



A Project Report on

**Fake Product Review Monitoring and Removal for
Genuine Online Product Reviews Using Opinion Mining**

Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering

by

Ankit Tripathi 1317180

Under the Guidance of

Chinthakunta Manjunath

and

Deepak Biradar

Department of Computer Science and Engineering

Faculty of Engineering, Christ University,

Kumbalagodu, Bengaluru - 560 074

April-2017



Faculty of Engineering

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that **Ankit Tripathi** has successfully completed the project work entitled “**Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining**” in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** during the year **2016-2017**.

Chinthakunta Manjunath

Guide

Deepak Biradar

Industry Guide

Dr. Balachandran K.

Head of the Department

Dr. Iven Jose

Associate Dean



Faculty of Engineering

Department of Computer Science and Engineering

BONAFIDE CERTIFICATE

It is to certify that this project titled "Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining" is the bonafide work of

Name	Register Number
Ankit Tripathi	1317180

Examiners [Name and Signature]

1.

2.

Name of the Candidate :

Register Number :

Date of Examination :



EFFENT TECHNOLOGIES

"An effective enterprise"

Certificate of Project Completion

This is to certify that **Mr. Ankit Tripathi (Reg no.1317180)** and **Mr. Utkarsh Gautam (Regno.1317192)** of **8th Semester B.Tech (CSE)** from **Christ University Faculty of Engineering, Bangalore** have successfully completed their 4 month internship project with our organization. They were part of organization as interns from **15th November to 15th March 2017**.

Their project title was **"Fake Product Review Monitoring And Removal For Genuine Online Product Reviews Using Opinion Mining"**

We wish them all the best in their future endeavours.

Deepak Biradar

M.Tech, (Ph.D)

Head, Operation



Acknowledgement

I would like to thank Christ University Vice Chancellor, **Dr. Rev. Fr. Thomas C Mathew**, Pro Vice Chancellor, **Dr. Rev. Fr. Abraham**, Director of Faculty of Engineering, **Fr. Benny Thomas** and the Associate Dean **Dr. Iven Jose** for their kind patronage.

I would like to express my sincere gratitude and appreciation to the Head of the Department, Department of Computer Science and Engineering, Faculty of Engineering, **Dr. Balachandran K.**, for giving me this opportunity to take up this project.

I am extremely grateful to my guide, **Chinthakunta Manjunath**, who has supported and helped to carry out the project. His constant monitoring and encouragement helped me keep up to the project schedule.

I am also extremely grateful to my co-guide, **Deepak Biradar**, who has supported and helped to carry out the project. His constant monitoring and encouragement helped me keep up to the project schedule.

I extend my sincere thanks to my family members and friends for their continuous support.

Declaration

I, hereby declare that the project titled “**Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining**” is a record of original project work undertaken by me for the award of the degree of **Bachelor of Technology in Computer Science and Engineering**. I have completed this study under the supervision of **Chinthakunta Manjunath**, Department of Computer Science and Engineering and **Deepak Biradar**, Effent Technology.

I also declare that this project report has not been submitted for the award of any degree, diploma, associate ship, fellowship or other title anywhere else. It has not been sent for any publication or presentation purpose.

Place: Faculty of Engineering, Christ University, Bengaluru

Date:

Name	Register Number	Signature
Ankit Tripathi	1317180	

Abstract

Sellers promoting products on the internet often need reviews from the customer side for the proper insight of their products like what is the current trend market is following. As e-commerce is rising and taking gigantic form day by day, the reviews received from customer about the items at the same time is also growing rapidly. The products which are sold in huge quantity will gain lot of reviews and popularity, which may or may not be in the positive sense. This creates huge confusion for the latent customer to as he being a newbie has to make a decision whether they should or should not buy the product. Problems also arise for the manufacturers of the product to keep record and manage customer opinions. Also extra problems are faced by the sellers because many other e-commerce sites may sell the same item at better ratings.

In this project, the main aim is to sum up all the customer reviews of a item and compare the products based on reviews at single place. This review task is different from traditional text encapsulation, because we only extract the information of that product on which the customers have given their opinions and whether the opinions are positive or negative. We do not sum up the reviews by selecting a rewrite of some of the original comment from the reviews to capture the main points as in the classic text summarization.

Contents

CERTIFICATE	i
BONAFIDE CERTIFICATE	ii
INDUSTRY CERTIFICATE	ii
ACKNOWLEDGEMENT	iv
DECLARATION	v
ABSTRACT	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi
GLOSSARY	xii
1 INTRODUCTION	1
1.1 Problem Formulation	1
1.2 Problem Identification	2
1.3 Problem Statement & Objectives	2
1.4 Existing System	2
1.5 Limitations	3
2 RESEARCH METHODOLOGY	4
3 LITERATURE SURVEY AND REVIEW	7
4 ACTUAL WORK	9
4.1 Architectural Strategies	9
4.1.1 Programming Language	9
4.1.2 Usage of Spring Framework and Industry Level Architecture . .	10
4.2 Data Flow Diagrams-Login Module	12
4.3 Collaborative Based Recommendations (user Based Rating based on Pearson Algorithms)	16
	vii

4.4	Objectives of Project	20
4.5	Testing	32
5	RESULTS, DISCUSSIONS AND CONCLUSIONS	33
5.1	SCREENSHOTS	33
5.2	Results & Analysis	41
5.3	Comparative Study	41
5.4	Cost Estimation	43
5.4.1	Cost In Terms Of Software And Hardware Requirement	45
5.4.2	Cost Estimation for sale of product	45
5.5	CONCLUSION	46
5.6	FUTURE WORK	47
	BIBLIOGRAPHY	48
A	ALGORITHMS	50
A.1	PEARSON RATING ALGORITHM	50

LIST OF FIGURES

4.1	Architecture Flow of Spring Framework	11
4.2	DFD Level 0	12
4.3	DFD Level 1	13
4.4	DFD Level 2	14
4.5	Methodology	15
4.6	Pearson Formula	16
4.7	Predicted Rating Formula	16
4.8	Item Based Rating	17
4.9	Review Based Recommendations	17
4.10	Content Based Recommendations	21
4.11	Review based Recommendations	22
4.12	Flowchart for Login Module	24
4.13	Flowchart for User Creation	25
4.14	Flowchart for Product Buying Creation	26
4.15	Flowchart for Content Based Module	27
4.16	Flowchart for Collaborative based Module	27
4.17	Noise Reduction Flowchart	28
4.18	Hybrid Recommendations	29
4.19	Sequence Diagram	30
4.20	Use Case Diagram	31
5.1	Registration	33
5.2	Login	34
5.3	Rating	34
5.4	Direct Rating Viewer	34
5.5	Rating Graph	35
5.6	Transaction	35
5.7	Settings	36
5.8	Review Submission	36
5.9	Reviews	37
5.10	Keywords	37
5.11	Keywords	38
5.12	Polarity Matrix	38
5.13	Positive Polarity	39
5.14	Negative Polarity	39

List of Figures

5.15 Neutral Polarity	40
5.16 Rating based on Pearson Algorithm	40
5.17 Average of Truthness and Fakeness in Review	41
5.18 Graphical Results	41
5.19 Comparative Results of all Attributes	42

LIST OF TABLES

4.1	Customer Table Description	18
5.1	Comparative Table of all Attributes	42

GLOSSARY

.

Item	Description
SQL	Structured Query Language
API	Application Programming Interfaces
CSV	Comma Separated Values
GUI	Graphical User Interface
PSO	Particle Swarm Optimization
IBCF	Item Based Collaborative Filtering
DBCF	Demographic Based Collaborative Filtering

Chapter 1

INTRODUCTION

This project basically scraps the data from several e-commerce websites and collects the reviews from them and identifies fake reviews and removes them. It also has a concept of ip address policy wherein only authorised customers are allowed to write reviews about products.

1.1 Problem Formulation

We know that posting fake reviews is a bad thing— especially if one is posting fake reviews in favor of his/her business. But what if one's business is affected by fake reviews posted by competitors and others in the market who have a negative intent towards his/her business? They will post negative reviews which may or may not be true, but if it is not true it hampers one's reputation in market even if they have a good position in market or any field. The best part is it is possible to get away with fake reviews. It's really important to address those fake reviews and remove them as soon as possible, this helps in building great reputation in the market and also prevents customers from getting misguided about any of the product. This has become a great problem for e-commerce world and should be taken care of. This is the problem we have identified and will be implementing a solution to this.

1.2 Problem Identification

Our aim is to develop an smart and intelligent system which automatically mines opinions and classify them into fake and real category. Our proposed idea is that opinion based spam review analyzer will automatically classify user opinions into fake or real reviews. This system can be useful to e-commerce organization as well as to customers. Business organizations can keep a track of their product selling by analyzing the trend and understand what is the feedback of customers about the products. Using this customers would be able to decide whether he/she should or should not buy the product. This would be helpful for people to purchase valuable products and spend their money on quality products.

1.3 Problem Statement & Objectives

The objectives of our projects are as follows:

- 1) Creating a simple and efficient user interface for easy identification and navigating through the pages.
- 2) Recommending relevant products to users by analyzing their profile and shopping history.
- 3) Showing relevant statistics to the user.
- 4) Adding security features so that the page is free from sql injection and other hacking activities.
- 5) Recommending top rated products in their ranking order.
- 6) Including ratings of trusted and certified users.
- 7) Adding admin module so that he can add or remove items and monitor reviews.

1.4 Existing System

Before going for an online shopping almost all the users research about the products they are going to buy in one or other way like reading customer reviews about the same product. A report of survey conducted by Yelp.com has shown that: More than 80 percent of users and shoppers depend on reviews given by the people, 50 percent depend on ratings of the online products, 30 percent of the users compare the product's reviews

with others product's reviews. Customers really consider the feedbacks given by other users and also the e-commerce companies do the same. Blogs, websites, discussion boards etc. are a very important platform for the users to comment about the products and also these become a very important source of information in the form of textual data. Therefore, today's customers believe a great deal in online reviews, which refers that people do a lot of analysis and research before buying a particular product.

It's a fact that after studying reviews or opinions for so long if the user gets a better idea about the product he might end up buying it. Normally it all depends on how much positive and good reviews are given for that particular product which makes a favourable condition for users to buy it. Similarly if the overall review is negative the customers would not buy the product. Now the situation might arise wherein the customers or sellers may give fake reviews about a particular product to increase its popularity so that its rating goes up. This is called spam reviews.

1.5 Limitations

The limitations of the project are:

1). Lack of time

The time duration was really less for the proper implementation of concepts.

2). Lack of data

Data collection was a real challenge from E-commerce websites.

3). Changing data

Continuous change of data was the real challenge.

4). Changing user preference

Changing user requirement is the real challenge.

5). Unpredictable items

Continuous change in market trend poses a great problem.

6). Complexity of the algorithm

Use of complex algorithm for calculation increases turnaround time.

Chapter 2

RESEARCH METHODOLOGY

Research about requirements will be ordered as: learning, investigation and implementation. Analysts might require online access to illustrious and periodic amazon information. Especially the chief sources conduct efficient research; Stanford amazon data access to far reaching unforgettable learning sets and collectively period access to sources. Potentially with a time delay, similar to Thomson Reuters and Bloomberg cash learning.

News learning

Access to memorable information and period news information sets, most likely through the development of 'instructive information licenses (cf. programming framework permit).

Open information

Access to scratched and chronicled key open information-, available through Reviewer Id, sites or remarks.

Programmable interfaces

Scientists conjointly might want access to direct application programming interfaces (APIs) to rub and store elective available information sources that may not be mechanically gathered.

As of now, amazon learning is normally either available through clear broad schedules or need the agent to program their examination in an extreme dialect like Python. Its specified higher than, analysts require:

Investigation dashboards

Non-programming interfaces are required for, giving what could be termed as -"profound"- access to "crude" learning. When all is said is done. Access to huge wellsprings of amazon learning is regularly confined and full business. access is costly.

Siloed learning

Most information sources have intrinsically disengaged information making it troublesome to consolidate with option information sources. Investigative instruments gave by merchants range unit normally fixing to 11 single information set. Maybe limited in expository capacity. and information charges make them costly to utilize.

There region unit partner degree expanding scope of capable business stages, similar to those prepared by Stanford University, however the costs range unit generally take precaution for instructive research. Either similar offices should be given by national science establishments or sellers might want to be influenced to present the build of partner degree 'instructive permit.'

Data Scraping

Data scraping is a technique in which a computer program extracts data from human readable output coming from another program.

Data Cleaning and Processing

Data Cleaning and Processing is the most important word in data mining experiments to eliminate irrelevant words ,determination of missing values and supporting data reduction by sampling and transaction elimination methods.

CSV files

A CSV is a comma seperated values file , which allows data to be stored in a table structured format.It is a type of spreadsheet but with a .CSV extension(it is a text tile containing information separated by commas) .

Database

Database is which allows the data to he stored in tabular forms which is extracted from .csv file which can be useful for the future purposes to store,define,delete and modify the data.

Visualization

It helps in visualisation of the desired output for an user to understand the execution of the work in simple GUI which is in the form of a graphs to analyse sales quarterly wise and have explained step wise.

Chapter 3

LITERATURE SURVEY AND REVIEW

Opinion mining has continuously been a hot topic and it's managed to draw in a larger deal of attention. However, the restricted work has been done on opinion spam (fake reviews). The matter is correspondent to spam within the net search[3]. Review spam is troublesome to discover, if not possible, acknowledge faux reviews by manually reading them. Therefore to search out a restricted drawback, for distinctive uncommon review patterns which might be suspicious behaviours of reviewers, we tend to develop the matter on finding out of the blue rules. The technique is domain freelance, victimization of the technique, Amazon.com review dataset is analysed and several sudden rules and rule teams which might indicate spam activities become active suddenly. Shoppers rate, review and analyse product on-line[8]. Therefore, websites of client reviews have become targets to opinion spam, whereas recent work have centered totally on manually specifiable instances of opinion spam during this work therefore on studying unreal opinion spam unreal opinions that are by design written within the sound trusted[10]. Group action work from science and procedure science, to develop and compare 3 approaches to search out deceptive opinion spam, and ultimately develop classifier that is nearly correct on our gold-standard opinion spam dataset. It supports these feature analysis of our learned models as well as created many theoretical contributions, relationship between deceptive opinions or ingenious writing. To spot and construct three-dimensional statistic that is supported mixture statistics, so as to show and mine correlations[1]. The formula additionally pinpoints such windows in several time resolutions that facilitates quicker human examination. Therefore to get the singleton review could be a big supply to spam reviews and for the most part affects the ratings of

on-line stores. Nowadays sizable amount of the merchandise reviews area unit denote to the net[4] . Such review area units are vital to customers,users and to firms. Customers use the reviews for deciding the standard of the merchandise to shop for.Companies and vendors use opinions to choose products online and enhance the sales in step with recent things done from alternative competitors[9].

All the reviews given by customers or users don't seem to be true reviews. These reviews are given to push or to kick down the merchandise. Some reviews are given on whole of product,as an attempt in the promotion of another product. There is need to ascertain what percentage of reviews are spam or non spam. Here this technique is employed for policing dishonest spam reviews, victimizing on n-gram language model and reviews for whole spam detection feature. Given system identifies spam and joins the result that shows spam and non spam reviews. For grading these there's got to live the degree of the spam for every reviewer ,And apply them on an Amazon review dataset.Then to pick out a set of extremely cautious reviewers for additional scrutiny by our user evaluators with the assistance of the net primarily based transmitter analysis package specially developed for user analysis experiments[7].Results show that projected ranking and supervised ways are effective in discovering spammers and crush alternative baseline methodology that are unit supported. Finally here we tend to show that the detected spammers have a lot of vital impact on ratings compared with these unhelpful reviewers.

Chapter 4

ACTUAL WORK

Designing is one of the key phases of software development. The design is a process in which a system organization is established that will fulfill the functional and non-functional system requirements. Large Systems are always disintegrated into sub-systems that provide required functionalities. The output of the this design process elucidates the architecture of the system.

4.1 Architectural Strategies

This section describes the design and strategies that influence the overall organization of the system and its approach. These policies will offer comprehension into the key opinions and devices used in the system architecture.

4.1.1 Programming Language

Java is used as a programming language for the development of this project. Java has the following advantages:

- Java is naive, informal , safe and strong Object oriented language.
- Java is a platform independent language and the program is deduced to byte code.
- Java is spread language which helps to develop the project in sectional fashion.

- Multithreaded programming has been smoothly integrated into java language, while

In other languages, operating system-specific procedures have to be called in order to enable multithreading.

- Java is multithreaded. Multithreading is the ability of a program to perform multiple tasks or several functions concurrently within a program.

4.1.2 Usage of Spring Framework and Industry Level Architecture

The architecture used in this project is as shown below. The early phase of designing process of identifying these sub-systems and making a framework for sub-system control and communication is called architecture design and the output of this process is a portrayal of the architecture. The architectural design process aims at establishing a basic physical framework for the system. It includes the classification of major components of the system and communications between these components.

The figure 4.1 shown below defines the system architecture which is the present trend in the industries for the growth of any routing software. The figure 4.1 shows that the UI is designed in the HTML/JSP pages and then the request is sent to the web container after which web container approves the request in the web.xml file by viewing first into the url prototype and then only forwards it to the servlet name and then searches for the alike servlet name in the servlet tag and looks into the servlet class and creates an object of Action Servlet and then the action servlet will allocate its job to Request Processor.

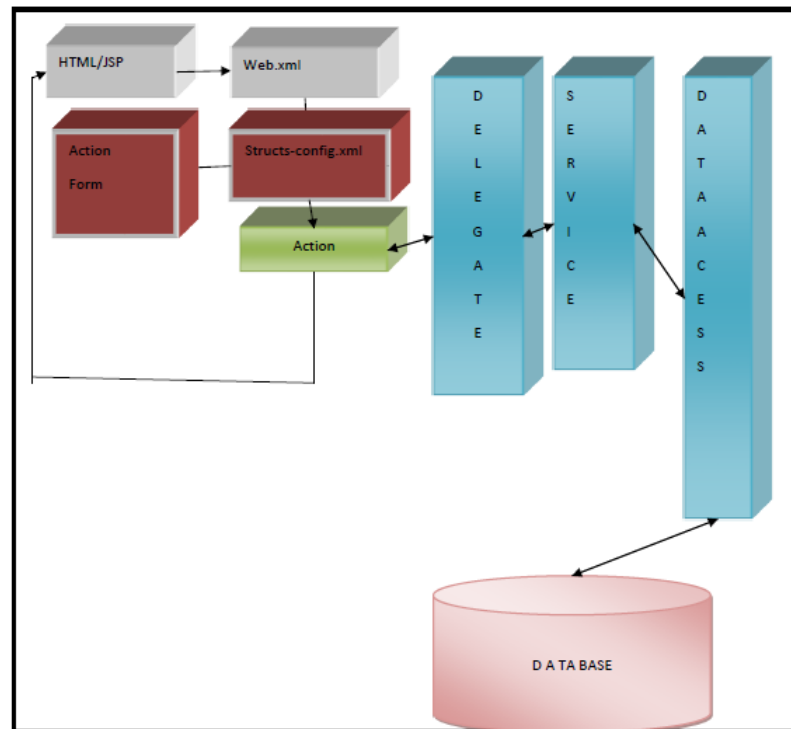


FIGURE 4.1: Architecture Flow of Spring Framework

The action class calls the delegate , then the delegate calls the service and service calls the Data Access layer and results go precisely in the reverse way and the subsequent JSP page is loaded

Action Form

This is the Plain Old Java Object which will have the getters and setters, setters get routinely called and data the user has provided will be reflected.

Action

This class is responsible for retrieving the data entered by users and then analyses it and calls the delegate layer and gets the results.

Delegate

Delegate layer comprises nothing but it's main job is to call an suitable service.

Service

This layer is accountable for the algorithm related implementation. Future service would need the help of Data Access Layer for few operations and many other classes.

Data Access Layer

This layer is accountable for the algorithm related implementation. Future service would need the help of Data Access Layer for few operations and many other classes.

Database

This is the place where all the routing tables and performance measurement tables would have been placed.

4.2 Data Flow Diagrams-Login Module

Data Flow Diagram (DFD) is a pictorial demonstration of the "flow" of data through an information system. It shows how data flows through a sequence of dispensation steps. The data is changed at each step before going to the next level. These allowance steps or alterations are program functions when Data Flow diagrams are used to make a software design. DFD diagram mostly contains four elements; process, data flow, external object and data supply.

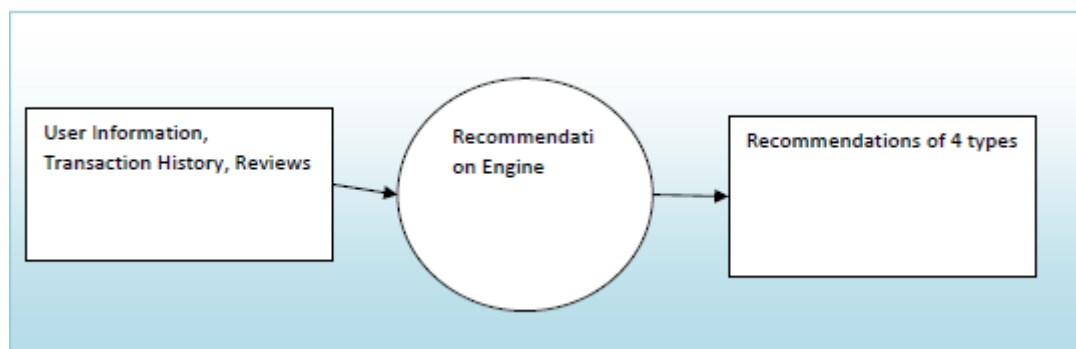


FIGURE 4.2: DFD Level 0

Fig. 4.2 shows DFD Level 0 shows the input namely User Information, Transaction History, Reviews. The process is the Recommendation Engine. The output is list of 4 types of recommendations namely Content Based Recommendations, Collaborative Recommendations, Association Rule Mining and Hybrid Recommendations.

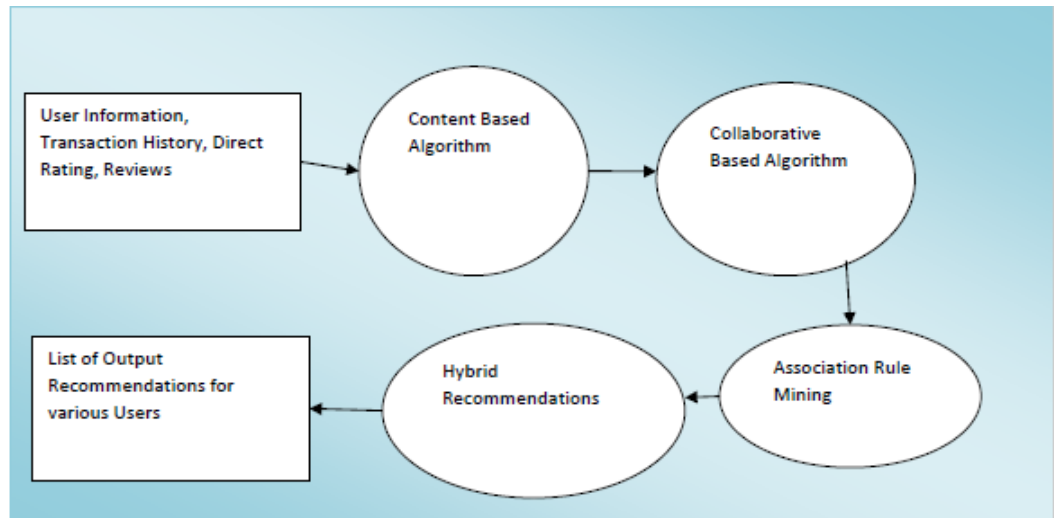


FIGURE 4.3: DFD Level 1

Fig. 4.3 shows DFD Level 1 shows the input namely User Information, Transaction History, Direct Ratings and Reviews. The process is the Content Based Algorithm, Collaborative Based Recommendations, and Association Rule Mining and Hybrid Recommendations. The output is list of 4 types of recommendations namely Content Based Recommendations, Collaborative Recommendations, Association Rule Mining and Hybrid Recommendations.

The fig. 4.4 given below shows the information related to customer like transaction history, pattern of buying, amount in account etc. It also mines the data and checks certain constraints based on which it provides collaborative based ranking of products. It also shows that reviews are collected and data is cleaned based on which negative, positive and neutral polarity is computed.

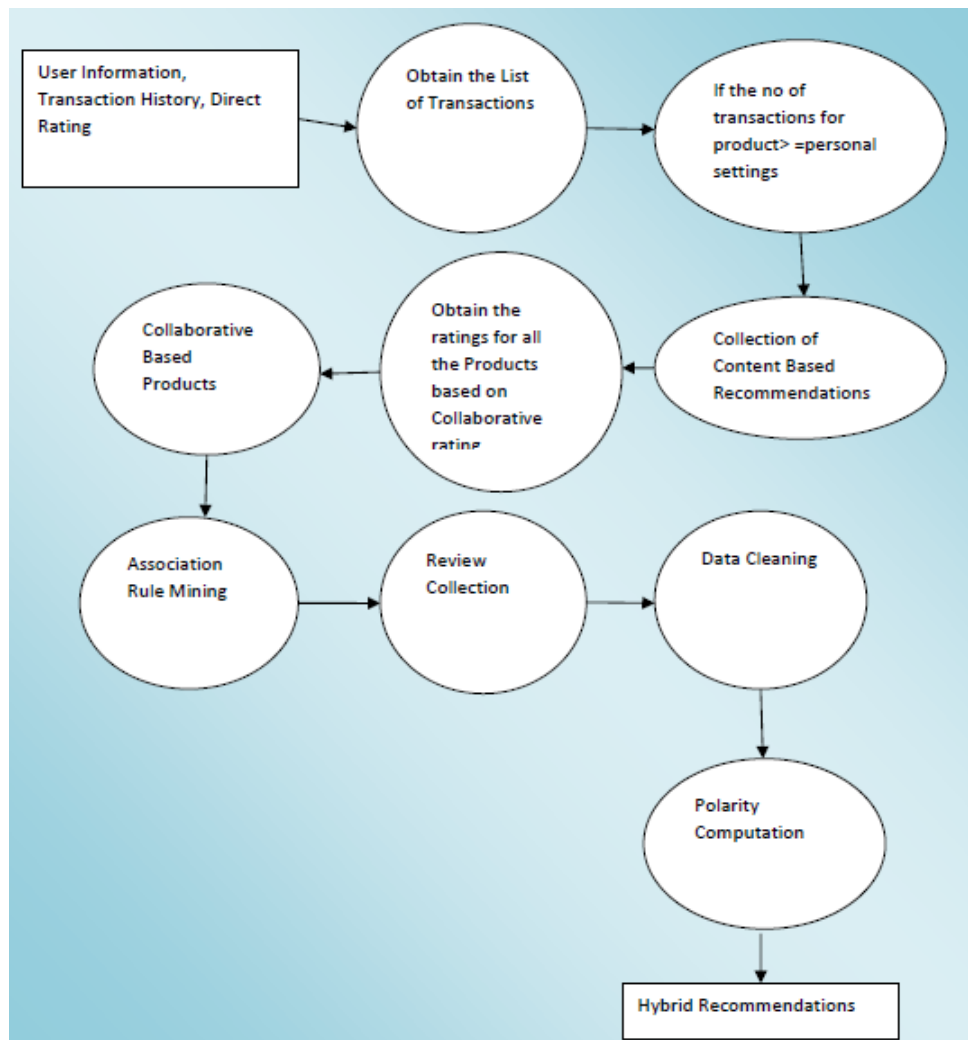


FIGURE 4.4: DFD Level 2

Methodology:

The Methodology can be described by the following diagram

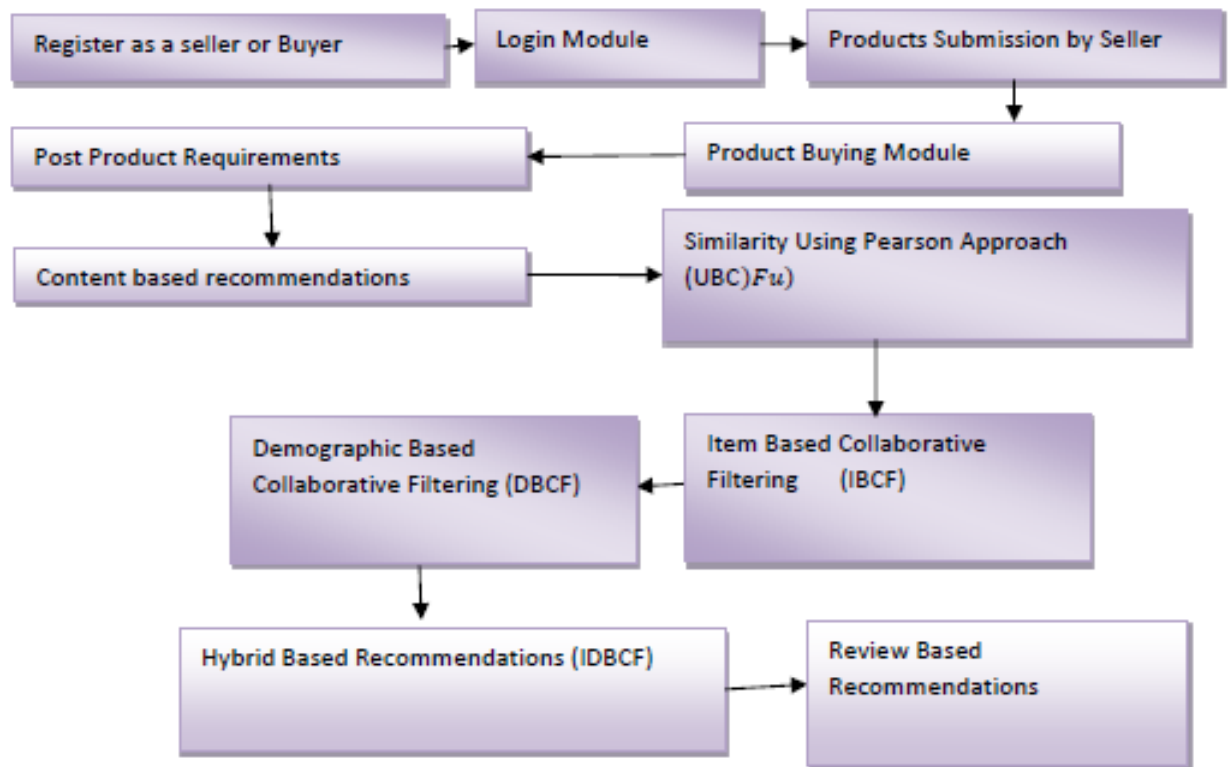


FIGURE 4.5: Methodology

Now this fig. 4.5 shows the different modules of the project to name few of them login module is the heart of this project which allows different users to login and submit reviews also it facilitates for admin to allow or disallow certain reviews, to add or remove products etc. it also shows demographic based rating.

4.3 Collaborative Based Recommendations (user Based Rating based on Pearson Algorithms)

This Module is responsible for gaining the rating across the registered as well as un-registered users. The recommendations are generated based on the aggregated rating across the users of the application. Finally the recommendations are generated based on the pearson formula

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

FIGURE 4.6: Pearson Formula

The predicted rating is given by the formula

$$P_{aj} = \bar{r}_a + \frac{\sum_{i \in NS_a} sim(a, i) * (r_{ij} - \bar{r}_i)}{\sum_{i \in NS_a} |sim(a, i)|}$$

FIGURE 4.7: Predicted Rating Formula

Item Based Rating

It determines the rating based on items purchased by various users

$$p(u, i) = \frac{\sum_{j \in S} s(i, j) r_{u, j}}{\sum_{j \in S} |s(i, j)|}$$

FIGURE 4.8: Item Based Rating

Demographic Based Rating

The users are grouped according to their demographic status and then items are rated as per demography

Hybrid based Recommendations

This Module is responsible for doing the intersection between the item based and demographic based recommendations

Review based recommendations

This Module is responsible for collecting the reviews for the products, cleaning of the reviews, the amount of positive polarity, negative polarity and neutral polarity and recommend products which has maximum positive polarity, neutral polarity and lowest negative polarity.

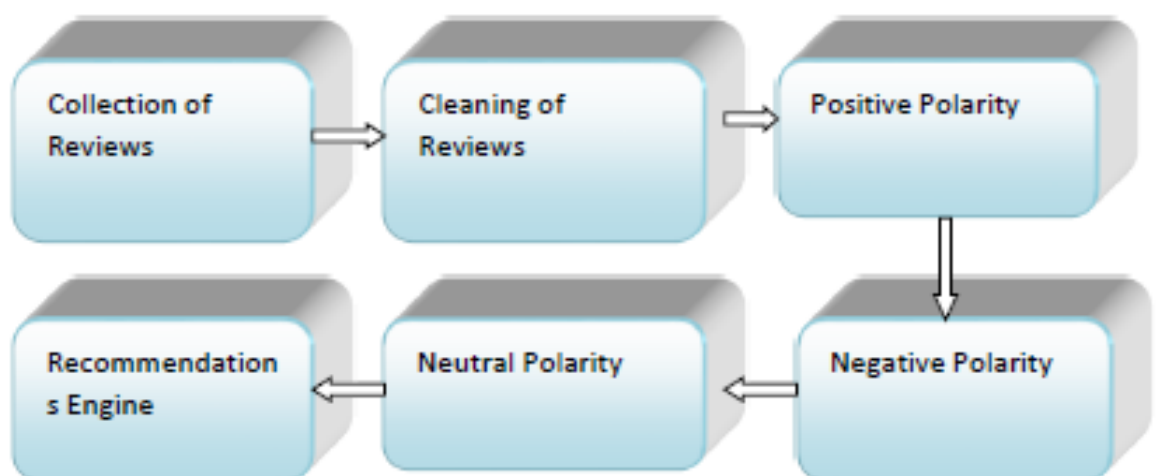


FIGURE 4.9: Review Based Recommendations

TABLE 4.1: Customer Table Description

FieldName	Description
Login ID	Customer ID
Name	Customer Name
Password	Customer Password
Email ID	Customer Email ID
Mobile Number	Customer Mobile Number

The Product Shopping website will make use of the following tables:

a)Customer Detail Table:

This table consists of basic customer informations like name ,id,passwords,email etc.

b. Product Detail Table:

This table contains all the information about the products like quantities sold,price,popularity etc.

c. Category Table:

This table stores the categoryId and the matching product's category name.

d. Order Information Tables:

These tables are used to stock data about the orders placedby customers.It stores the orderId, ProductId, quantity etc.

Instructions to User:

Directions to User:

1. Only those users who are registered members will be allowed to post reviews.
2. They can find out the different types of the product that the buyer has bought formerly from the buyer's profile.
3. They can find out the subcategory of the product if there is any as they are registered users.
4. They can perform gratified based filtering in different category or subcategory of products found in step1 and 2, the products that are much similar to the products that the buyer has bought earlier based on the feedbacks from the buyers and past history records can be easily identified.
5. As a result of step 3 products are arranged in the descending order of recommendations. In this step system actually assess the worth of the recommended products based on the rating given to those products by the other buyers.
6. From the product transaction database can find all the transactions whose category and sub category is same as found in step1 and step2.
7. Step 6 is the outcome of all which is the final recommendation for the buyer. All these functionalities are performed when the buyer is logged out and the results are stored in the buyers web profile. When the buyer comes back online next time the references will be produced spontaneously.

4.4 Objectives of Project

1. Design and Development of Login Module for the customer in order to login into the application by providing the valid username and password.
2. Design and Development of Sign Up Module for the customer where the customer would be entering login id, password, Account Number and Internet Pin. If the customer already exists it would not allow to create.
3. Design and Development of Product Buying Module by customer where customer would be capable of buying the product.
4. Design and Development of Content Based Recommendations algorithms for product recommendations.
5. Design and Development of Collaborative Based Recommendations based on Pearson formula algorithms for product recommendations.
6. Design and Development of Hybrid Recommendations which provides products which have maximum positive polarity, maximum neutral polarity and minimum negative polarity.

Register as Seller or Buyer

This module is responsible for any user to register in the application by entering the personal details like Username, Password, Email, Phone No, Type (Buyer and Seller) .

Login Module

This module is responsible for any user to perform login into the application.

Product Submission by Seller

This Module is responsible for submitting of the products by giving details like Product Name, Product Description, Product Cost, Category (Mobile, Book, Tablet, Laptop, and Electronics)

Product Buying Module

This Module is responsible for purchase of products by entering the credit card number and expiration date.

Content Based Recommendations

This Module analyzes the transaction history of the buyers through the transaction logs. A user customized threshold is set. If any of the products exceeds the specific threshold then recommendations of the product are suggested.

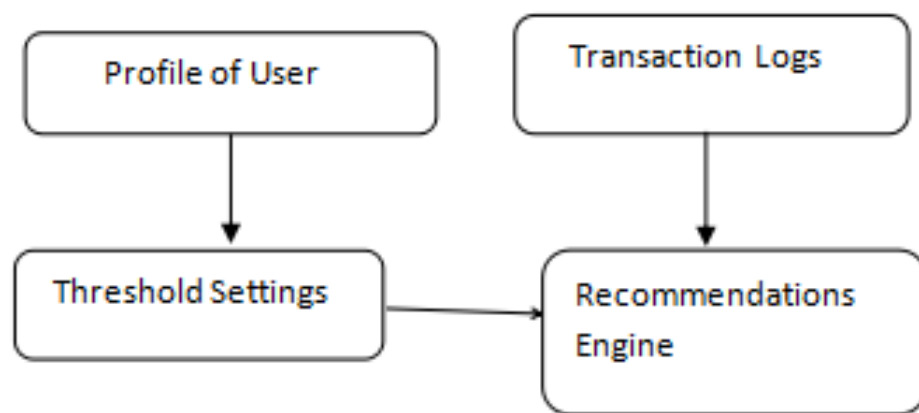


FIGURE 4.10: Content Based Recommendations

Above Fig. 4.12 shows that the profile of the user and his/her threshold settings along with transaction logs acts as an input for recommendations engine and recommendation engine filters the transaction based on threshold criteria.

Collaborative Based Recommendations

This Module is responsible for gaining the rating across the registered as well as un-registered users. The recommendations are generated based on the aggregated rating across the users of the application.

Hybrid based Recommendations

This Module is responsible for doing the intersection between the content based recommendations and collaborative based recommendations.

Review based recommendations

This Module is responsible for collecting the reviews for the products, cleaning of the reviews, the amount of positive polarity, negative polarity and neutral polarity and recommend products which has maximum positive polarity, neutral polarity and lowest negative polarity.

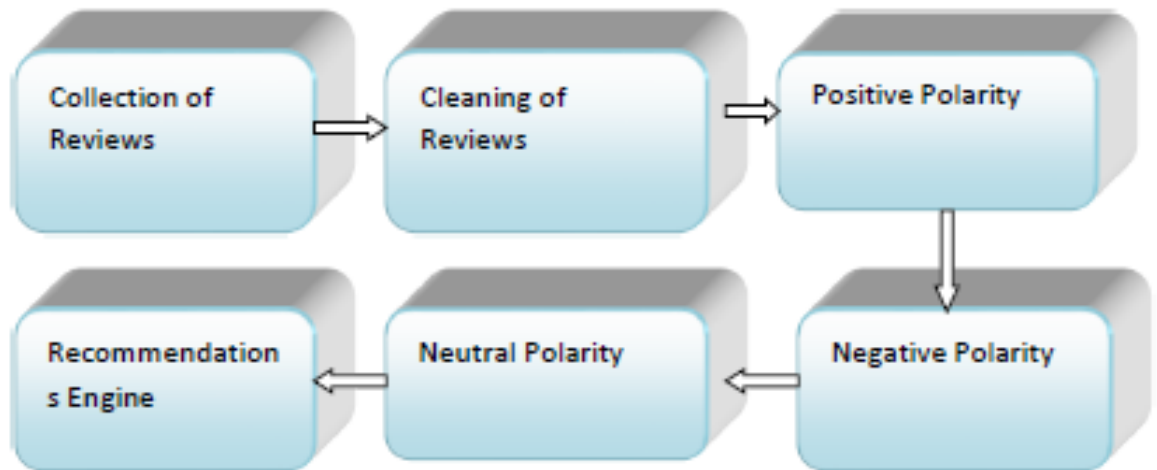


FIGURE 4.11: Review based Recommendations

Collection of Reviews

This Module is responsible for collecting the reviews from the registered users. This module will collect reviews based on actions that are required on the products.

Cleaning of Reviews

This Module is responsible for cleaning of each of the reviews by removing unused or unwanted words.

Positive Polarity Computation

This is a process in which all the reviews are disjointed by a delimiter. Each of the reviews given by the users are analysed and compared against the positive keywords stored in the database after which their frequency is calculated and it is fed to the algorithm and polarity is computed.

Negative Polarity Computation

This is a process in which each of the reviews are detached by a delimiter. . Each of the reviews given by the users are analysed and compared against the negative keywords stored in the database after which their frequency is calculated and it is fed to the algorithm and polarity is computed.

Neutral Polarity Computation

In this process each of the reviews provided by the users are allocated a neutral review if the positive and negative keywords both are not present.

Recommendations Engine

In this module the products are ranked based on their polarity value,like products with positive polarity are ranked first followed by neutral and negative polarity products.

Detailed Design

Web usage data scraping is one of the most important field of web mining. A lot of research is being done in this field which indicates the prominence of web usage mining towards search engines.Speed and precision are the most necessary features.

Flowchart for Login Module

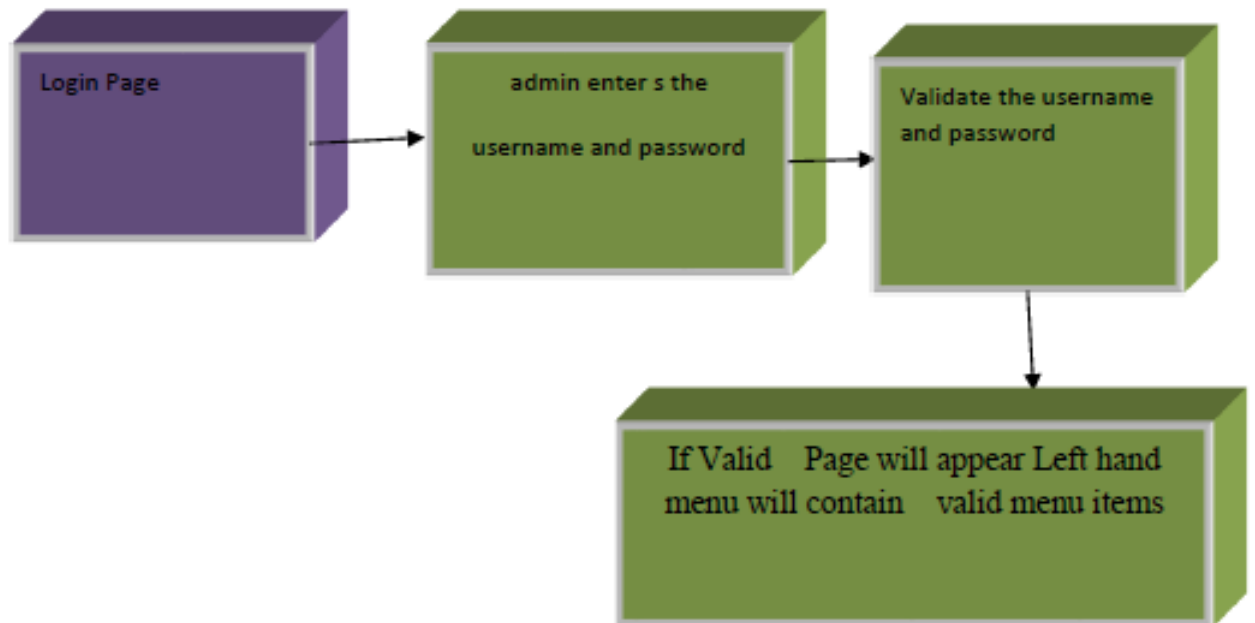


FIGURE 4.12: Flowchart for Login Module

Input- Username, Password and Login Type Output- Validation of Username and Password is done and page is appropriately routed The figure shows the login module for the admin/user. If the admin/user enters the valid username and password then the he/she would be able to login otherwise not.

Flowchart for User creation

Input- Username, Password

Output- Validation of data is done and page is appropriately routed

The below figure shows the User Creation module for customer.

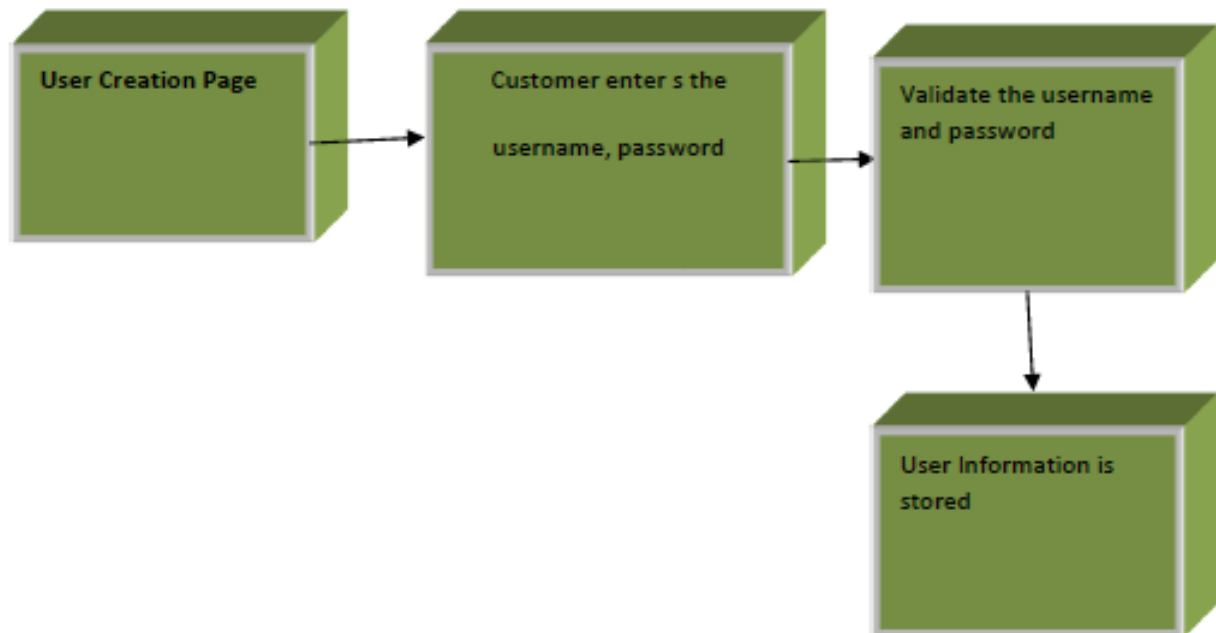


FIGURE 4.13: Flowchart for User Creation

Flowchart for Product Buying Creation

Input- Product Selected by the User

Logic- The Product ID of the product selected is send and cost of the product is retrieved. Bank account of the user is verified and if the account balance is above the product cost transaction is successful.

Output- Transaction Complete

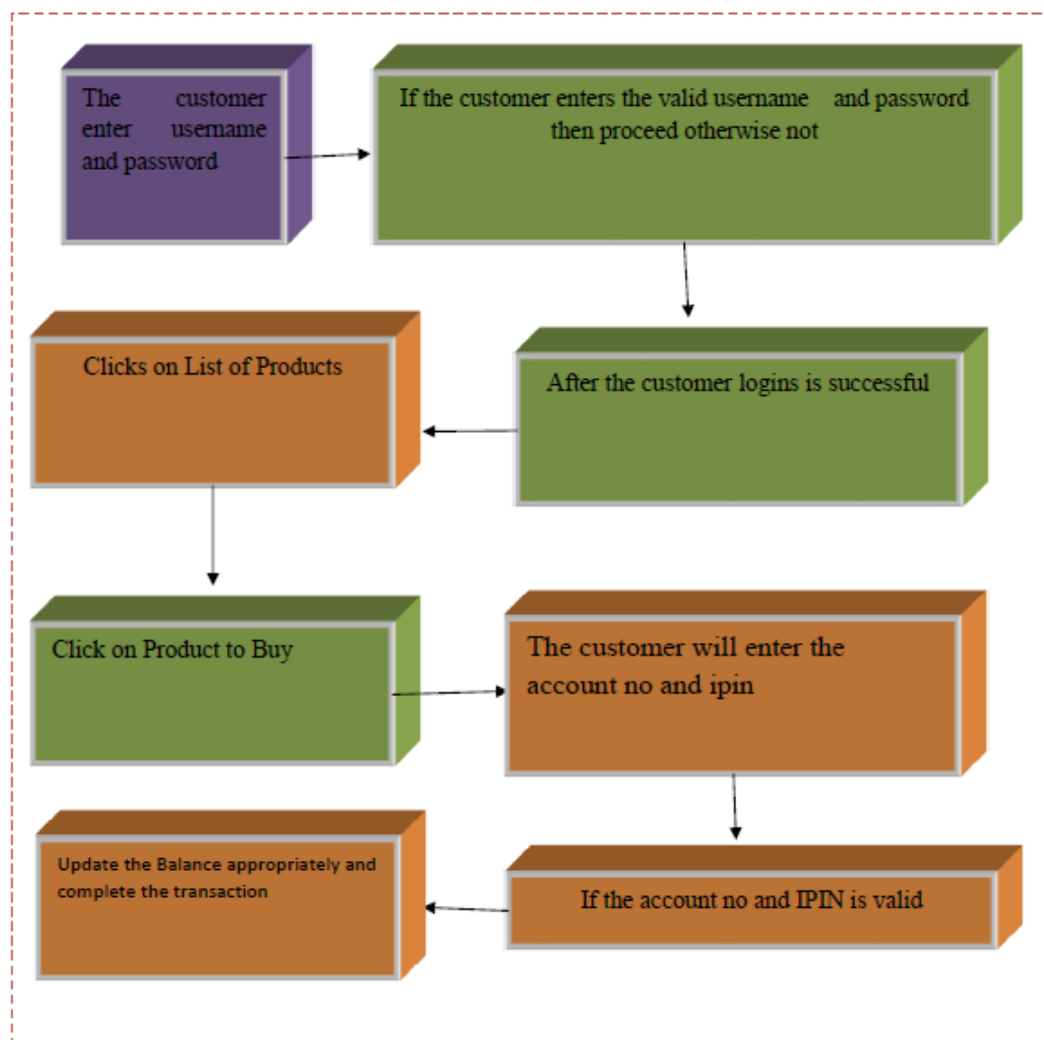


FIGURE 4.14: Flowchart for Product Buying Creation

Flowchart for Content Based Module

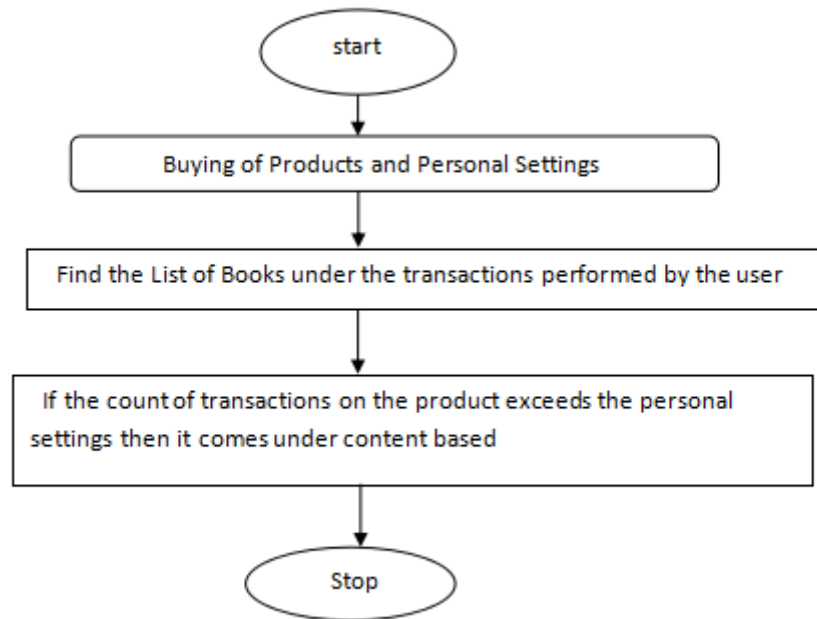


FIGURE 4.15: Flowchart for Content Based Module

Flowchart for Collaborative based Module

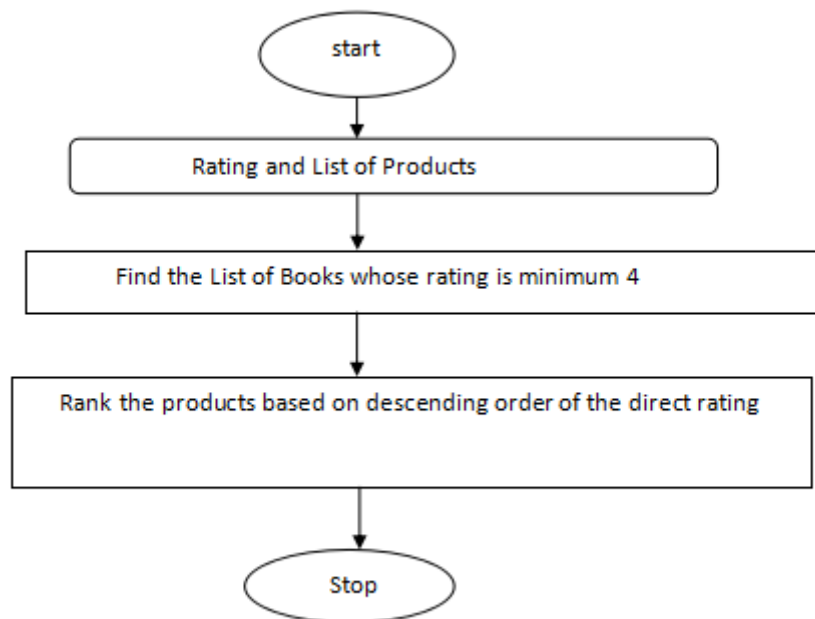


FIGURE 4.16: Flowchart for Collaborative based Module

Noise Reduction

This process is responsible for removal of noise the stop words that are present in the given tweet.

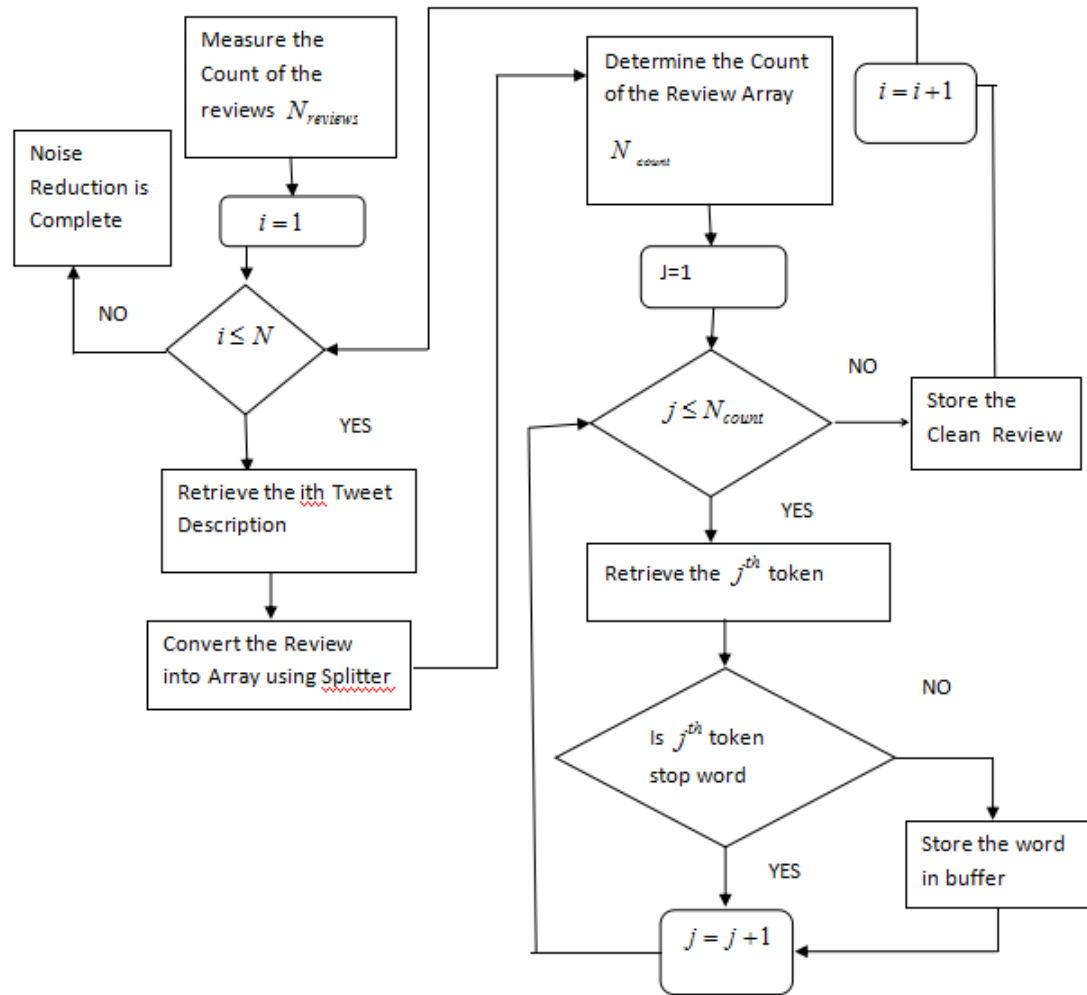


FIGURE 4.17: Noise Reduction Flowchart

Fig shows the Noise Reduction process

- 1) Determine the count of the number of reviews.
- 2) For each of the review the cleaning is performed and the stop words are removed and one can obtain the clean review.

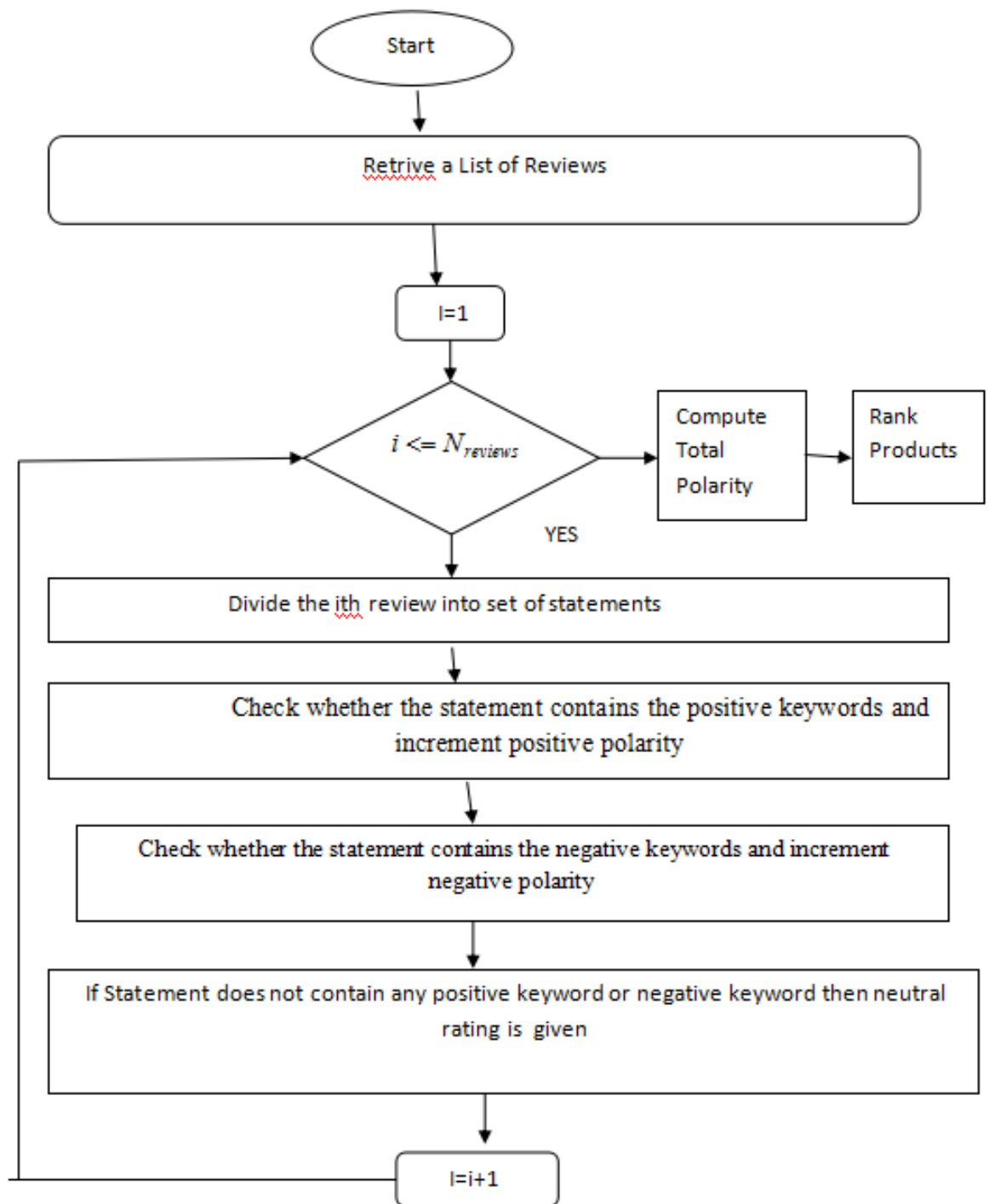


FIGURE 4.18: Hybrid Recommendations

Implementation

The following chapter gives information about the implementation details of the project in terms of sequence diagram and class diagram and database diagram

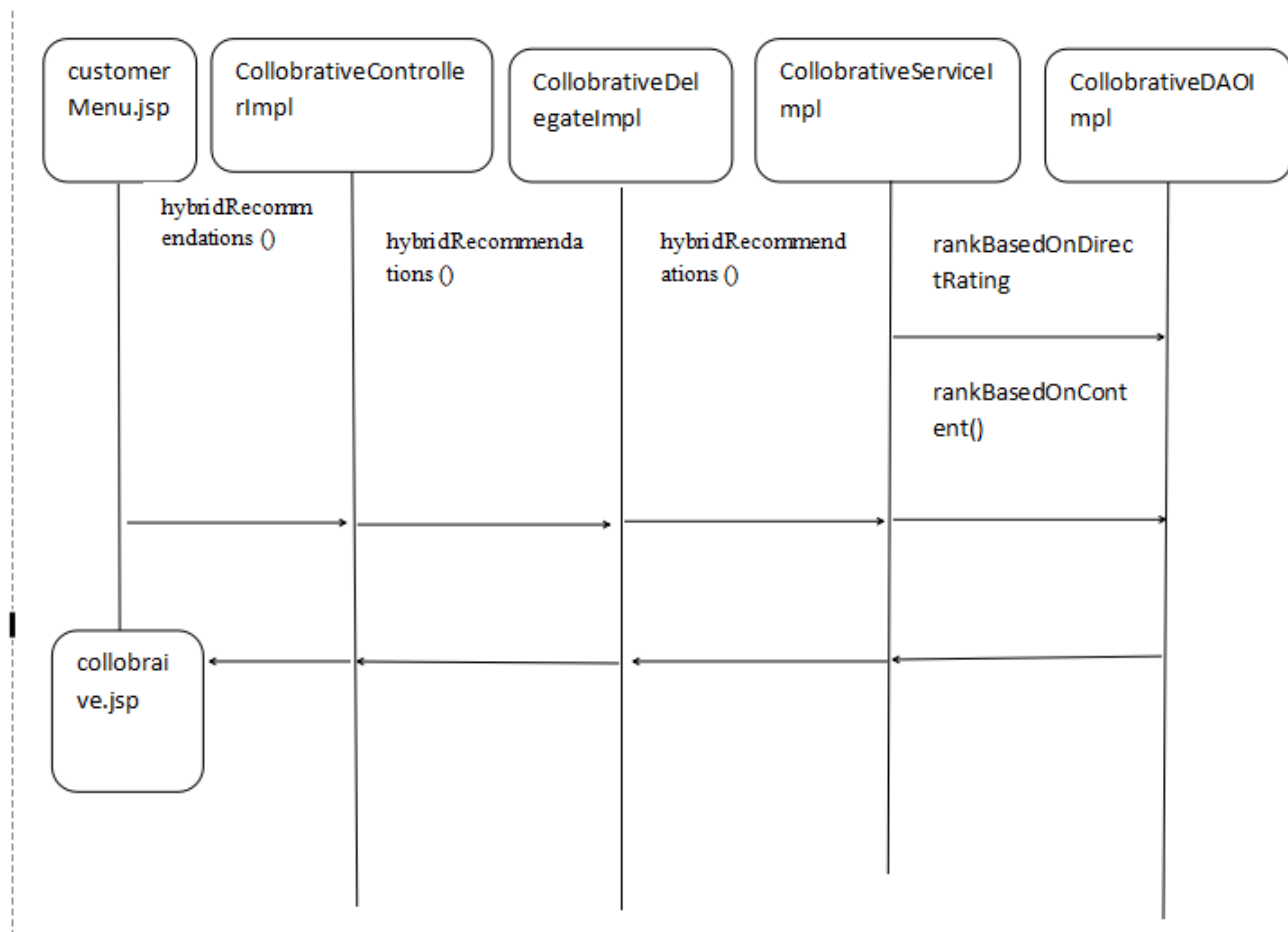


FIGURE 4.19: Sequence Diagram

Use Case Diagrams:

This is the function which is used to rank the products based on personal settings and the list of transactions for the user. If the number of transactions exceeds personal settings then add the List of Content Based Filtering.

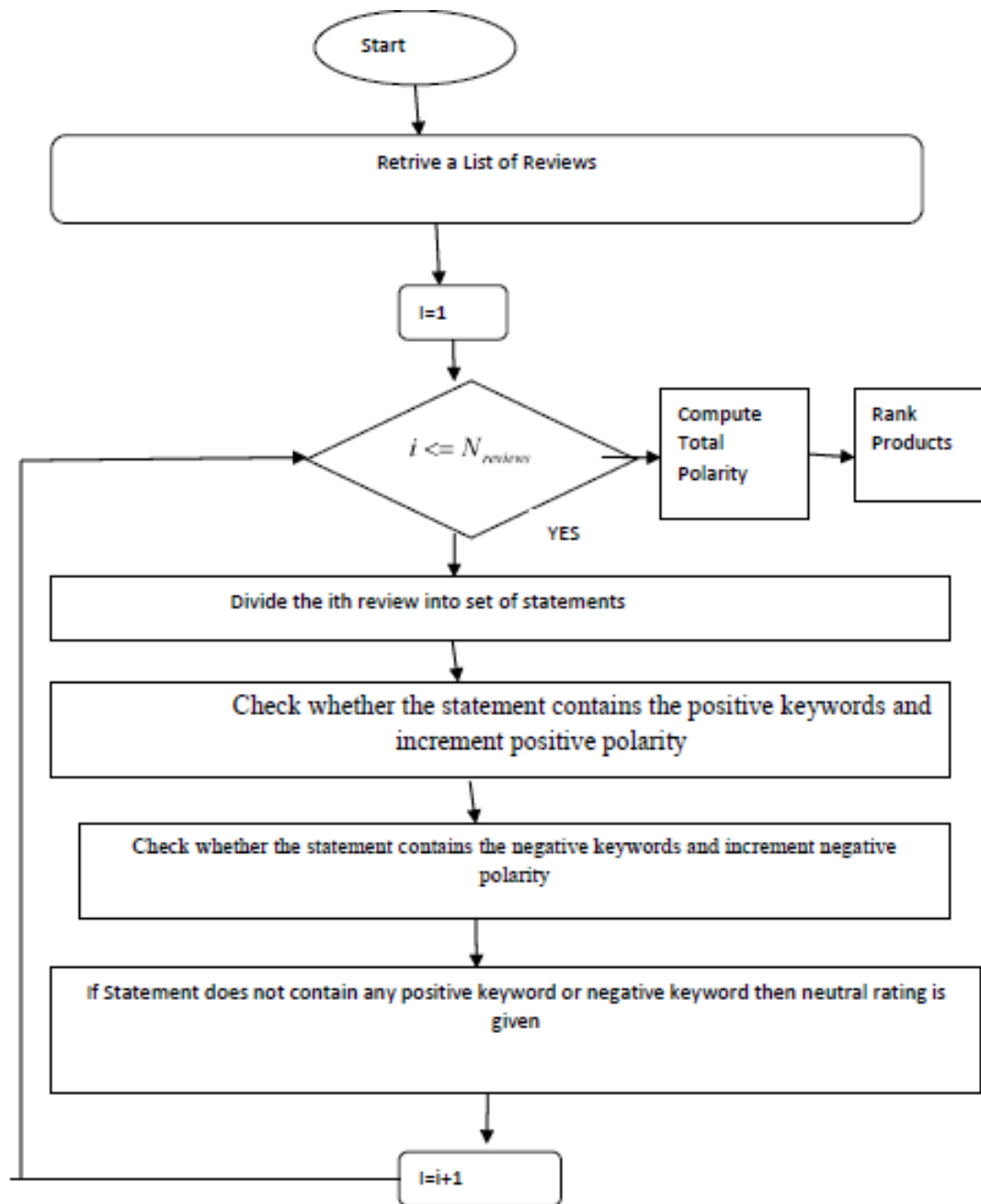


FIGURE 4.20: Use Case Diagram

4.5 Testing

Unit test:

Each report must be tested. There should be someone else other than the developer who should go through the report bit by bit, verifying and testing calculations, report structure, user contributions, and outcomes. If achievable, compare the results from other, autonomous source from unfamiliar to the current DW/BI system.

System test:

Once the report is believed to work on a separate basis, it must be observed how it works with the current system. Based on your front-end tool and the complication of your repository environment, there can be a large number of elements that need to be verified and validated, with time or event-based planning, delivery, and failure warning procedures. Support a large user group with these reports, carry out some stress testing as well. Including enough time in one's project and planning to open it for the rest of large user communities

User test:

If the users are not yet familiar with reports, or one would like to get feedbacks from non-technical users, this could be achieved by including a task in your approach to give a opportunity to users to inspect and support them. This may be a practical session, or it may take the form of a sample to a group of users with time for questions and answers. Or, it can be somewhat users do from their end with a simple web review form or email reply system.

Chapter 5

RESULTS, DISCUSSIONS AND CONCLUSIONS

5.1 SCREENSHOTS

Registration Module UI

A screenshot of a web registration form. The form has a light blue background and a white border. It contains five text input fields for 'First Name', 'Last Name', 'User Name', 'Password', and 'Email ID', each preceded by a label in blue text. Below these are three dropdown menus for 'Country', 'State', and 'City'. At the bottom left is a red 'Register' button.

Enter the First Name:

Enter the Last Name:

Enter the Desired User Name:

Enter the Password:

Enter the Email ID:

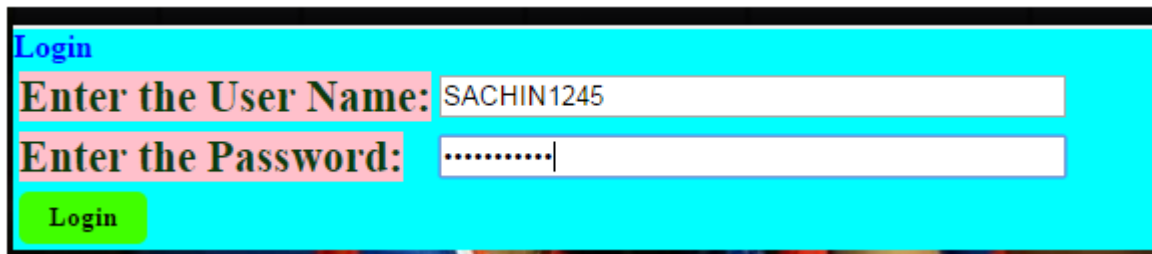
Select Country ▼

Select State ▼ Select City ▼

Register

FIGURE 5.1: Registration

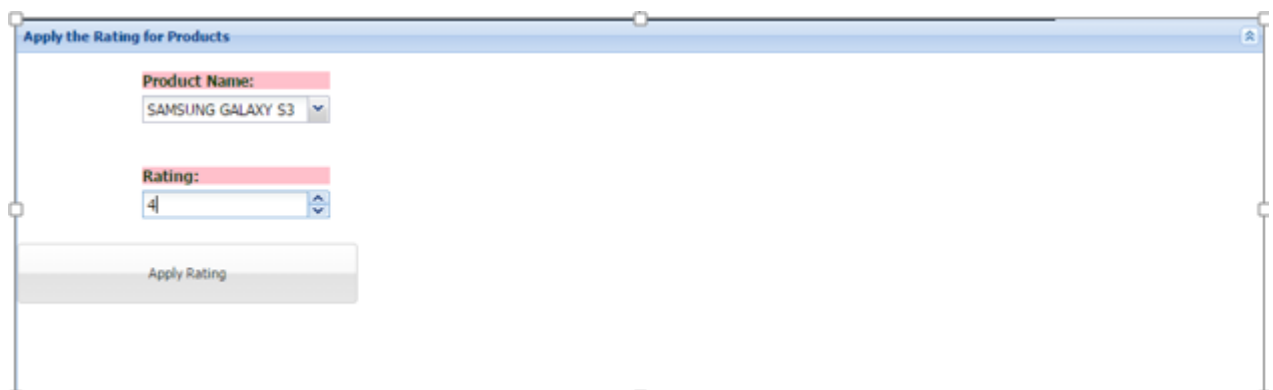
Login Module



A login form with a blue header bar containing the word "Login" in blue text. Below the header, there are two input fields. The first is labeled "Enter the User Name:" in red text, and the second is labeled "Enter the Password:" in red text. The first input field contains the text "SACHIN1245". The second input field contains a series of dots. Below the input fields is a green button labeled "Login" in black text.

FIGURE 5.2: Login

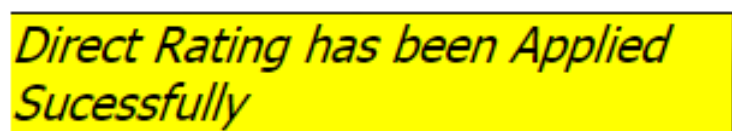
Direct Rating of Products



A form titled "Apply the Rating for Products" in a blue header bar. Below the header, there are two input fields. The first is labeled "Product Name:" in red text, and the second is labeled "Rating:" in red text. The first input field contains the text "SAMSUNG GALAXY S3". The second input field contains the number "4". Below the input fields is a button labeled "Apply Rating".

FIGURE 5.3: Rating

View Direct Rating



A yellow rectangular box with a black border containing the text "Direct Rating has been Applied Successfully" in a black, italicized font.

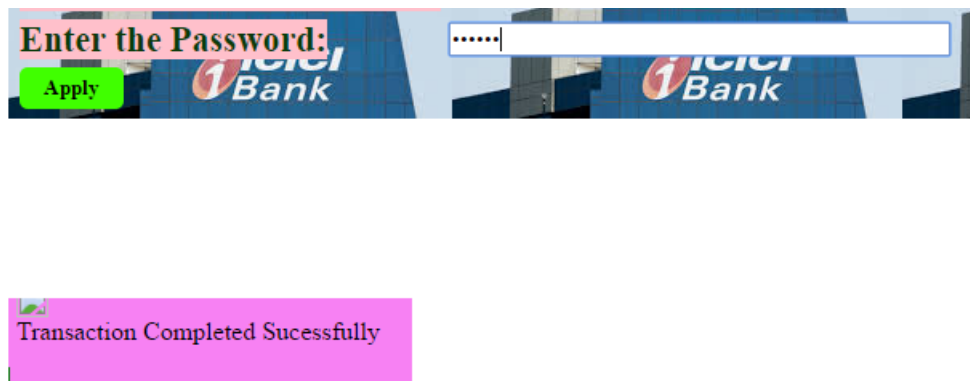
FIGURE 5.4: Direct Rating Viewer

Direct Rating Graphs

Ratings Output		
Product ID	Product NAME	RATING
3	NOKIA LUMINA	9
1	SAMSUNG GALAXY S1	4
2	SAMSUNG GALAXY S3	4
4	LG	0
5	Apple iPhone 6	0

FIGURE 5.5: Rating Graph

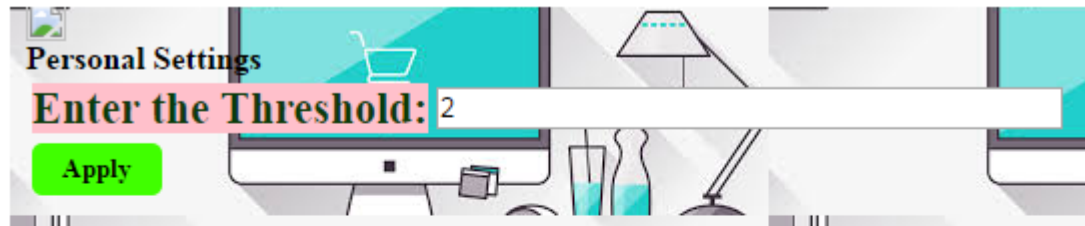
Transaction Page



The screenshot shows a transaction page for '1st Bank'. It features a password prompt 'Enter the Password:' with a green 'Apply' button and a password input field. Below this is a pink confirmation box that reads 'Transaction Completed Sucessfully'.

FIGURE 5.6: Transaction

Personal Settings



Personal Settings

Enter the Threshold: 2

Apply

The image shows a web form titled 'Personal Settings' with a shopping cart icon. A text input field contains the number '2'. Below the field is a green 'Apply' button. The background features a stylized illustration of a desk with a monitor, lamp, and books.

Settings has been Stored Sucessfully

FIGURE 5.7: Settings

Customer Review Submission



Personal Settings

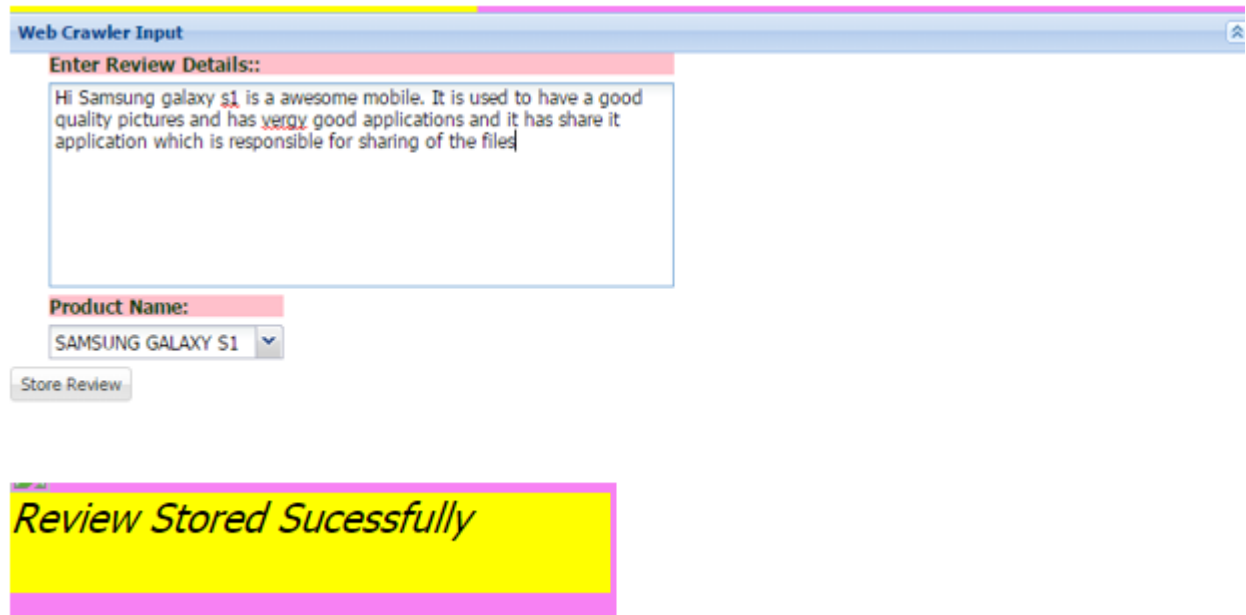
Enter the Threshold: 2

Apply

This image is identical to Figure 5.7, showing the 'Personal Settings' form with the threshold value of 2 and the 'Apply' button.

FIGURE 5.8: Review Submission

Reviews Collected Online and Offline



The screenshot shows a web application interface for storing a review. It features a title bar 'Web Crawler Input', a text area for 'Enter Review Details::' containing a review about the Samsung Galaxy S1, a 'Product Name:' dropdown menu set to 'SAMSUNG GALAXY S1', and a 'Store Review' button. Below the form is a yellow confirmation box with the text 'Review Stored Sucessfully'.

Web Crawler Input

Enter Review Details::

Hi Samsung galaxy s1 is a awesome mobile. It is used to have a good quality pictures and has vergy good applications and it has share it application which is responsible for sharing of the files

Product Name:

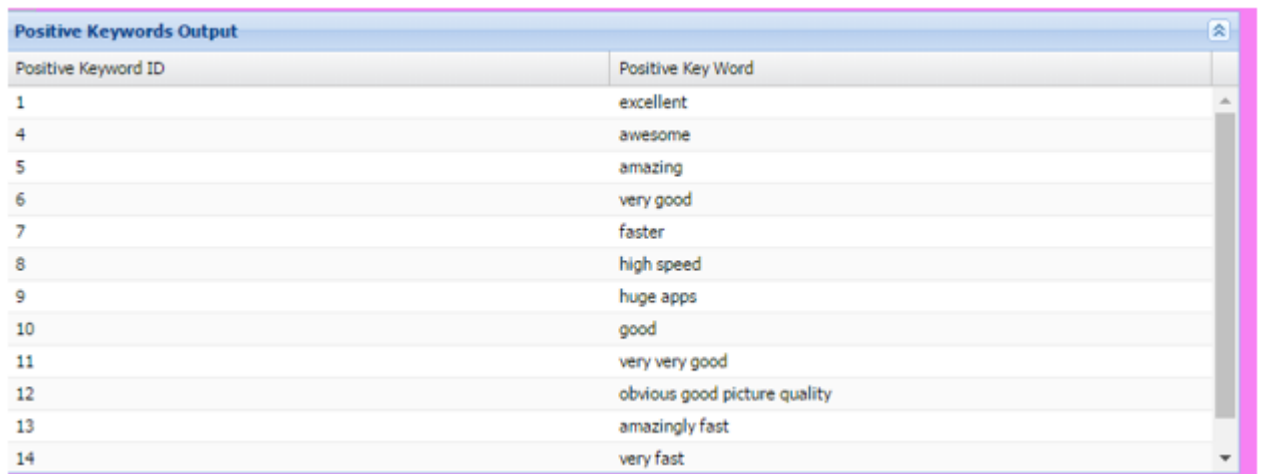
SAMSUNG GALAXY S1

Store Review

Review Stored Sucessfully

FIGURE 5.9: Reviews

Viewing Positive Keywords

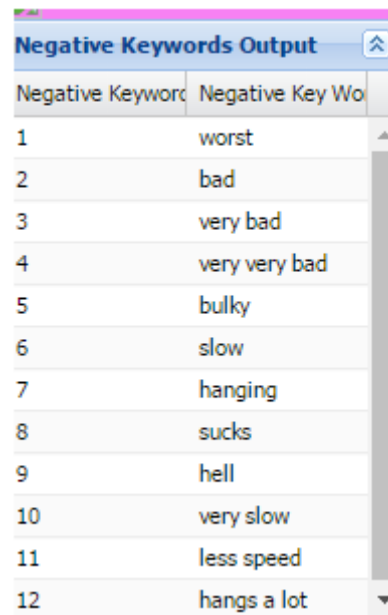


The screenshot displays a table titled 'Positive Keywords Output'. The table has two columns: 'Positive Keyword ID' and 'Positive Key Word'. It lists 14 keywords extracted from the reviews, such as 'excellent', 'awesome', 'amazing', 'very good', 'faster', 'high speed', 'huge apps', 'good', 'very very good', 'obvious good picture quality', 'amazingly fast', and 'very fast'.

Positive Keyword ID	Positive Key Word
1	excellent
4	awesome
5	amazing
6	very good
7	faster
8	high speed
9	huge apps
10	good
11	very very good
12	obvious good picture quality
13	amazingly fast
14	very fast

FIGURE 5.10: Keywords

Negative Keywords Analysis

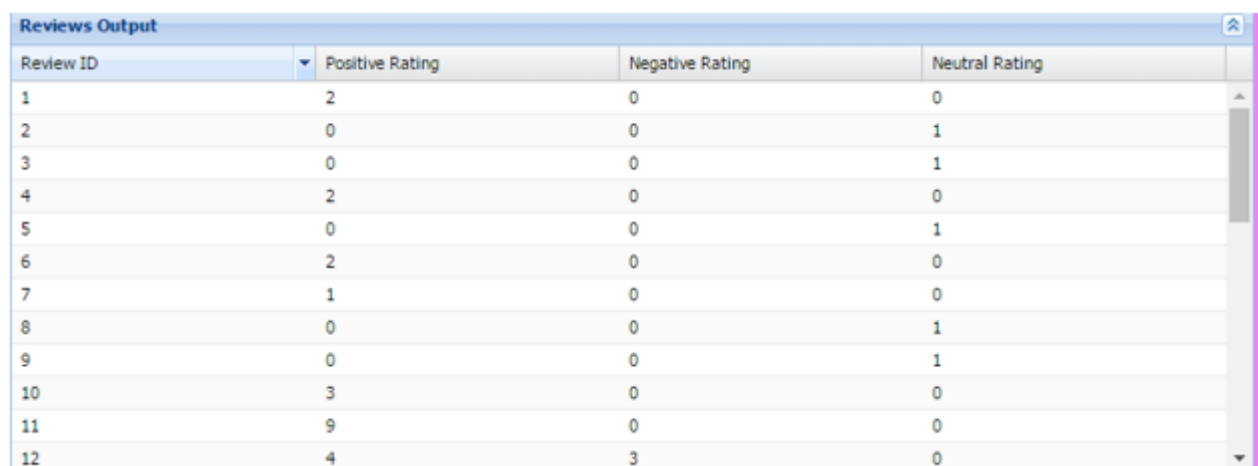


A screenshot of a software window titled "Negative Keywords Output". It contains a table with two columns: "Negative Keyword" and "Negative Key Wo". The table lists 12 items, numbered 1 through 12. The keywords are: worst, bad, very bad, very very bad, bulky, slow, hanging, sucks, hell, very slow, less speed, and hangs a lot. The table has a scrollbar on the right side.

Negative Keyword	Negative Key Wo
1	worst
2	bad
3	very bad
4	very very bad
5	bulky
6	slow
7	hanging
8	sucks
9	hell
10	very slow
11	less speed
12	hangs a lot

FIGURE 5.11: Keywords

Review Based Polarity Matrix



A screenshot of a software window titled "Reviews Output". It contains a table with four columns: "Review ID", "Positive Rating", "Negative Rating", and "Neutral Rating". The table lists 12 items, numbered 1 through 12. The ratings are: Review 1 (2, 0, 0), Review 2 (0, 0, 1), Review 3 (0, 0, 1), Review 4 (2, 0, 0), Review 5 (0, 0, 1), Review 6 (2, 0, 0), Review 7 (1, 0, 0), Review 8 (0, 0, 1), Review 9 (0, 0, 1), Review 10 (3, 0, 0), Review 11 (9, 0, 0), and Review 12 (4, 3, 0). The table has a scrollbar on the right side.

Review ID	Positive Rating	Negative Rating	Neutral Rating
1	2	0	0
2	0	0	1
3	0	0	1
4	2	0	0
5	0	0	1
6	2	0	0
7	1	0	0
8	0	0	1
9	0	0	1
10	3	0	0
11	9	0	0
12	4	3	0

FIGURE 5.12: Polarity Matrix

Polarity Graphs

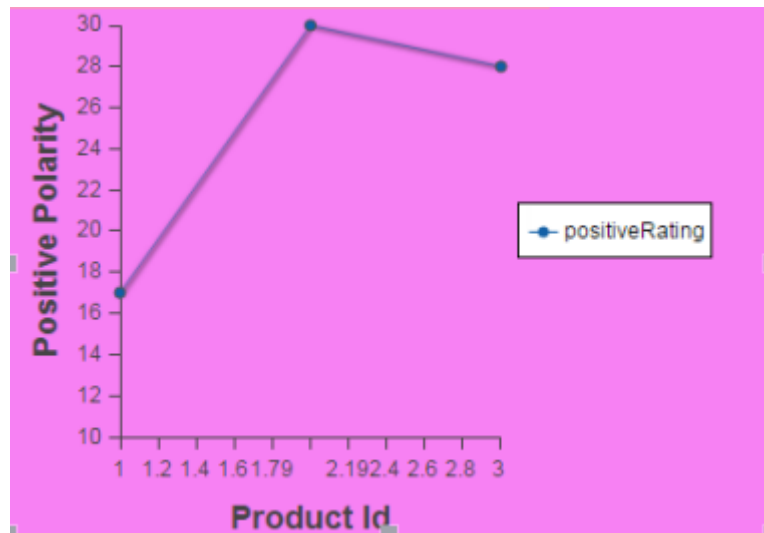


FIGURE 5.13: Positive Polarity

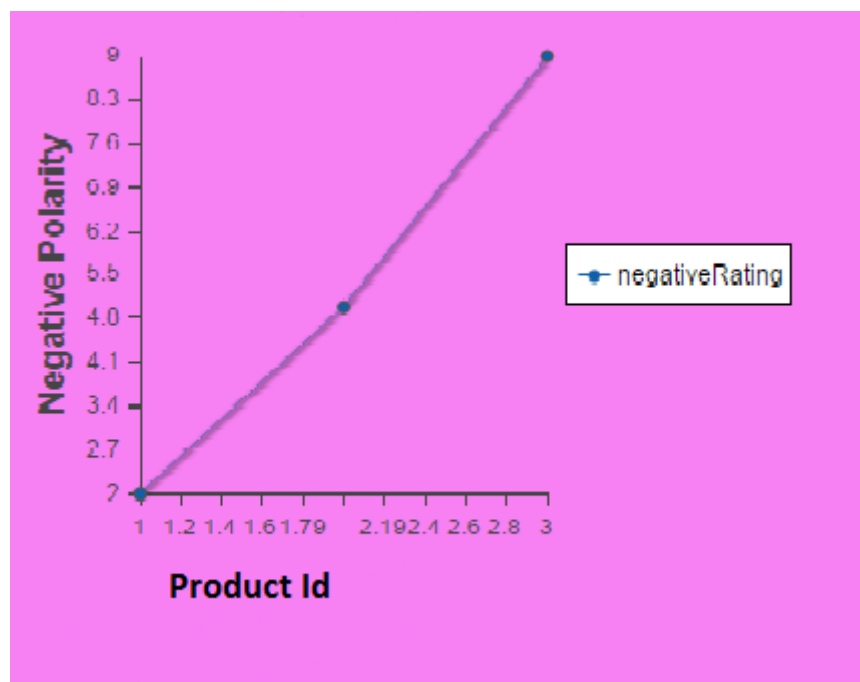


FIGURE 5.14: Negative Polarity

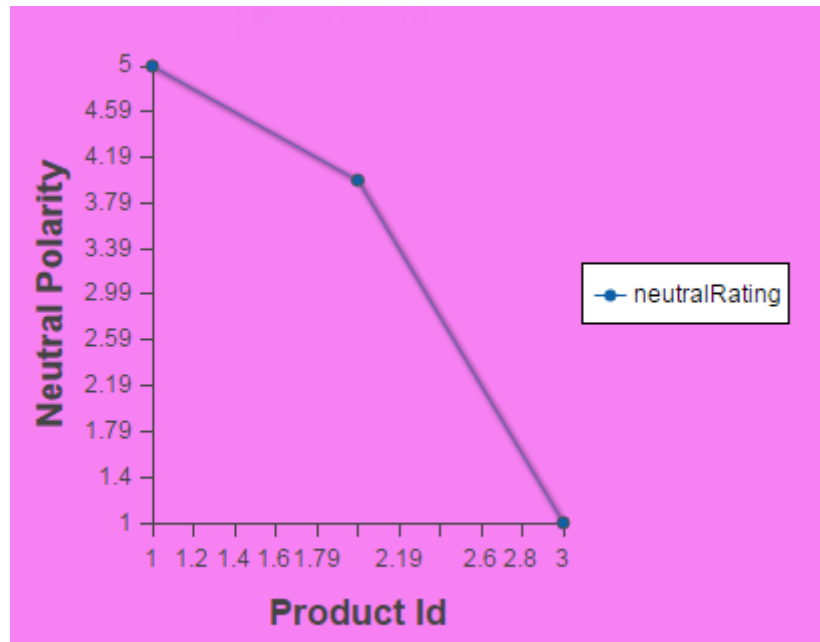


FIGURE 5.15: Neutral Polarity

Pearson Rating Recommendations

Product Name		Product ID	Predicted Ratings
NOKIA LUMINA		3	0.9333333333333333
Apple I Phone6		5	0.1666666666666667
SAMSUNG GALAXY S1		1	-0.2333333333333333
SAMSUNG GALAXY S3		2	-0.6666666666666667
LG		4	-1.4

Product ID	Rating Product	Average Product Rating User	Other Rating User	Average Other Product Rating	Pearson Rating	Predicted Rating
1	1	1.2	3	1.8333333333333333	-0.2333333333333333	0.9666666666666667
2	2	1.2	1	1.8333333333333333	-0.6666666666666666	0.5333333333333333
3	2	1.2	3	1.8333333333333333	0.9333333333333335	2.1333333333333333
5	1	1.2	1	1.8333333333333333	0.1666666666666666	1.3666666666666665
4	0	1.2	3	1.8333333333333333	-1.4000000000000001	0.20000000000000018

FIGURE 5.16: Rating based on Pearson Algorithm

5.2 Results & Analysis

In the result section we are discussed the results obtained by the system for detecting fake and truthful reviews given by the users. Following diagram shows the number of reviews of user. In the following diagram, we have fetch total 80 reviews from which the red region shows the truthful review and blue region shows the fake review detected by the proposed system.

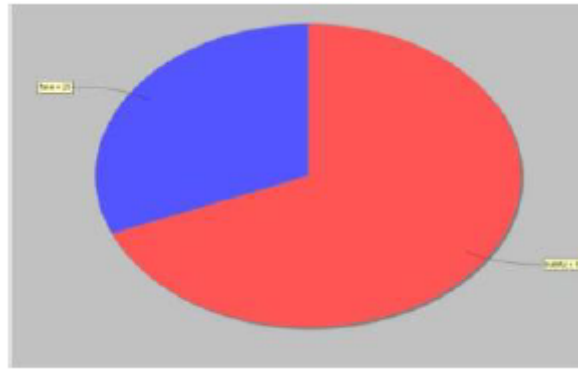


FIGURE 5.17: Average of Truthness and Fakeness in Review

5.3 Comparative Study

The following Table. II shows comparative table for all attributes. The comparison is made between J48 and ICRM. True Positive Rate, True negative Rate, Accuracy, Rule, Condition per Rule are the attributes which are considered for comparison.

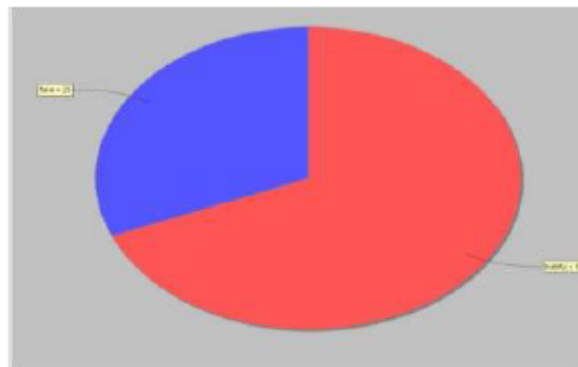


FIGURE 5.18: Graphical Results

TABLE 5.1: Comparative Table of all Attributes

Confusion Matrix	J48	ICRM
True Positive Rate	97.4	97.5
True Negative Rate	53.3	75
Accuracy	93.4	95.5
Rules	31	7.6
Condition per Rules	3.1	1.9

The following graph shows the comparative results of all attributes:

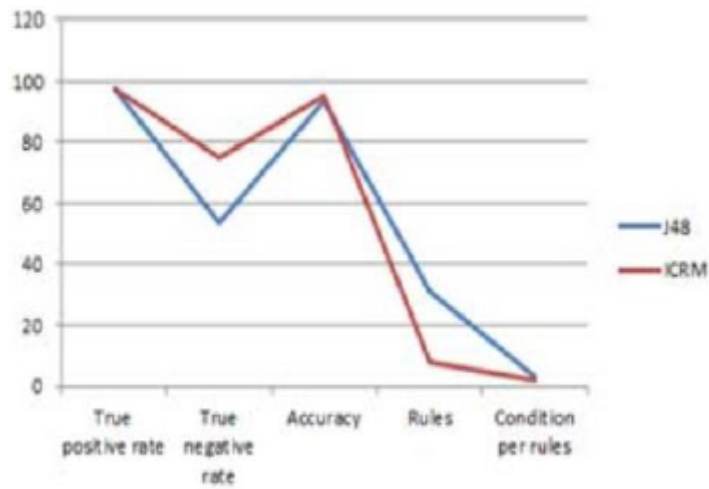


FIGURE 5.19: Comparative Results of all Attributes

5.4 Cost Estimation

Project underestimation of capitals and charges is one of the most important backers to project failure. As such, project managers should be well informed of and reflect the numerous industry practices and tools in the description and performance of project cost estimation. As defined by the Project Management Body of Knowledge (PMBOK), cost estimation is the regular course of evolving an approximation of the fiscal resources required to carry out project activities. Project teams should guess costs for all capitals that will be charged to the project. This comprises of:

- Workforce
- Resources
- Apparatus
- Facility
- Software
- Hardware
- Implementation Costs

The following list comprises of collective tools and methods used in project cost estimation:

Parallel Estimating

Knowledge gained from past project management involvement is used. Proficient judgement, in aggregation with factual estimation approach, provides appreciated information about the structural situation and evidence from previous similar projects.

Analogous Estimating

The metrics from a prior, similar project is set as a foundation for the carrying out estimation for the current project. Parallel approximating takes the genuine cost of previous, similar projects as a starting point and then figures out the shortcomings (such as size, complexity, scope, duration, etc.).

Parametric Estimating

A statistical connection between past data and other variables to calculate an estimate for activity considerations such as scope, cost, budget, and period is used.

Bottom-Up Estimating

All discrete work correspondences / doings with the utmost level of detail, briefing higher-level approximations with the mixture of the individual approximations. The correctness of bottom-up estimating is enhanced when discrete works are expressed in detail.

Three-point Opinion Approximation

We have used three approximations to govern a range for an activity's cost: the best-case approximation, the most favourable approximation, and the worst-case approximation.

Stand-in Analysis

Determination of eventuality stashes to account for cost indecision.

Project Administration Approximating Software

Use of project administration price approximating software applications, computerized spread-sheets, replication, and arithmetical tools. Such tools can permit for quick deliberation of numerous cost approximation replacements.

Vendor Bid Analysis

Determining what the project should cost based on a appraisal of vendor offers / suggestions. This method may be castoff in combination with other cost approximation approaches to guarantee that cost estimates are inclusive. Connected links whereas the implementation of suitable cost approximation methods surely adds to the accuracy of cost approximations, other project administration knowledge areas also play a vital role in cost approximation accurately.

Quality Management

If group members do not settle clearly upon deliverable characteristic criteria in the early phases of project, they may take much time to meet potentials, needlessly causing a schedule delay and consistent cost besieges.

Communications Management If group members do not fully know their roles and tasks on the project, project work may require longer time to complete, thus postponing the agenda and accumulative prices.

Scope Management

If the functionality and requirement of a project is not clear properly and project is delivered without the same, then it might happen that project might be needed to be prepared again, which will cause unnecessary delay and expenditure.

Human Resource Management

If team members do not have the mandatory services or knowledge to carry out project work, it may cause longer delivery period, causing delays and cost besieges.

Risk Management

If group mates do not conduct risk management from time to time, cost impacting problems that could have been avoided may arise.

Time Management

If group mates do not precisely approximate the time to carry out the activities, the project may experience postponements and cost besieges.

5.4.1 Cost In Terms Of Software And Hardware Requirement

SOFTWARE REQUIREMENTS

- Windows and above
- My Sql
- JSP

HARDWARE REQUIREMENTS

- Processor – i3
- Hard Disk – 5 GB
- Memory – 1GB RAM

5.4.2 Cost Estimation for sale of product

This project of us is definitely going to be of real use to several e-commerce based companies. So for the same if we are going to sell our product to them we have kept a cost of 25000 for the upgraded version of this software also we will provide online support system free of cost to our clients.

5.5 CONCLUSION

This project suggests behavioural tactic to identify fake reviewers who try to alter or do not provide correct ratings on some important products. We develop an cumulative behaviour scoring approach for reviewers keeping in mind the amount of spamming behaviours they show. So in order to evaluate our planned methods, that performs test or take reviews from different e-commerce websites containing reviews of different manufactured products. We identified that this proposed method generally overtakes the reference point method based divisions. We additionally learn a reversion model from the user considered ground truth reviewers. The feedback and views are used for decision making by users and companies to get par with their competitors. But these reactions come under the negatives like bad promotion and then it is really hard to reach right people and identify true reviews. It becomes compulsory that to sense opinion spam and fake reviewer. This project focuses on review based spam detection which provides greater focus on review collection and it's analysis. As a part of future work, we can include fake reviewer detection. Finding ways to learn behavioural outlines connected to fake reviews so as to advance the correctness of the current regression model is also an thought-provoking research area.

5.6 FUTURE WORK

There is a urgent need in the industry,for a mechanism to identify the fake reviews and reviewers in order to keep a proper track of product compatibility with market so thate-commerce websites can provide worthy informations. This project suggests an opinion spam analyzer which spontaneously separates review data collected into either fake or non-fake category.The planned system will use machine learning chore of deducing a function from a labeled data.The selected algorithm based on recreation work is Sup-port Vector Machine (SVM). A course for future research is to employ the system and check execution by applying proposed methods to various standard data sets. Comparing routine of different categorization methods to find the best one for our projected opinion spam classification method could be another future research area. However, there exist lot of other kinds of review or reviewer related features that are expected to make a impact to the prediction task.In the future further investigation is needed in order to incorporate different kinds of features to make more exact estimates.

Bibliography

- [1] L. Akoglu, R. Chandy, and C. Faloutsos, “Opinion fraud detection in online reviews by network effects,” in *International Conference on Web and Social Media*, 2013.
- [2] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, “Exploiting burstiness in reviews for review spammer detection,” in *International Conference on Web and Social Media*, 2013.
- [3] D. Hernandez, R. Guzman, M. Montes Gomez, and P. Rosso, “Using pu-learning to detect deceptive opinion spam,” in *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Georgia, pp. 38–45, June 2013.
- [4] N. Jindal and B. Liu, “Opinion spam and analysis,” in *Web Search and Data Mining*, pp. 219–230, 2008.
- [5] X. Kong, P. S. Yu, Y. Ding, and D. J. Wild, “Meta path-based collective classification in heterogeneous information networks,” in *Conference on Information and Knowledge Management*, pp. 1567–1571, 2012.
- [6] B. Liu, Y. Dai, X. Li, W. S. Lee, and P. S. Yu, “Building text classifiers using positive and unlabeled examples,” in *International Conference on Data Mining*, pp. 179–188, 2003.

- [7] Q. Lu and L. Getoor, “Link-based classification.” in *International Conference on Machine Learning*, pp. 496–503, 2003.
- [8] A. Mukherjee, V. Venkataraman, B. Liu, and N. S. Glance, “What yelp fake review filter might be doing?” in *International Conference on Web and Social Media*, 2013.
- [9] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. EliassiRad, “Collective classification in network data,” *AI Magazine*, vol. 29, no. 3, pp. 93–106, 2008.
- [10] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, “Pathsim: Meta path-based top-k similarity search in heterogeneous information networks,” *The Proc. of the VLDB Endowment*, vol. 4, no. 11, pp. 992–1003, 2011.
- [11] G. Wang, S. Xie, B. Liu, and P. S. Yu, “Review graph based online store review spammer detection,” in *International Conference on Data Mining*, pp. 1242–1247, 2011.

Appendix A

ALGORITHMS

A.1 PEARSON RATING ALGORITHM

```
public PearsonRatingResults rankBasedOnCustomerRating(String ↵
    userName) {

    PearsonRatingResults pearsonRatingResults = new ↵
        PearsonRatingResults();
    List<ProductRating> productRatingList = null;
    try
    {
        productRatingList = new ArrayList<ProductRating>();

        List<Integer> productIds = ↵
contentDao.retrieveUniqueProductIdsFromPearsonRating();

        if (null == productIds || productIds.isEmpty()
            || productIds.size() == 0)
        {
            return null;
        }

        // Individual Rating Computed

        int totalRatingUser = 0;
```

```

int totalRatingOtherUser = 0;

int userValuesCount = 1;

int otherUserValuesCount = 1;

List<PearsonRatingInfo> pearsonRatingInfoList = new ←
ArrayList<PearsonRatingInfo>();

for (Integer productId : productIds)
{

    PearsonRatingInfo pearsonRatingInfo = new PearsonRatingInfo();

    int countForProduct = contentDao
        .retriveCountForProductIdAndUserName(productId,
        userName);

    if (countForProduct > 0)
    {
        userValuesCount = userValuesCount + 1;
    }

    totalRatingUser = totalRatingUser + countForProduct;

    pearsonRatingInfo.setProductId(productId);
    pearsonRatingInfo.setRatingProduct(countForProduct);

    int otherUserRating = contentDao
        .retriveCountForProductIdAndNotUserName(productId,
        userName);

    if (otherUserRating > 0)
    {
        otherUserValuesCount = otherUserValuesCount + 1;
    }

    totalRatingOtherUser = totalRatingOtherUser + otherUserRating;

```



```

        pearsonRatingInfo.setOtherUserRating(otherUserRating);

        pearsonRatingInfoList.add(pearsonRatingInfo);
    }

// Individual Rating Computed with mean and numerator

    double userMean = (double) ((double) totalRatingUser / (double) ←
userValuesCount);

    double otherUsersMean = (double) ((double) totalRatingOtherUser ←
/ (double) otherUserValuesCount);

    double denomratorValue = 0;

    List<PearsonRatingInfo> pearsonRatingInfoListNew = new ←
ArrayList<PearsonRatingInfo>();

    for (PearsonRatingInfo pearsonRatingInfo : pearsonRatingInfoList)
    {

        PearsonRatingInfo pearsonRatingInfo2 = new PearsonRatingInfo();

        pearsonRatingInfo2.setProductId(pearsonRatingInfo
.getProductId());
        pearsonRatingInfo2.setRatingProduct(pearsonRatingInfo
.getRatingProduct());
        pearsonRatingInfo2.setOtherUserRating(pearsonRatingInfo
.getOtherUserRating());
        pearsonRatingInfo2.setAverageProductRatingUser(userMean);
        ←
        pearsonRatingInfo2.setAverageOtherProductRating(otherUsersMean);

        double numerator1 = (pearsonRatingInfo.getRatingProduct() - ←
userMean)
            * (pearsonRatingInfo.getOtherUserRating() - otherUsersMean);

        pearsonRatingInfo2.setNumerator(numerator1);
    }

```

```

        denomratorValue = denomratorValue + numerator1;

        pearsonRatingInfoListNew.add(pearsonRatingInfo2);

    }
    if (denomratorValue <= 0)
    {
        denomratorValue = 1;
    }

    // Adding the Pearson

    List<PearsonRatingInfo> pearsonRatingInfos = new ↵
    ArrayList<PearsonRatingInfo>();

    double pearsonDenom = 0;

    for (PearsonRatingInfo pearsonRatingInfo : ↵
    pearsonRatingInfoListNew)
    {

        PearsonRatingInfo pearsonRatingInfo3 = new PearsonRatingInfo();

        pearsonRatingInfo3
        .setAverageOtherProductRating(pearsonRatingInfo
        .getAverageOtherProductRating());
        pearsonRatingInfo3
        .setAverageProductRatingUser(pearsonRatingInfo
        .getAverageProductRatingUser());
        pearsonRatingInfo3.setOtherUserRating(pearsonRatingInfo
        .getOtherUserRating());
        pearsonRatingInfo3.setRatingProduct(pearsonRatingInfo
        .getRatingProduct());
        pearsonRatingInfo3.setProductId(pearsonRatingInfo
        .getProductId());
        pearsonRatingInfo3.setNumerator(pearsonRatingInfo
        .getNumerator());
        pearsonRatingInfo3.setTotalSum(denomratorValue);
    }
}

```

```

        double pearson = (double) ((double) pearsonRatingInfo
            .getNumerator() / (double) denomratorValue);

        pearsonDenom = pearsonDenom + pearson;

        pearsonRatingInfo3.setPearsonRating(pearson);

        pearsonRatingInfos.add(pearsonRatingInfo3);

    }

    if (pearsonDenom <= 0)
    {
        pearsonDenom = 1;
    }

//

    List<Double> predictedRatingList = new ArrayList<Double>();

    List<PearsonRatingInfo> pearsonRatingInfosNewList = new ↵
    ArrayList<PearsonRatingInfo>();

    for (PearsonRatingInfo pearsonRatingInfo : pearsonRatingInfos)
    {

        PearsonRatingInfo pearsonRatingInfo3 = new PearsonRatingInfo();

        pearsonRatingInfo3
            .setAverageOtherProductRating(pearsonRatingInfo
            .getAverageOtherProductRating());
        pearsonRatingInfo3
            .setAverageProductRatingUser(pearsonRatingInfo
            .getAverageProductRatingUser());
        pearsonRatingInfo3.setOtherUserRating(pearsonRatingInfo
            .getOtherUserRating());
        pearsonRatingInfo3.setRatingProduct(pearsonRatingInfo
            .getRatingProduct());
        pearsonRatingInfo3.setProductId(pearsonRatingInfo

```

```

        .getProductId());
        pearsonRatingInfo3.setNumerator(pearsonRatingInfo
        .getNumerator());
        pearsonRatingInfo3.setTotalSum(denomratorValue);
        pearsonRatingInfo3.setPearsonRating(pearsonRatingInfo
        .getPearsonRating());

        double predictedRating = (double) ((double) pearsonRatingInfo
        .getAverageProductRatingUser() + ((double) pearsonRatingInfo
        .getPearsonRating() / (double) pearsonDenom));

        if (predictedRating < 0)
        {
            predictedRating = predictedRating * -1;
        }

        predictedRatingList.add(predictedRating);

        pearsonRatingInfo3.setPredictedRating(predictedRating);

        pearsonRatingInfosNewList.add(pearsonRatingInfo3);
    }

    StatusInfo statusInfo = contentDao.deleteTotalPearsonRatings();

    if (!statusInfo.isStatus())
    {

        return null;
    }

    statusInfo = contentDao
    .insertPearsonRatingRank(pearsonRatingInfosNewList);

    if (!statusInfo.isStatus())
    {

        return null;
    }

```

```

    }

    List<PearsonRatingRank> pearsonRatingRanks = contentDao
        .retrivePearsonRatingRank();

    if (null == pearsonRatingRanks)
    {

        return null;
    }

    List<PearsonMainInfo> pearsonMainInfoList = new ↵
        ArrayList<PearsonMainInfo>();
    for (PearsonRatingRank predictedRating : pearsonRatingRanks)
    {

        PearsonMainInfo pearsonMainInfo = new PearsonMainInfo();
        pearsonMainInfo.setPredictedRating(predictedRating
            .getPredictedRating());
        pearsonMainInfo.setProductId(predictedRating.getProductId());
        String productName = contentDao
            .retriveProductNameForId(predictedRating.getProductId());
        pearsonMainInfo.setProductName(productName);

        pearsonMainInfoList.add(pearsonMainInfo);
    }

    pearsonRatingResults
        .setPearsonRatingList(pearsonRatingInfosNewList);
    pearsonRatingResults.setPearsonMainInfoList(pearsonMainInfoList);

    return pearsonRatingResults;

}

catch (Exception e)
{
    e.printStackTrace();
    System.out.println("Exception " + e.getMessage());
}

```

```
        return null;  
    }  
}
```