indicates that the product overall rating is good and the customer satisfies to the maximum extent. If it goes to the least number the customer completely disappointed with the product and it is not up to the range of customer's needs and the remaining rating process gives the satisfactory ranges of the moderate customers who ever partially satisfy with that product. Let we see some other example for the product overall rating in the bellow figure. Fig. 2.

This describes that the customer is satisfied with the product and they have given the rating 5 in the form of stars. We'll try to undergo some more examples of this traditional process of giving the ratings by using some sample reviews in Amazon website. These examples will illustrate how the review process will lead to a False_Positive mechanism. Fig. 3.

Fig. 4 is an example of the False_Positive situation. This describes that a particular customer has given the rating for the product as 5 and in the description part he has written: "the product is worst". We can say these tow situations are contradicted to each other and also it is very difficult to the users who are trying to buy that product. This type of reviews and ratings will take the customers to a confusion state. Here there is no clarification on the product quality and features. Fig. 5.

This is also another kind of rating process for the Flase_Positive situation that describes a customer have given the rating as 1 and they describes the product as good. Such kind of rating process also takes the customer to a confusion state to buy a product or not.

There is no clarity in rating process of such kind of examples; this will take the customers to confusion state and will show the negative impact on buying the products. The actual problem with such kind of rating is, there is no clarity on the description of a product and also they did not provide the feature level of rating. The word feature level rating will give the description oriented rating for the products and we are introducing this kind of mechanism in our proposed approach.

Here the main problem is we can select the rating of the products in the form of stars scaling from 1 to 5. In this process 5 is for I Love it and 1 is for I Hate it. And even is we select 5 which means I Love the product and we can write in the review as "The Product is Worst". This will lead the customers to False_Positive situation.

3. Proposed work

By analysing all the existing review and rating processes we have observed that there is no clarity on the products rating and there is no feature level rating of the product which helps the customers to buy a product based on a specific feature. In our proposed approach we are trying to concentrate on the feature level rating process of the product by considering the reviews given by the users [4]. Here we are trying to identify the "key_terms" from the reviews. These key terms are derived by the help of association rule mining in the big data analytics. After identifying the key terms we will relate those to the products and also we'll try to identify the key patterns from the detailed reviews of the customers. Later we will establish a connection between all the key terms and key patterns for generating the feature level rating. The keywords are the terms which describe the product nature in the form of positive nature like "good, awesome, fine, fabulous" etc [5] can be identified and the key patterns are in the form of product

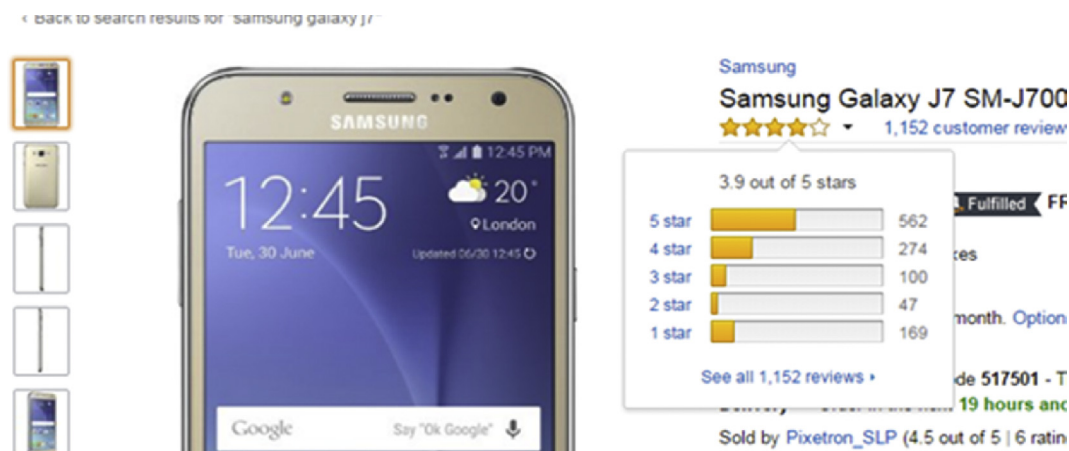


Fig. 1. Overall product rating by the customers.

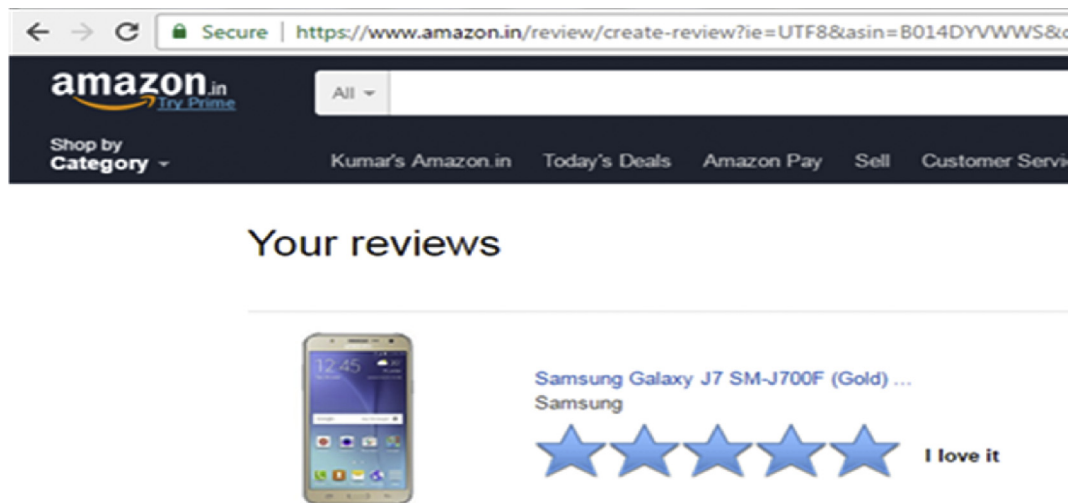


Fig. 2. Opinion of a customer on the product in the form of ratings 5-star rating.

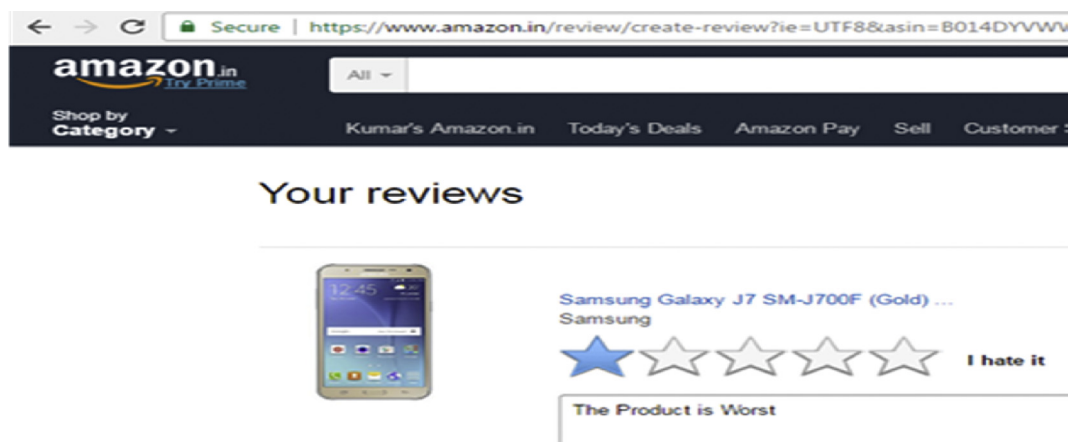


Fig. 3. Opinion of a customer on the product in the form of ratings 1-star rating.

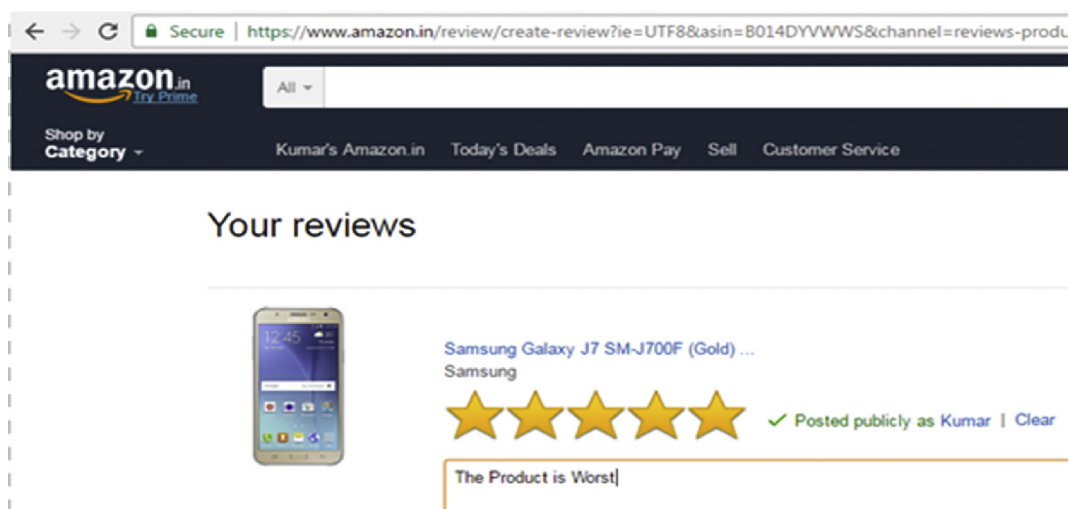


Fig. 4. False_Positive situation for a product in terms of high rating and disappointed review.

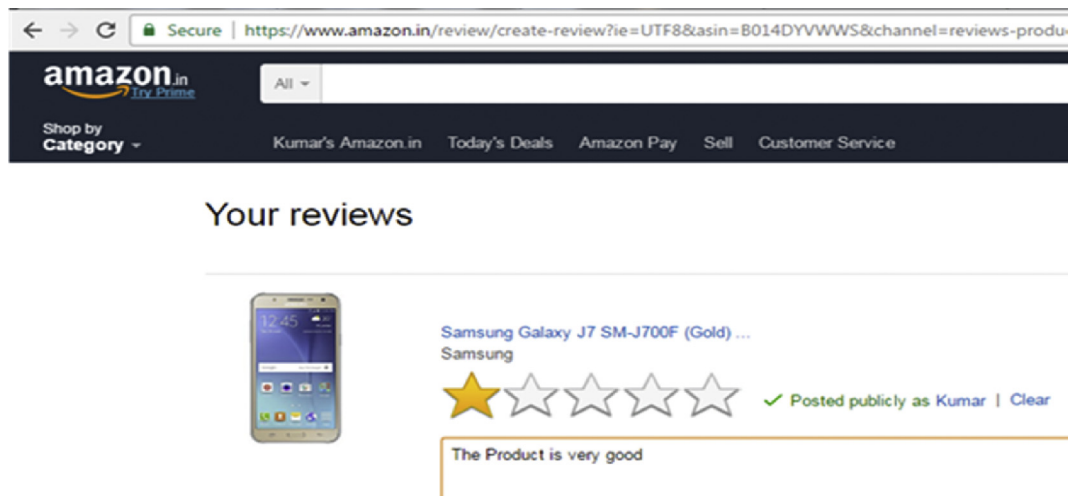


Fig. 5. The False_Positive situation for a product in terms of less rating and satisfied review.

features which are derived from the detailed reviews of the customers.

After identifying these two we will relate each keyword with each key pattern in a table. From that table, we can easily provide the feature level rating on the products. Here also the traditional star rating system is used to judge the product quality but this is in terms of feature orientation.

3.1. Steps to follow

- Gather the reviews of the products in terms of data set
- Process that data set to get key term and key patterns by using any FP Growth algorithm.
- Construct the feature level table with the Feature with its quality (quality means the quality of the rating like good, bad, average) based on the association results.
- Assign quality related words to the key patterns in the table
- Generate the result which shows the feature level ratings.

As we discussed above, we have to refine the data set for identifying structured reviews from the Amazon website. In that refining process we are going to use association rule mining for gathering the product related reviews.

Opinion mining is the process of extracting opinions from the Web. An opinion is a private state of an individual, and as such it represents the individual's ideas, beliefs, assessments, judgements and evaluations about a specific subject/topic/item. Opinions can be taken by customer from friends, relatives and even user reviews before or after buying of the product. Opinions can be collected from different sources like individual interactions, newspapers, television, internet and many.

An opinion has three main components. They are:

1. Opinion holder or source of the opinion.
2. Object about which opinion is expressed.

3. Evaluation and appraisal that is the opinion.

Opinions can be expressed in different ways. Following are some examples.

- Shahid Afridi is a good player.
- The hotel was expensive.
- Coffee is expensive, but tea is cheap.

With the advent of using World Wide Web (WWW), the content on the web is increasing day by day with lots of topic discussions and their comments also called opinions. Opinions have a great impact on and provide guidance for individuals, governments, organizations and social communities during the decision making process.

There are two kinds of opinion sentences are there they are:

1. Structured Sentences: Sentences which have defined a format of vocabulary rules are can be called as Structured Sentences [3].

Camera is good for photographers

Noun Adjective Adverb

2. Unstructured Sentences: Sentences which do not have any defined format of vocabulary rules are can be called as Unstructured Sentences.

Example: I purchased this for my father the first need for purchasing this product is that he can operate it very easily. But at the first they didn't mention anywhere like it is not suitable for the people who are not able to operate touch interfaced mobiles and then after purchasing the product it was working properly but later on it is not working effectively [3]. Fig. 6.

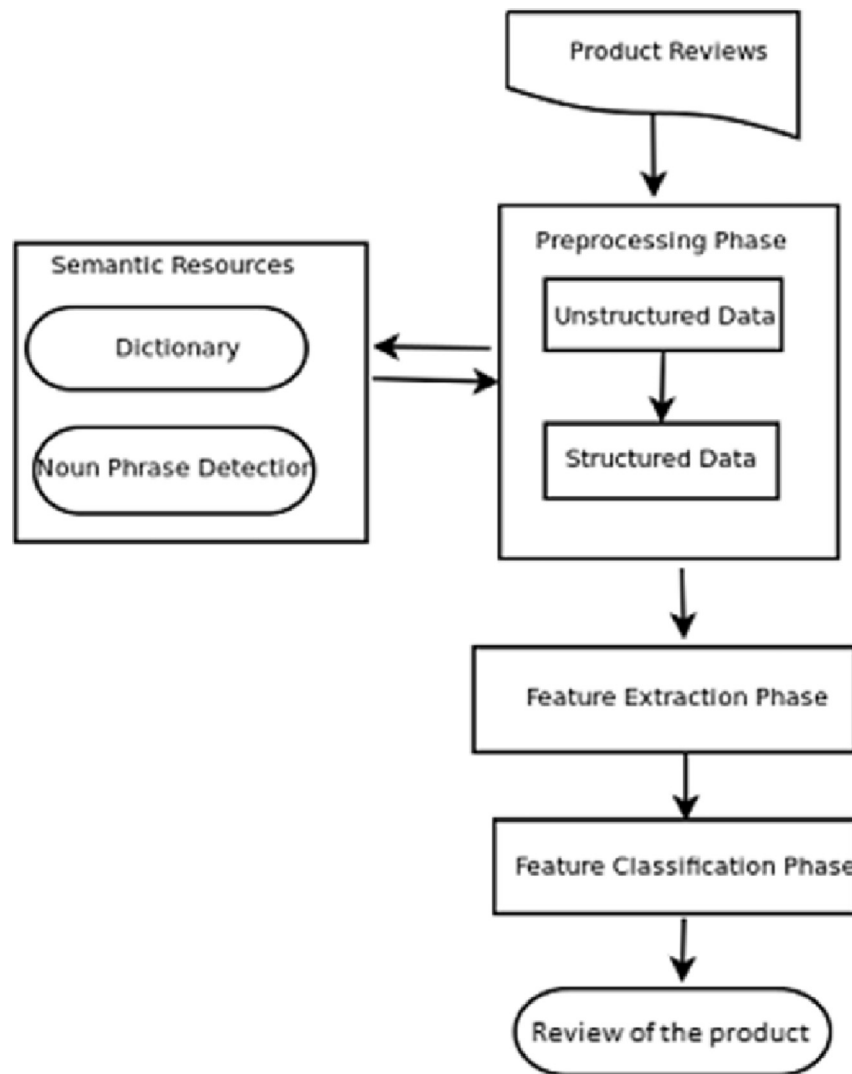


Fig. 6. Flow chart of the Algorithm.

4. Methodology

The ratings of the products should be given by evaluating the reviews written by the customers. For evaluating the reviews we are proposing a generic algorithm which will consider the reviews written by the customers and rate the products.

5. FLRRA algorithm

Require: Product Review Document.

Ensure: Rating of the product.

1. Read the Document.
2. Convert the unstructured document to structured document.
3. Split the text document into sentences.
4. Tokenize the sentences into keywords.
5. Eliminate Stop words and tag the tokens using POS tagger.

6. Tag the tokens using POS Tagger and identify the Noun phrases.
7. Identify frequently occurring Noun phrases using term frequency distribution.
8. Mine feature descriptors and linguistic hedges using the dictionary.
9. If term is not in the dictionary check for the synonyms of candidate terms.
10. Extract opinionated Phrases.
11. Return review and review score of given text.

In pre-processing the review data, we parsed all the review documents in each data set using the well-known Stanford Parser, which has been previously shown to perform well in working out grammatical dependency structures of texts in practice [8].

1. <http://mpqa.cs.pitt.edu/lexicons/subjlexicon>
2. <http://liu.cs.uic.edu/download/data>
3. <http://www.amazon.com>

4. <http://sifaka.cs.uiuc.edu/wang296/Data/index.html>
5. <http://www.tripadvisor.com>
6. <http://nlp.stanford.edu/software/lexparser.html>

The data which we get from the E-Commerce websites have to be refined to work with the data. Refining involves the data to be converted to a suitable format to work with. For example, we have taken the JSON format data from Amazon and converted that data to. CSV file format system to give as an input to Rapid Miner tool.

6. Pseudo code

```

{
  Create Feature Level Table
  //Table will contain a list of all the products available in the
  E-Commerce Website. Product Name is the one that which we
  have to give according to our preference-give the contents of
  the table//
}
Else
{
  //go to the initial stage of the base file and fetch the another
  name
}
For (i=0;i<a[n];i++)
  //a[n] is total number of products available on the website//
  {
    Mine the table with Association Rule Mining Technique
    using FP-growth algorithm.
    //The major advantages of FP-Growth algorithm is, Uses
    compact data structure and Eliminates repeated database scan
    and it is cheap compared to Apriori Algorithm [7]//
    If Key_Word == Good
    {
      Rating=5
    }
    Else If
    {
      Key_Word==Bad
    }
    Rating=1.
  }
  Else{
    Key_Word!= Good&&Key_Word!=Bad
    Rating=3//WHAT ABOUT OTHER RATINGS
  }
}
}

```

7. Implementation

For implementing the algorithm automatically in any data mining tools like Rapid Miner, Weka etc we require full licensed version of the tool. Because of the availability of data in large amounts of size it is not possible for us to work with these tools even though they are freely available for use.

Manually the algorithm was tested on Samsung Galaxy J7 product with the reviews of total 8778 and classified into two categories as Positive_Reviews 6903 and Negative_Reviews 1875. Now to test the process given two reviews with 5 star rating and review as the product is bad, then surprisingly we found that the total reviews was 8781 with 6905 Positive_Reviews and 1875 Negative_Reviews.

After looking this copied all the reviews and analysed them manually and divided them into two categories then we got 5974 Positive_Reviews and 2807 Negative_Reviews using Feature Level Review Rating Algorithm and came to know that which leads to False_Positive results and makes the customers take wrong decisions. With the proposed algorithm in the paper we will get good results compared to the existing system.

8. Results

Rating system used	Positive reviews	Negative reviews
Normal rating system	6905	1875
Using FLRRA	5974	2807

Total number of reviews available for Samsung Galaxy J7 are 8781 and without using any algorithm and existing system classified them by using the star ratings as 6905 are positive reviews and 1875 reviews are negative which is wrong. Using FLRRA the same 8781 reviews are classified as 5974 are positive reviews and 2807 are negative reviews which proved that existing rating system is no effective.

9. Conclusion

This paper proposes a new algorithm to understand the individual's opinion on a certain feature of a product this will be used to analyse the productivity of the particular feature of the product. It is used as a decision-making strategy for a company to take decisions on the product and the features of the product. Our method is useful to generate feature level rating of any product and this will avoid False_Positive ratings of a product with an existing approach which is being used by almost all E-Commerce websites.

References

- [1] Kumar Raja DR, Pushpa S, Naveen Kumar BS. "Multidimensional distributed opinion extraction for sentiment analysis – a novel Approach", International Conference on Advances in Electrical, Electronics, Information, Communication, and Bio-Informatics (AEEICB16) 978-1-4673-9745-2 ©2016 IEEE.
- [2] Ups and downs: modeling the visual evolution of fashion trends with one-class collaborative WWW 2016, April 11–15, 2016, Montréal, Québec, Canada ACM 978-1-4503-4143-1/16/04. <https://doi.org/10.1145/2872427.2883037>.

- [3] Dhana lakshmi P, Ramani K, Eswara reddy B. Feature relevance analysis in on-line marketing to improve productivity. *I-Manager 's J Software Eng* January – March 2015;9(3).
- [4] Dependency Parsing Tutorial at COLING-ACL, Sydney 2006 Joakim Nivre Sandra Kubler Uppsala University and Växjö University, Sweden. Eberhard-Karls Universität at Tübingen, Germany.
- [5] Yang Jie, Yecies Brian, Yang H. Mining Chinese social media UGC: a big-data framework for analyzing douban movie reviews and Yecies. *Journal of Big Data* 2016;3:3. <https://doi.org/10.1186/s40537-015-0037-9>.
- [7] Implementation of web usage mining using APRIORI and FP growth algorithms. *Int J Adv Netw Appl* 2010;01(06).
- [8] Maas AL, Daly RE, Pham PT, Huang D, Ng AY, Potts C. Learning word vectors for sentiment analysis. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies – Volume 1*, ser. HLT'11. Stroudsburg, PA, USA: Association for Computational Linguistics; 2011. p. 142–50.