

A Project/Dissertation Review Report

on

PRODUCT REVIEW ANALYSIS

*Submitted in partial fulfillment of the
requirement for the award of the degree of*

MCA (CSE)



**Under The Supervision of
DR.K.M.Baalamurugan
Assistant professor**

Submitted By

1. Abhay Pratap Singh 20SCSE2030057
2. Shivam Kumar 20SCSE2030093



**SCHOOL OF COMPUTING SCIENCE AND
ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA**

CANDIDATE’S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled “**PRODUCT REVIEW ANALYSIS** ” in partial fulfillment of the requirements for the award of the MCA-submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of

DR.K.M.Baalamurugan, Assistant professor, Department of Computer Science and Engineering, of School of Computing Science and Engineering , Galgotias University, Greater Noida .

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Abhay Pratap Singh, 20SCSE2030057

Shivam Kumar, 20SCSE2030093

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

DR.K.M.Baalamurugan

Assistant professor

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Abhay Pratap Singh(20scse2030057) Shivam kumar (20scse2030093) has been held on 13th MAY 2022 and his/her work is recommended for the award of MCA.

Signature of Examiner(s)

Signature of Supervisor(s)

Signature of Project Coordinator

Signature of Dean

Date: 13th MAY ,2022

Place: Greater Noida

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
May,2022**

Abstract

-

Online reviews have become a major factor in people's decision-making and business decisions. A retailer that sells products to the web often asks or takes reviews from customers about the products they have purchased. As e-commerce grows and becomes popular day by day, the number of updates received from a customer about a product is growing rapidly. With a famous product, updates can be up to thousands. The growing popularity of online reviews also encourages the business of writing a fake review, which refers to paid human writers who produce fraudulent reviews to influence readers' opinions. Our project addresses this issue with construction a separator that takes the text of the review and the basic information of its reviewer as input and determines whether the update is reliable. This creates difficulties for the customer who may read and decide whether to buy the product or not. Problems also arise so that the manufacturer can follow and manage customer feedback. And the more difficulties he faces manufacturer because many other retailer sites can sell the same product at good prices and the manufacturer often produces many types of products. In this study, we aim to summarize all customer reviews of the product and compare based products in the review can be done in one place. This summary function is different from the standard text summary, because we are the only one information about that product in which customers have expressed their views and whether the ideas are correct or not negative. We summarize the review by choosing to rewrite some of the original comments, from reviews to great snap. Points as in the summary of the ancient text. Our experimental results using a review of the number of products sold online are indicative strategic efficiency.

Table of Contents

Title	Page No.
Abstract	I
List of Table	II
List of Figures	III
Chapter 1 Introduction	4
1.1 Introduction	5
1.2 Formulation of Problem	6
1.2.1 Tool and Technology Used	8
1.3 Frameworks used	8
Chapter 2 Literature Survey/Project Design	10
Chapter 3 Methodology	11
Chapter 4 Results and Discussion	16
Chapter 5 Conclusion and Future Scope	18
5.1 Conclusion	18
5.2 Future Scope	18
Reference	19
Publication/Copyright/Product	20

CHAPTER-1

1.1 Introduction

Web-based life has become part of everyone's daily life. Used to communicate with friends, family members, people who live far away and next to this, it allows people to express their opinions about things by commenting on the E commercial area. The purpose of this paper is to express the feelings of the author or client within the nature of the audit; it will be good or bad for something. These thoughts or feelings that are expressed by individuals are called emotions. Emotional Analysis also referred to as Opinion Mines is an in-house field of Natural Language Processing (NLP) that integrates frameworks that attempt to differentiate and differentiate exploration within content. As data that is accessible and confidential via the Internet grows exponentially, an equal number of articles exposing customer surveys are available at research sites, circles, online journals, and web-based life. From a client's point of view, individuals are more technologically savvy and more willing to post their content through web-based social networking sites, for example, interpersonal and social networking sites. In today's developed world, most customers are prone to e-commerce, because of their value propositions, but especially because they need a review and performance response to rate the product and measure it according to the user's feelings. In addition, it is not uncommon for customers to post reviews about what they are buying, for better or for worse.

1.2 FORMULATION OF PROBLEM

This paper refers to an experimental website event where mind-boggling and analytical methods are used. The test website contains products such as clothing and shoes where the user can purchase products and submit updates equally. It has two panels, namely, user and administrator. User Panel It is the responsibility of any user to register on the website. Only then can he explore the website.

1.2.1 TOOLS USED

The user can add products to cart and also has the choice to purchase them. He can filter the products supported categories like brands, color and price. He can go through other customers review and supported the ratings can decide whether he/she wants to shop the product. After using it customer can rate the product and provide star ratings. Admin Panel Admin is that the one who handles the website and allows only registered users to access the website. He has the authority to add/delete products, edit product information and view user. The different processing components of the system are as follows:

- A. Extract Suitable Text After the user has logged in and posted a review for any product(s), this component extracts all the text from the review section of the product.
- B. List of Products This module will display an inventory of products from which reviews are going to be extracted upon selection of certain products.
- C. Show Review List This module will show all the reviews of the given product.
- D. Sentiment Visualization This module shows a dynamic bar chart representation measured in terms of positive and negative sentiments for a given product.

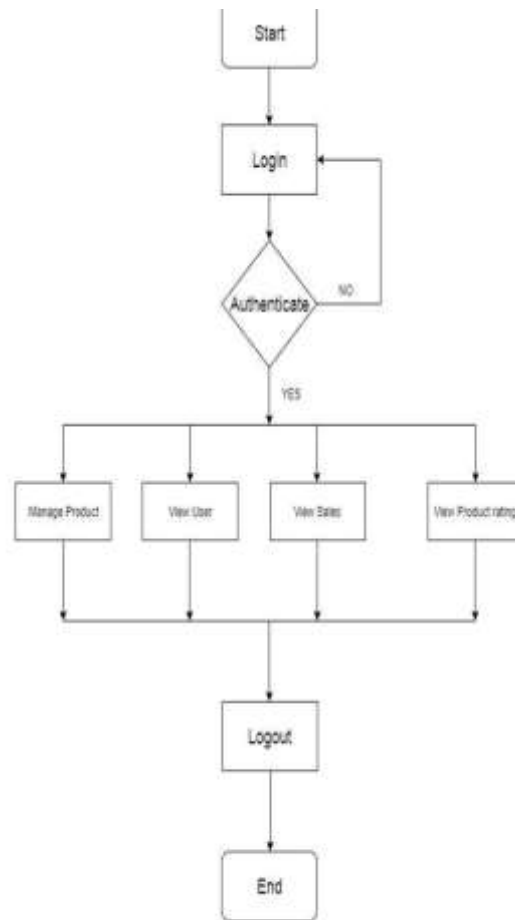


Fig -1: Admin Activity Diagram

Fig -1 represents the Admin Panel of the website where the admin manages the site infrastructure. In Manage product, he/she can add/edit, view or delete the products based on the brands. In View User, as the name suggests the admin has the privilege to access the details of the user i.e. username, email id, contact details and user address. In View Sales, the admin can view the user's sales i.e. the products bought from the site and in View Product Rating, can access each product's ratings.

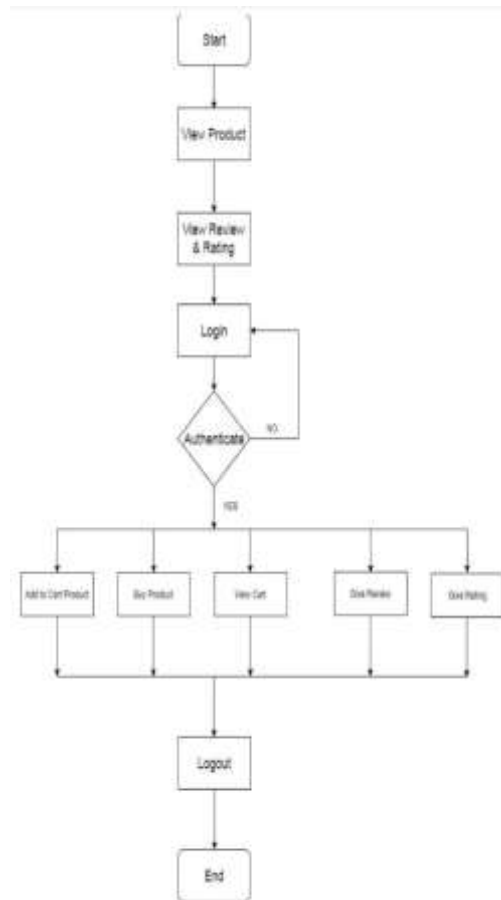


Fig -2: User Activity Diagram

Fig -2 represents the user panel of the website. The user must authenticate himself for carrying out above privileges. In Add to Cart product, the user can filter out products based on the brands, color, price and can add the product to their cart. Buy Product means the user can simply purchase the product and he/she would be redirected to the payment page. In View Cart, the user could view the products in his cart and perform out subsequent operations. The user could add reviews or ratings after using each product.

1.3 FRAMEWORKS USED

1. .NET: The framework used in our emotion project is .NET. .NET is a free open source developer platform used to build many types of technology applications. With .NET, you are empowered to use multiple languages, editors to work better, and libraries you can create for the web, mobile, desktop. This easy-to-use techno framework can be used to design, develop, integrate, build, and supply an app with its useful plugins for compilers, code libraries, etc. Net supports various programming languages such as Visual Basic, C #, C, Java.

2. Sentiment Intensity Analyzer Package: In the review system we used the Sentiment analysis model using Vader Sharp which is the best emotional analysis tool in C #. VADER is widely known as a lexicon analysis tool and legal framework that is explicitly adapted to extract hidden emotions in texts displayed on social media. By using it, we should only enter the package at the top of the page and then simply launch the Sentiment Intensity Analyzer example and call its Polarity Scores method. Here we have used the Sentiment Intensity Analyzer package item that pulls text updates and assigns polarity points to each of them. It then organizes the entire content survey into positive, negative or impartial based on high scores and predicts the exact result of each client survey. This may help other registered users to encourage in-depth feedback on whether to purchase for sale or not.

CHAPTER-2

Literature Survey

Web-based life has become part of everyone's daily life. Used to communicate with friends, family members, people who live far away and next to this, it allows people to express their opinions about things by commenting on the E commercial area. The purpose of this paper is to express the feelings of the author or client within the nature of the audit; it will be good or bad for something. These thoughts or feelings that are expressed by individuals are called emotions. Emotional Analysis also referred to as Opinion Mines is an in-house field of Natural Language Processing (NLP) that integrates frameworks that attempt to differentiate and differentiate exploration within content.

Here we propose an advanced product review system that provides a platform for registered users to evaluate one or more products using this system. The system uses product update analysis to achieve the desired performance. Product Review Analysis is a web application that combines multiple products added by the administrator for review to evaluate and review. The system takes reviews of various users, based on their own opinion, the system will determine whether the product posted is good, bad, or very bad. We use a database of emotion-based keywords and the direct or indirect weight on the website and based on these emotion keywords extracted from user reviews are rated. When a user logs into the system they look at multiple products and provide an update on the product. The user can view the product description, price and links to purchase the product. The system will use the website and will match the update to the site keywords and will list the update. The role of the director is to add new products, their description and provide a link to purchase the product on the website. The administrator can also view additional products, view registered users and view system-related feedback from registered users. This app is also useful for users who want to purchase a new product. This program helps you find a good product based on good reviews of many users.

In today's developed world, most customers are prone to e-commerce, because of their value propositions, but especially because they need a review and performance response to rate the product and measure it according to the user's feelings. In addition, it is not uncommon for customers to post reviews about what they are buying, for better or for worse.

CHAPTER-3

Methodology

Data preprocessing

Dataset used in this research was dataset from another research conducted by McAuley and Leskovec (2013), which was a product review from Amazon.com from June 1995 - March 2013, data retrieved from <https://snap.stanford.edu/data/web-Amazon.html>.

Examples of data of (as shown in the figure-1) unpopular product, ie a product that has less than 3 reviews. The fourth stage is the removal of inactive users, ie users that has less than 3 review. After the filtering stage finished, the data generated is clean and are ready to be analyzed.

```
product/productId: B000FTPOMK
product/title: 14k Yellow Gold Butterfly Pendant, 16"
product/price: unknown
review/userId: A3T4KKUM1JATUX
review/profileName: Disappointed Sony customer
review/helpfulness: 1/1
review/score: 4.0
review/time: 1178150400
review/summary: pretty necklace
review/text: It is a nice made necklace, and the butterfly pendant looks beautiful. I love it.
```

Figure-1

ICF ++ method

ICF ++ method took the name of a product, the name of a reviewer, the content of a review, the number of iteration and minsup as parameter. After taking those values, this method initializes the counter to 0, and honesty value and reliability value to 1. After that, the next process are pos tagging, creation of transaction file, FP Growth, polarity generation, calculation of agreement value. The next step is iteration. Inside in iteration, the method continues by calculating the honesty value, calculating the trustiness value, calculating the reliability value, updating agreement value using updated honesty value, updating reliability value and increment counter. Iteration in this research is 3, because the algorithm converges quite fast in practice. The process are described in the figure-2.

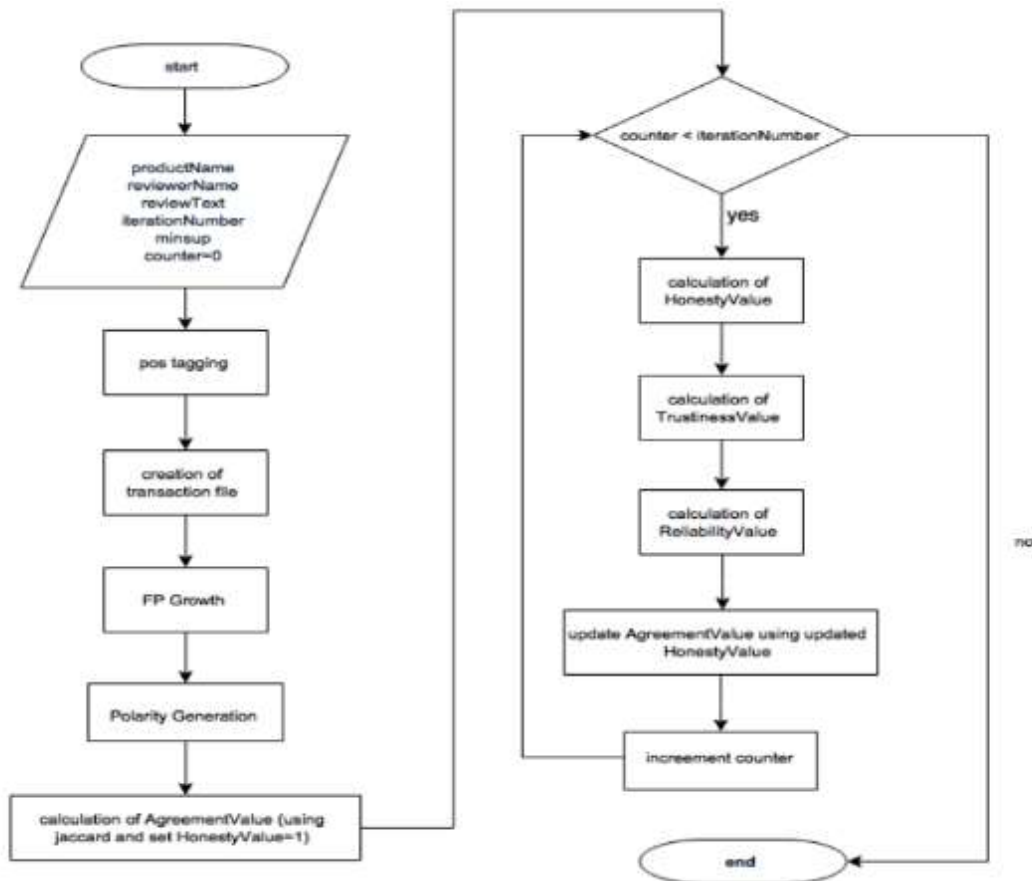


Figure-2

POS tagging

POS tagging was conducted by utilizing the Stanford Part-Of-Speech (POS) tagger from stanfordcorenlp 3.5.1. Stanford Pos Tagger is a software that reads text and determine the part of speech for each token.

This process produced a tag in the form of abbreviations, for english tagger, the resulted abbreviations used Penn Treebank standard.

For this process, a desktop application was developed using Java. This process started with training the tagger model. Input from this process is a sentence from review and produce POS tag for each token that is stored into the database.

Creation of transaction file

Transaction file in this research contains all tokens that are stored in the database from a products that it's tag value is noun, either NN, NNP, NNPS or NNS.

Each row in the transaction file was a noun from every sentence for a product. This transaction file became the input for the FPGrowth process.

FPGrowth

This process aimed to identify the features that have the most comment. In this study, the features are properties or attributes from a product. For example, for camera, the feature can be its battery, memory card etc.

This information actually available on amazon, but, datasets related to this information were not publicly available, so to get information about product's feature, this research used fp-growth algorithm from association rule mining technique. This study used spmf library developed by (Fournier-Viger, et al., 2014).

The minsup used in this research was 0,1 and the confidence was 0,8. The minsup and confidence values ranging from 0 to 1. The minsup (minimum support) used for the comparison against the support that was owned by an item that compose a rule. Rule produced by this algorithm should consist of items that it's support value are exceed the minsup value. Support value was calculated from the number of item appearance in a transaction. For example, there was an item that appears

two times in 5 transactions, then the support was $2/5$ which is 0.4. If the minsup value entered was 0.1, then this item would appear in the output, but if the minsup value entered is 0.5 then this item will not appear in the output.

The FP Growth process in this research described as follow, input for this process are the transaction file that contains nouns. The next stage, are FP-Tree generation. FP tree is data structure that contains specific information about item, number of occurrence of the item and link to other node. The next stage, this fp tree will be scanned to generate frequent pattern in the form of rule. The resulting rule will be broken down to obtain each phrase. Each of these phrases is an attribute of a product. The last stage is, the attribute will be saved to the transaction file and to the database.

Polarity generation

The feature or attribute resulted from the previous process will be identified its opinion orientation. For example, for a camera, the feature or attribute that can be identified is the battery, the comment related to the battery from userId X is “this battery drain pretty fast, and you do not want to be stuck somewhere with a dead battery”. The purpose of this process is to identify the opinion orientation from a sentence that contains the attributes that were identified from the previous process. Whether its orientation positive, negative or neutral. From the example above, its opinion orientation is negative. Sentiment prediction/opinion orientation prediction can use variety of techniques, from the words level, sentence level and others. According to Socher, etal., (2013), sentiment prediction only from words often misses, because the order of each word was ignored. As in the illustration below, the words funny and witty each are positively oriented, but in the sentence “This movie was actually neither that funny, nor super witty”, causing the orientation of the overall sentence to be negative. The input for this process are two parameters, ie a list of all tokens from a review of a current product (sentences that have been POS tagged and stored into the database), named list i. The second parameter is a list of all the features from a current product, named list f, which is obtained from the database. This process will check whether index i-th from list i contain a token from list f. If this condition is met, it will look for the orientation by using `stanford.nlp.sentiment` from `stanford-corenlp-3.5.1`. library. The output of this process is a list of

features contained in a single sentence along with the orientation of the feature, as the following examples “necklace pos pendant neg”.

Calculation of agreement value

The next process was calculating the agreement value that illustrates how similar the feature and its opinion orientation resulted from the polarity generation process with the other reviews. This process uses Jaccard similarity. Results from Jaccard similarity was in the range of (0,1), 0 means that the second review didn't have the same opinion orientation for any feature, 1 means both reviews had the same opinion orientation for all the features. In this research, both of the reviews agreed each other if their Jaccard value exceed 0.8. The surrounding set of review v (S_v) was the set of v 's surrounding reviews about a product.

$$S_v = S_{v,a} \cup S_{v,d}$$

where $S_{v,a}$ is set of the reviews that agree with each other, $S_{v,d}$ is a set of reviews that disagree with each other. From the definition above, $S_{v,a}$ is in the form of a list of a reviewID that contains the other review which it's jaccard value exceed 0.8. Whereas, contains other review which it's jaccard value less than 0.8.

The agreement value from a review v within a period Δt is described in the following equation

$$A_n(v, \Delta t) = \frac{2}{1 + e^{-A(v, \Delta t)}}$$

$$A(v, \Delta t) = \sum_{i \in S_{v,a}} T(\kappa_i) - \sum_{j \in S_{v,d}} T(\kappa_j)$$

where A is the agreement value of the target review. In this research, we used 3 months before and after the posting time of the target review as the time window Δt .

Calculation of honesty value

Reviewer's honesty value is described in the following equation

$$H(v) = |R(\Gamma_v)|A_n(v, \Delta t)$$

Where $R(\Gamma_v)$ is product's reliability value that is being reviewed.

Calculation of trustiness value

Reviewer's trustiness value is defined in the following equation

$$T(r) = \frac{2}{1 + e^{-Hr}} - 1$$

Where $T(r)$ is the reviewer's trustiness value.

Calculation of reliability value

Product's reliability value ($R(p)$) is described in the following equation

$$R(p) = \frac{2}{1 + e^{-\theta}} - 1$$

Where

$$\theta = \sum_{v \in U_s, T(\kappa_v) > 0} T(\kappa_v)(\Psi_v - \mu)$$

Where κ_v is the rating from review v , while μ is the median value of the rating, which is 3, in the range of 1-5 rating.

CHAPTER-4

RESULT AND DISCUSSION

There are two types of tests in this research, which is verification of results test and performance test. The purpose of the test of verification of the results is to check whether the calculations performed by the application and manually calculated has the same process and the same result. This test also intend to check if the given input has the correct output. While the aim of the performance test is to compare results from the ICF ++ method with another fake review detection methods. Method from Wang, et al., (2011) is re implemented using java to see the performance. Then the fake review detection methods in Wang's research used as a reference comparison with method ICF ++.

This idea is difficult to implement in real life because it requires the cooperation from banks, which aren't willing to share information due to their market competition, and also due to legal reasons and protection of data of their users. All individuals mentioned in this list had their cards closed to avoid any risk due to their high-risk profile. The condition is more complex for the other list. The level 2 list is still restricted adequately to be checked on a case by case basis. Credit and collection officers considered that half of the cases in this list could be considered as suspicious fraudulent behaviour. For the last list and the largest, the work is equitably heavy. Less than a third of them are suspicious.

CHAPTER-5

5.1 CONCLUSION

We furnished results for Sentiment and Emotional Analysis on twitter data .On applying Logistic regression, Bernouille Naive Bayes and Multinomial Naive Bayes for sentiment analysis Multinomial Naive Bayes stands out with 96.4% accuracy at test split=0.3.Users topic of interest for sentiment analysis has been considered ,So that they may get to know the statistics of sentiment behind the topic of their own interest. We firmly conclude that implementing sentiment analysis and emotional analysis using these algorithms will help in deeper understanding of textual data which can essentially serve a potential platform for businesses.

5.2 FUTURE SCOPE

In future work , we aim to handle emoticons , dive deep into emotional analysis to further detect idiomatic statements .We will also explore richer linguistic analysis such as parsing and semantic analysis.

REFERENCES

- [1]. Shiv Dhar, Suyog Pednekar, Kishan Borad- Methods for Sentiment Analysis Computer Engineering, VIVA Institute of Technology, University of Mumbai, India
- [2]. Elsharif Elmurungi, Abdelouahed Gherbi, “Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques”, École de Technologie Supérieure Montreal, Canada, DATA ANALYTICS 2017: The Sixth International Conference on Data Analytics W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
- [3]. B. Jiang, R. Cao, B. Chen, “Detecting Product Review Spammer using Activity Model,” Proceeding of International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013), pp. 650-653,

RESEARCH PAPER PUBLICATION

TITLE: PRODUCT REVIEW ANALYSIS

<http://www.ijser.org>