A Project Report On

E-COMMERCE PRODUCT REVIEW MANAGEMENT SYSTEM BASED ON OPINION MINING

Submitted in partial fulfillment of the requirement

of the degree of

Bachelor in Engineering

by

Sujith Nair	BE-4-38
Apresh Pandit	BE-4-42
Sanal Pillai	BE-4-46
Onkaar Sawant	BE-4-57

Under the guidance of

Prof. Rupali Kale

DEPARTMENT OF COMPUTER ENGINEERING

SHAH AND ANCHOR KUTCHHI ENGINEERING COLLEGE

CHEMBUR, MUMBAI-400088

2017 - 2018

www.shahandanchor.com Tel: 022 2558 0854

Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Mahavir Education Trust Chowk, W.T. Patil Marg, Chembur, Mumbai 400 088

Affiliated to University of Mumbai, Approved by D.T.E. & A.I.C.T.E.

Awarded provisional accreditation for Computer & Electronics Engineering by NBA

(for 2 years from 06-08-2014)

Certificate

This is to certify that the report of the project entitled

E-COMMERCE PRODUCT REVIEW MANAGEMENT SYSTEM BASED ON OPINION MINING

is a bonafide work of

Sujith Nair	BE-4-38
Apresh Pandit	BE-4-42
Sanal Pillai	BE-4-46
Onkaar Sawant	BE-4-57

submitted to the

UNIVERSITY OF MUMBAI

during semester VIII in partial fulfilment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER ENGINEERING.

(Prof. Rupali Kale)

Guide

(Prof. Uday Bhave)

I/c Head of Department

(Dr. Bhavesh Patel)

Principal

Approval for Project Report for B. E. semester VIII

This project report entitled E-COMMERCE PRODUCT REVIEW MANAGEMENT SYSTEM BASED ON OPINION MINING by Sujith Nair, Apresh Pandit, Sanal Pillai, Onkaar Sawant is approved for semester VIII in partial fulfilment of the requirement for the award of the degree of Bachelor of Engineering.

Examiners

1._____

2._____

Guide

Prof Rupali Kale: _____

Date:

Place: MUMBAI

Attendance Certificate

Date:

To,
The Principal
Shah and Anchor Kutchhi Engineering College,
Chembur, Mumbai-88

Subject: Confirmation of Attendance

Respected Sir,

This is to certify that Final year (BE) students
Sujith Nair
Apresh Pandit
Sanal Pillai
Onkaar Sawant

have duly attended the sessions on the day allotted to them during the period from January to March for performing the Project titled E-COMMERCE PRODUCT REVIEW MANAGEMENT SYSTEM BASED ON OPINION MINING.

They were punctual and regular in their attendance. Following is the detailed record of the student's attendance.

Attendance Record:

Date	Sujith Nair	Apresh Pandit	Sanal Pillai	Onkaar Sawant
17/1/18	Present	Present	Present	Present
8/2/18	Present	Present	Present	Present
28/2/18	Present	Present	Present	Present
21/3/18	Present	Present	Present	Present
28/3/18	Present	Present	Present	Present

Signature and Name of Internal Guide

Prof. Rupali Kale

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of student	Roll No.	Signature
Sujith Nair	BE-4-38	_____
Apresh Pandit	BE-4-42	_____
Sanal Pillai	BE-4-46	_____
Onkaar Sawant	BE-4-57	_____

Date:

Abstract

The growth of E-commerce has led to the invention of several websites that market and sells products as well as allows users to post reviews. It is typical for an online buyer to refer to these reviews before making a buying decision. Hence, automatic summarization of users' reviews has a great commercial significance. However, since the product reviews are written by non-experts in an unstructured, natural language text, the task of summarizing them is challenging. This Project presents a semi supervised approach for mining online user reviews to generate comparative feature-based statistical summaries that can guide a user in making an online purchase. It includes various phases like preprocessing and feature extraction and pruning followed by feature based opinion summarization and overall opinion sentiment classification. Results indicate that the approach used can identify opinionated sentences from user reviews with an acceptable accuracy.

Table of contents

List of Figures

List of Tables

Chapter 1

Introduction

There are many users who purchase products through E-commerce websites. Through online shopping many E-commerce enterprises were unable to know whether the customers are satisfied by the services provided by the firm. This boosts us to develop a system where various customers give reviews about the product and online shopping services, which in turn help the E-commerce enterprises and manufacturers to get customer opinion to improve service and merchandise through mining customer reviews. An algorithm could be used to track and manage customer reviews, through mining topics and sentiment orientation from online customer reviews.

1.1 Objective

The objective of this project is to develop a system which will help customers who are buying products online to understand the popularity and efficiency of the product through the sentiments of customers who have earlier used that product and reviewed them.

1.2 Methodology used

We generated an opinion review database by crawling some popular websites that categorically post product reviews by actual users. Our product opinion summarizer has three main phases. These phases are

(1) preprocessing phase,

(2) feature extraction phase, and

(3) opinion summarization and classification phase.

These phases are briefly described next.

Preprocessing Phase:

Online blog reviews posted by users frequently contain spelling errors and incorrect punctuation. Our next phase—the feature-extraction phase—requires parts-of- speech tagging which works at the sentence level. Thus, it becomes important to detect end of sentences. So, in this phase we performed basic cleaning tasks like sentence boundary detection and spell-error correction. Sentences normally end with punctuations like period (.), question mark (?), or exclamation mark (!). Sometimes bloggers overuse the "?" and "!" symbols for emphasis. For example, a blogger may post a review that says "It's surprising that the eBook reader does not have a touch screen !!!!"

In such cases we conflate the repetitive punctuation symbols to a single occurrence (i.e., "!!!!" is replaced by a single "!").

Several other considerations arise during the Preprocessing phase. The period (.) requires to be disambiguated as it may mean a full stop or a decimal point or an abbreviation (e.g., "Dr.," "Ltd."). Sometimes a single sentence straddles multiple lines as the user presses unnecessary return keys. In such cases we apply the sentence merge rules.

After sentence boundary detection, we perform spell-error correction using a word processor.

Feature Extraction Phase:

In this phase we extract opinion features from the pre-processed review text obtained from the previous phase. We treat frequently occurring nouns (N) and noun phrases (NP) as possible opinion features and associated adjectives describing them as indicators of their opinion orientation.

We perform parts-of- speech (POS) tagging on the review sentences using the Link Grammar Parser [3]. The Link Grammar Parser is a well-known and efficient syntactic parser for English language. First, we extract all nouns (N) and noun phrases (NP) tagged by the Link Grammar Parser and identify the frequently occurring N and NP as possible opinion features. By frequently occurring N and NP we mean those Ns and NP which occur at least five times in the users' reviews. We do not extract frequent itemsets from review sentence database using the Apriori based approach, since this method mines frequent features using a BOW (bag-of- words) approach and does not take into account the order in which the words of a phrase occur. Moreover, mining in this way would require ordering besides compactness and redundancy pruning [4, 8]. We also do not use the seed-set expansion approach as it would require prior domain knowledge to specify a seed set. Instead we generate a frequent feature set using the multiword approach.

Opinion Summarization and Classification Phase:

In the previous phase we extracted opinion features, adjectives describing them, and any modifiers if present. We also generate a statistical feature-wise summary for each product which enables comparison of different brands selling similar products. In order to determine the sentiment polarity of an adjective describing an opinion feature we make use of SentiWordNet which is a lexical resource for opinion mining. SentiWordNet assigns three normalized sentiment scores: positivity, objectivity, and negativity to each synset of WordNet. Let us revisit the review sentence:

"The processor[.n] is[.v] significantly faster[.a], and the text[.n] is[.v] clear[.a]."

In this example, the SentiWordNet scores assigned to the appropriate usage of adjective clear is indicated as (P:0.625; O: 0.375; N: 0). Since the value of the positive polarity is highest, the adjective "clear" can be assigned a positive polarity.

In this way, we generate a feature-orientation table (FO table) that records the opinion features and their corresponding descriptors of positive and negative polarities. The Table 1 shows the FO table entries for some of the features of product "Tablet." The FO table, thus generated, enables us to generate feature-wise summary of a product or comparative summaries of different brands of similar products.

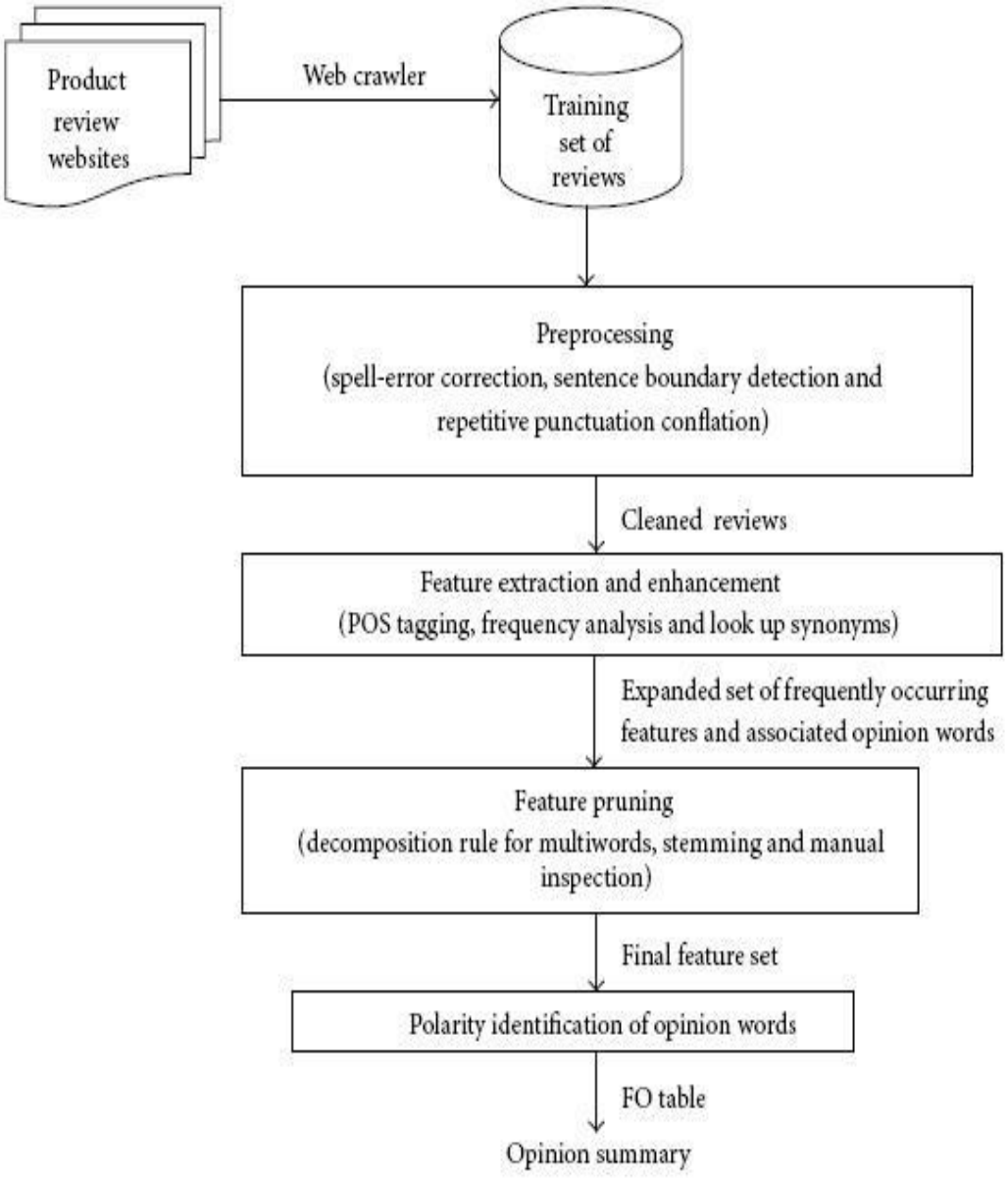


Fig 1.1: System Architecture Diagram [1]

Chapter 2

Literature review

Classification and summarization of online reviews are very important to the growth of E-commerce and social networking applications. Earlier work on automatic text summarization has mainly focused on extraction of sentences that are more significant in comparison to others in a document corpus. The main approaches used to generate extractive summaries are (1) combinations of heuristics such as cue words, key words, title words or position (2) lexical chains, and (3) rhetorical parsing theory. [1]

However, it is important to note that the task of summarizing online product reviews is very different from traditional text summarization, as it does not involve extracting significant sentences from the source text. Instead, while summarizing user reviews, the aim is to first of all identify semantic features of products and next to generate a comparative summary of products based on feature-wise sentiment classification of the reviews that will guide the user in making a buying decision. The authors have demonstrated that traditional unsupervised text classification techniques like naive Bayes, maximum entropy, and support vector machine do not perform well on sentiment or opinion classification and pointed out the necessity for feature-oriented classification. Thus, recent research work in opinion mining has focused on feature based extraction and summarization. [3]

Opinion mining from users' reviews involves two main tasks— (1) identification of the opinion feature set and (2) sentiment analysis of users' opinions based on the identified features.

It has been observed that nouns and noun phrases (N and NP) frequently occurring in reviews are useful opinion features, while the adjectives and adverbs describing them are useful in classifying sentiment. [4]

In order to extract nouns, noun phrases, and adjectives from review text, parts-of-speech (POS) tagging is performed. However, all nouns and noun phrases are not useful in mining and cannot directly be included in the feature set. So, the feature set is subsequently extracted using approaches that involve frequency analysis and/or use of domain knowledge as is discussed next. [4]

Various methods exist in the literature to associate features with their corresponding descriptors. Hu and Liu proposed the nearby-adjective heuristic. Although this method is simple and fast, it may result in inaccuracies. So, supervised approaches to determine association have been proposed in recent years such as syntactic dependency parsing and syntactic tree templates. [4]

In "Fined-grained product features extraction and categorization in reviews opinion mining", focuses on review mining and sentiment analysis on Amazon website. Users of the online shopping site Amazon are encouraged to post reviews of the products that they purchase. Amazon employs a 1-to-5 scale for all products, regardless of their category, and it becomes challenging to determine the advantages and disadvantages to different parts of a product. [7]

In "Mining popular menu items of a restaurant from web reviews", aim is to automate the process of gathering online end user reviews for restaurants and analyzing those reviews in terms of the sentiments expressed about specific features. It is simply an approach to make use of the vast amount of growing user domination on the Internet in reviewing important business decisions, policies, services etc by analyzing and visualizing the data in a user-friendly manner, to-the-point. [3]

According to "Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews ", the aspect identification and their importance ranking is measured and sentiment analysis on those aspects is performed. [8]

"A product features mining method based on association rules and the degree of property co-occurrence ",studied that a prototype system can be used to track and manage customer reviews. With the rapid development of e-commerce, customer reviews will become more and more important for e-commerce enterprises and manufacturers. The prototype system model can be a reference for e-commerce enterprises, which is a cost-effective solution available to manage and analyze online reviews. [6]

Chapter 3

Problem Statement

The main goal of our project is to provide a prototype system that could be used to track and manage customer reviews, the opinion of customers who bought that product, and also provide detailed comparison of features through data mining techniques and sentiment analysis from online customer reviews. The project will help determine the overall sentiment of a user's review and provide a graphical representation of the same.

Chapter 4

System Requirements

4.1 Hardware Requirements:

- 1 GB RAM.
- 200 GB HDD.
- Intel 1.66 GHz Processor Pentium 4

4.2 Software Requirements:

- Windows XP, Windows 7,8
- Visual Studio 2010
- MS SQL Server 2008
- Windows Operating System

Chapter 5

Design

This section describes the fundamental design aspects of the project. Software design is a process to transform user requirements into suitable form.

5.1 System Block Diagram

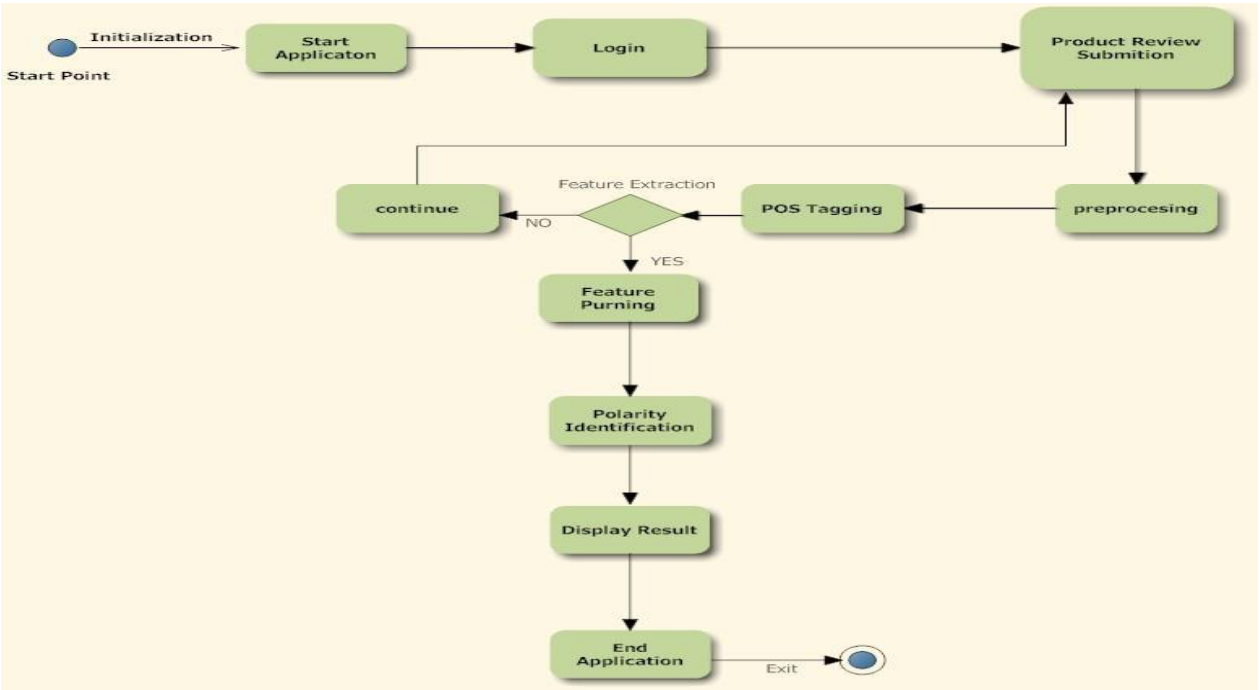


Fig 5.1: System Block Diagram

A block diagram is a diagram of a system in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks. They are heavily used in engineering in hardware design, electronic design, software design, and process flow diagrams.

Block diagrams are typically used for higher level, less detailed descriptions that are intended to clarify overall concepts without concern for the details of implementation. Contrast this with the schematic diagrams and layout diagrams used in electrical engineering, which show the implementation details of electrical components and physical construction.

The system basically starts with the user logging in to the system. It then moves to "Product Review Submission" which allows the user to submit his/her own reviews on a product. The review is then pre-processed and POS (part-of-speech) tagging is performed. The system further performs feature identification from the review, followed by polarity identification of the review.

The result of the analysis is displayed as a graph showing polarity analysis or the polarity of the review is displayed. The system stops after user logs out of the system.

5.2 Flow Chart

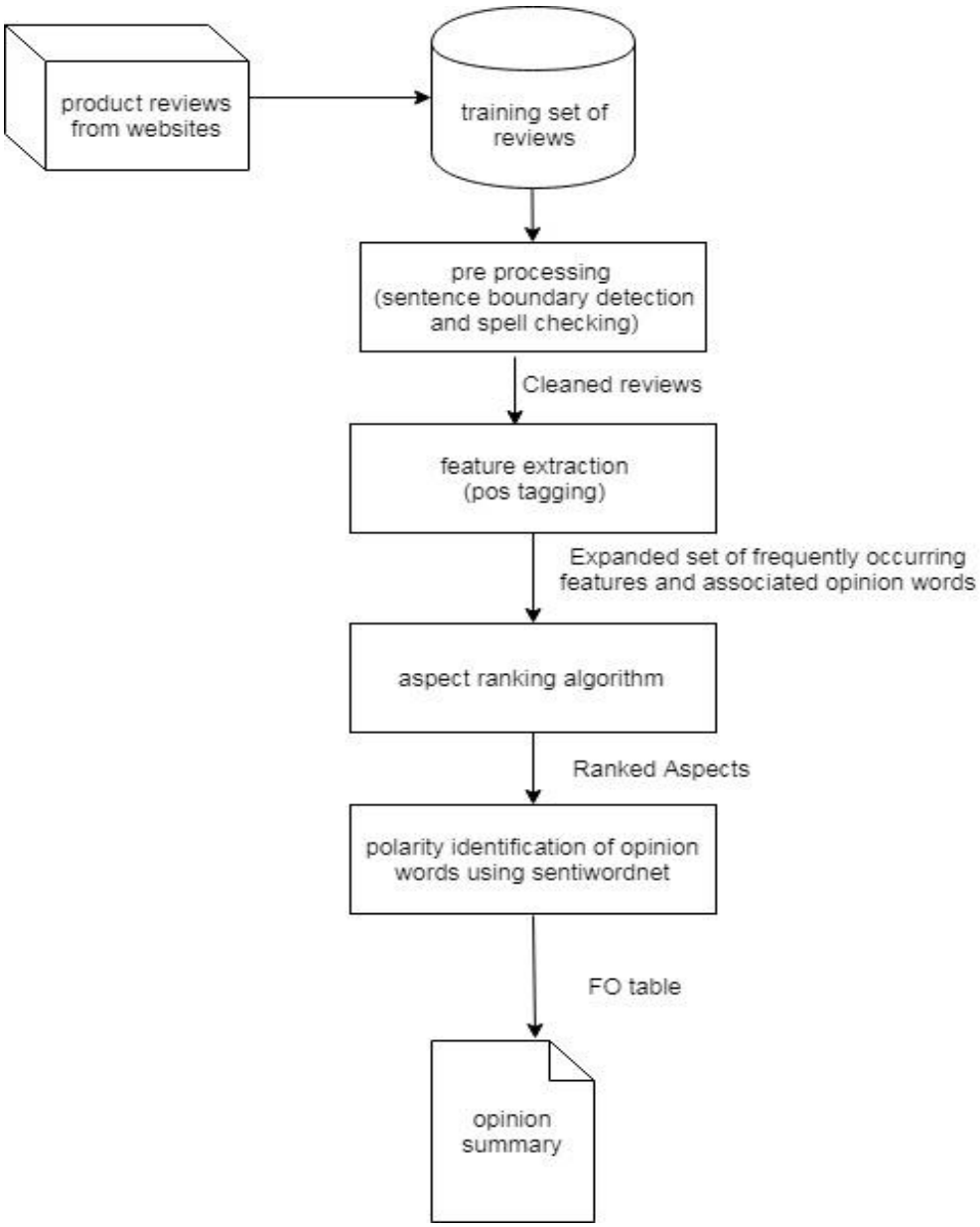


Fig 5.2: Flow Chart

The product reviews are gathered from the website and are preprocessed to eliminate spelling mistakes and to detect the sentence boundary. POS tagging is done on these 'clean' reviews and then the aspects (features) are ranked with respect to the frequency with which they occur.

The polarity of the opinion words is found using Sentiwordnet, and they are then summarized in the feature opinion (FO) table.

5.3 Algorithm

Aspect Ranking Algorithm:

a) Terms used in Algorithm

- D= {r1, r2,r3…rn} be the set of reviews.
- Ak= {a1, a2, a3,…… an} be the set of aspect
- Ca,D is the number of times aspect term a occurs in review dataset D.
- Pa is the number of comments in the positively labeled set with aspect term a.
- |P| is the number of comments in the positively labeled set.
- Na is the number of comments in the negatively labeled set with aspect term a.
- |N| is the number of comments in the negatively labeled set.
- Va,D is the feature value for aspect term a in review dataset D.
- Let Φ = set of positive words

  Φ= {P1, P2, P3 ... Pn}

- Let ψ = set of negative words

  ψ = {N1, N2, N3 ... Nn}

- P(Φ)=probability of Φ
- P(ψ) =probability of ψ
- ω weight of aspect a

b) Algorithm Steps

- Calculate *the value of aspect a, given by*

$$V_{a,D} = C_{a,D} * \log_2(|P|/ P_a) - C_{a,D} * \log_2(|N|/ N_a)$$

$$= C_{a,D} * \log_2((|P| N_a / P_a|N|)$$

$$= C_{a,D} * \log_2(N_a / P_a)$$

- Calculate *the occurrence probability of each positively opinionated word.*

$$\alpha = \sum_{i=1}^{n} P(\Phi i) * W(\Phi i)$$

- Calculate *the occurrence probability of each negatively opinionate word.*

$$\beta = \sum_{i=1}^{n} P(\psi i) * W(\psi i)$$

- Calculate *weight,*

$$\omega = V_{a,D} - \sum_{i=1}^{D} (\alpha - \beta)$$

Chapter 6

Implementation Details

6.1 Tools and Technologies Used

ASP.NET

ASP.NET is more than the next version of Active Server Pages (ASP); it is a unified Web development platform that provides the services necessary for developers to build enterprise-class Web applications. While ASP.NET is largely syntax-compatible with ASP, it also provides a new programming model and infrastructure that enables a powerful new class of applications. You can migrate your existing ASP applications by incrementally adding ASP.NET functionality to them. ASP.NET is a compiled .NET framework-based environment. You can author applications in any .NET Framework compatible language, including Visual Basic and Visual C#. Additionally, the entire .NET Framework platform is available to any ASP.NET application. Developers can easily access the benefits of the .NET Framework, which include a fully managed, protected, and feature-rich application execution environment, simplified development and deployment, and seamless integration with a wide variety of languages.

Microsoft SQL Server

Business today demands a different kind of data management solution. Performance scalability, and reliability are essential, but businesses now expect more from their key IT investment.

SQL Server 2008 exceeds dependability requirements and provides innovative capabilities that increase employee effectiveness, integrate heterogeneous IT ecosystems, and maximize capital and operating budgets. SQL Server 2008 provides the enterprise data management platform your organization needs to adapt quickly in a fast-changing environment.

Benchmarked for scalability, speed, and performance, SQL Server 2008 is a fully enterprise-class database product, providing core support for Extensible Markup Language (XML) and internet queries.

Microsoft Visual Studio

Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs, as well as web sites, web apps, web services and mobile apps. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store and Microsoft Silverlight. It can produce both native code and managed code.

Visual Studio includes a code editor supporting IntelliSense (the code completion component) as well as code refactoring. The integrated debugger works both as a source-level debugger and a machine-level debugger. Other built-in tools include a code profiler, forms designer for building GUI applications, web designer, class designer, and database schema designer. It accepts plug-ins that enhance the functionality at almost every level—

including adding support for source control systems (like Subversion) and adding new toolsets like editors and visual designers for domain-specific languages or toolsets for other aspects of the software development lifecycle.

C Sharp

C# is a multi-paradigm programming language encompassing strong typing, imperative, declarative, functional, generic, object-oriented (class-based), and component-oriented programming disciplines. It was developed by Microsoft within its .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO/IEC 23270:2006.

C# is one of the programming languages designed for the Common Language Infrastructure. C# is a general-purpose, object-oriented programming language. Its development team is led by Anders Hejlsberg. The most recent version is C# 7.2, which was released in 2017 along with Visual Studio 2017 version 15.5.

HTML & CSS

Hypertext Markup Language (HTML) is the standard markup language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript, it forms a triad of cornerstone technologies for the World Wide Web.

HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects such as interactive forms may be embedded into the rendered page. HTML provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items.

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language like HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

CSS is designed to enable the separation of presentation and content, including layout, colors, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple

web pages to share formatting by specifying the relevant CSS in a separate .css file and reduce complexity and repetition in the structural content.

6.2 Modules

6.2.1 Web Scraping Module:

This module is responsible for scraping reviews from a particular website. Module takes product details as input, gains access to the reviews page of that product and then scrapes the review to text form for local usage. This module is implemented in C#. This basically involves extracting only the relevant review part from the HTML file being fetched.

Code:

```
protected void btnSubmit_Click(object sender, EventArgs e)
 {
    ProductInfo.ProductName = DropDownList2.SelectedItem.Text;
    //Old Code
    #region
    string htmlCode;
    List<string> lst = new List<string>();

    string uri = String.Format("http://reviewing.net/s/prd/{0}", DropDownList2.SelectedItem);
    var getHtmlWeb = new HtmlWeb();
    var document = getHtmlWeb.Load(uri);
    Literal Literal1 = new Literal();
    Literal1.ID = "Literal1";
    foreach (HtmlNode para in document.DocumentNode.SelectNodes("//div[@class='review-summary']"))
    {
        Literal1.Text += String.Format(@"<p>{0}</p>", para.InnerText);
        D.Add((string)(para.InnerText));
    }
    // Panel1.Controls.Add(Literal1);
    Literal1.Text = Literal1.Text.Replace("<p>", " ");
    txtPara.Text = Literal1.Text.Replace("</p>", " ");
    allReview = "";
    allReview = txtPara.Text;
    #endregion

    //New Code
    #region
}
```

6.2.2 Parts-of-Speech Tagging Module:

POS tagging is done with the help of Maxent tagger. It is a tool created by Stanford University to tag various parts of a sentence as noun, verb etc.

```
private void TagReader(Reader reader)
{
    var tagger = new MaxentTagger(Model);
    //List obj = (List)MaxentTagger.tokenizeText(reader);
    foreach (ArrayList sentence in MaxentTagger.tokenizeText(reader).toArray())
    {
        var tSentence = tagger.tagSentence(sentence);
        System.Console.WriteLine(Sentence.listToString(tSentence, false));
        posstring = (Sentence.listToString(tSentence, false));
        newString = newString + posstring;
        System.Console.WriteLine();
    }

}
```

6.2.3 Feature Extraction Module:

In order to determine the sentiment polarity of an adjective describing an opinion feature, SentiWordNet is used, which is a lexical resource for opinion mining. SentiWordNet assigns three normalized sentiment scores: positivity, objectivity, and negativity to each synset of WordNet.

The SentiWordNet scores are assigned to the appropriate usage of adjective indicated as (P:0.625; O: 0.375; N: 0). Since the value of the positive polarity is highest, the adjective can be assigned a positive polarity.

The SentiWordNet score for over 35000 words are stored in a database.

	pos	Id	PosScore	NegScore	SynsetTerms
1	a	00001740	0.12500	0.00000	able
2	a	00002098	0.00000	0.75000	unable
3	a	00002312	0.00000	0.00000	dorsal
4	a	00002527	0.00000	0.00000	ventral
5	a	00002730	0.00000	0.00000	acroscopic
6	a	00002843	0.00000	0.00000	basiscopic
7	a	00002956	0.00000	0.00000	abducting
8	a	00003131	0.00000	0.00000	adductive

Query executed successfully. | HOME-PC (11.0 RTM) | Home-PC\Home (51) | master | 00:00:00 | 35552 rows

Table 6.1: Sentiwordnet Library

A sample FO table may look like this:

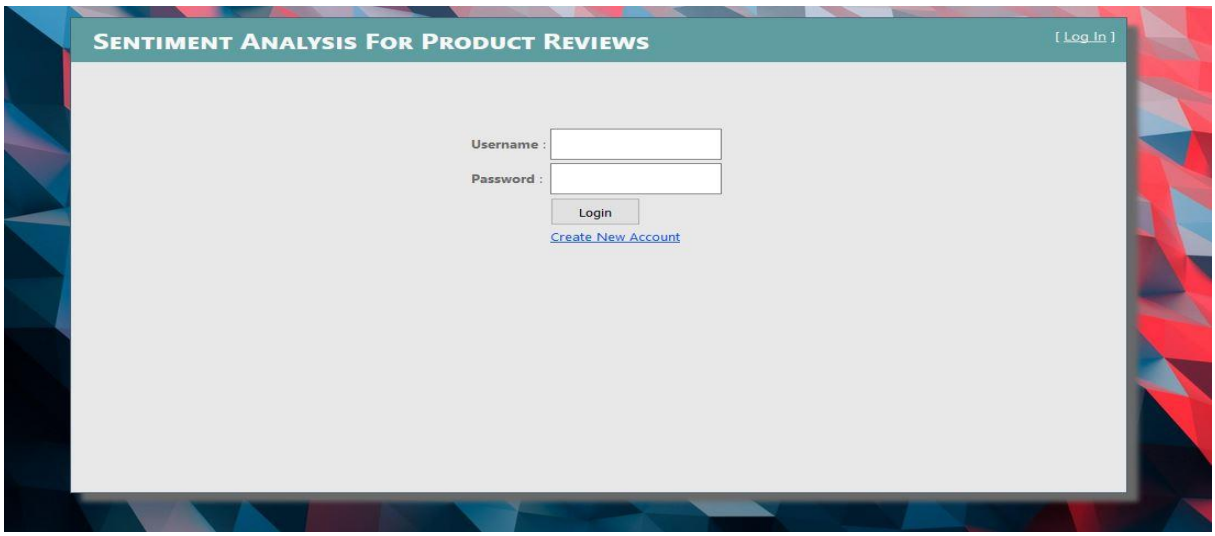
	id	Feature	PositivePolarity	NegativePolarity	Value	Weight
1	1	music	NULL	1	other	2
2	2	battery	1	NULL	Great	0.666666666666667
3	3	Video	1	NULL	helpful	0.666666666666667
4	4	camera	1	NULL	TIRED	0.666666666666667

Table 6.2: Feature Opinion Table

A feature-orientation table (FO table) is generated that records the opinion features and their corresponding descriptors of positive and negative polarities. The FO table, thus generated, enables us to generate feature-wise summary of a product or comparative summaries of different brands of similar products.

6.3 Snapshots

The first page of our review system is the login page. The login page provides functionality for user to login to the system and use the functions provided by our system. Also, if a new user wants to use the system, then he/she can create a new account from link given in the homepage.



This page shows the first stage in the application. The user is provided with the option to select what product they want to find the overall sentiment of. Products are decided on basis of Category selected and Brand Name selected.

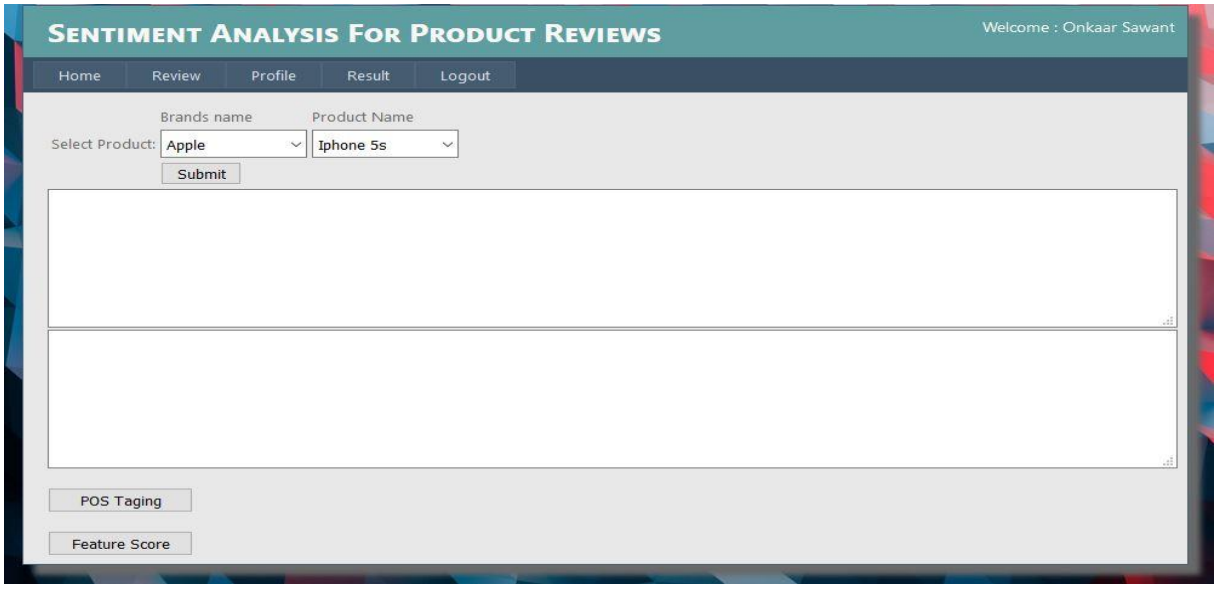


Fig 6.3: Web Scraping Module

Extracted Reviews:

The reviews for the product which was selected by the user is fetched and displayed on the screen. The fetched reviews are live and not stored locally on the database. Reviews are fetched dynamically based on user's request. The number of reviews fetched depends on the number of user reviews present online.

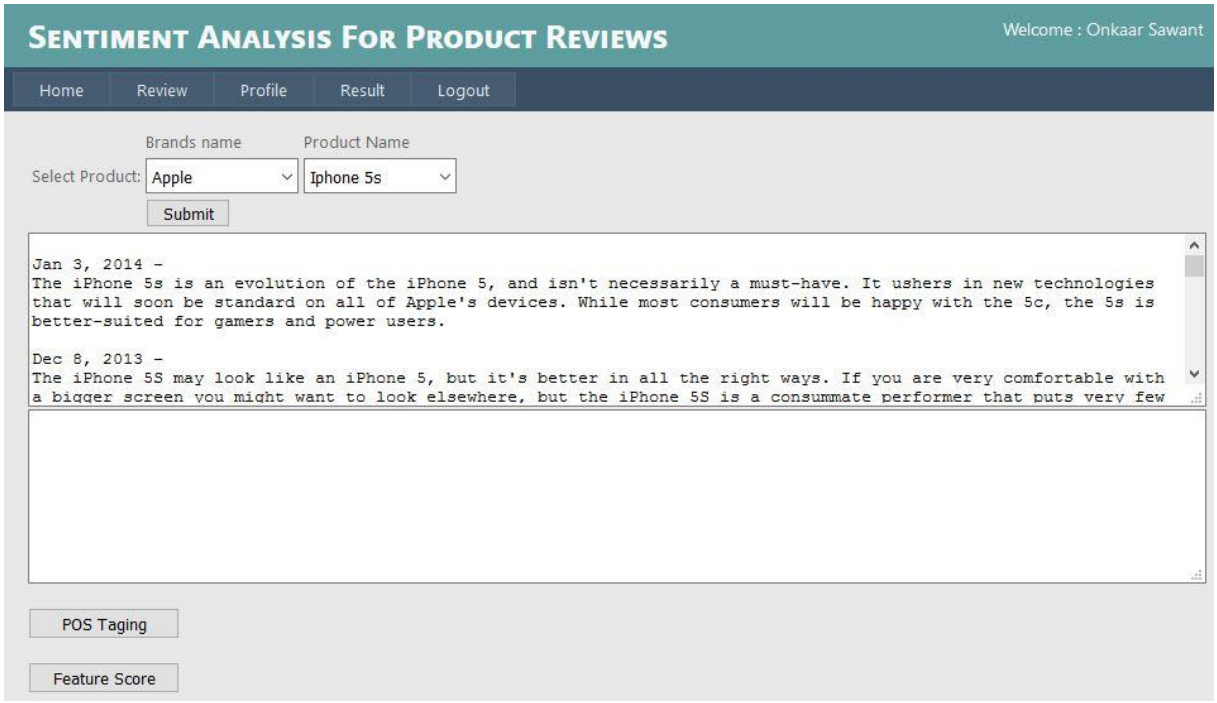


Fig 6.4: Output of Web Scraping Module

Reviews are scraped from a specified website. Scraping, though, requires user to pass parameter for the scraper to work. This is entirely handled by the system after user selects what product they want to see.

POS Tagging:

The POS tagging output is basically Pats-of-speech tagging of the reviews that were fetched previously. Each tag defines what part-of-speech it is.

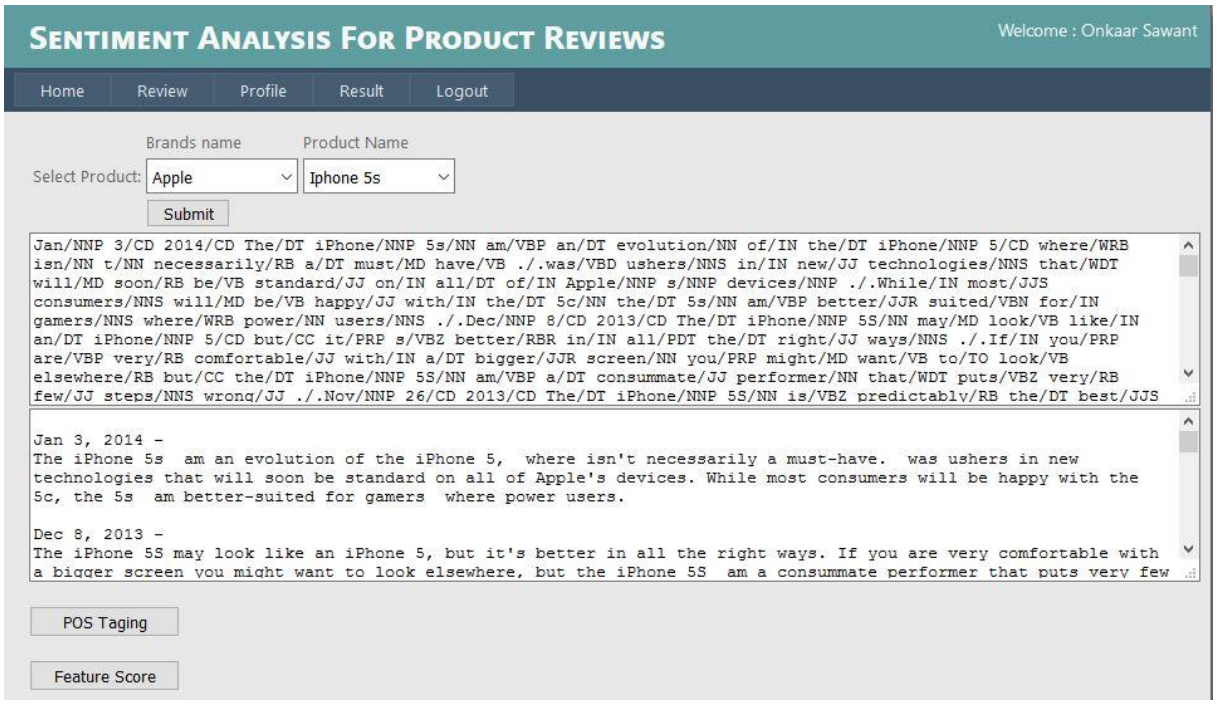


Fig 6.5: POS Tagging Module

POS tagger tags various tokens in the sentence as an adjective, noun, verb etc. For example,

DT stands for Determiner

NN stands for noun

JJ stands for adjective and so on.

Final Output:

The final output is shown in terms of overall sentiment of the reviews that were processed. A table is displayed which consists of the features that were talked about in the reviews fetched for the product, and also shows how many times the feature of the product was used in a positive or negative way.

A graph is also presented showing how the review sentiment score is distributed and the overall positivity score for the product is displayed.

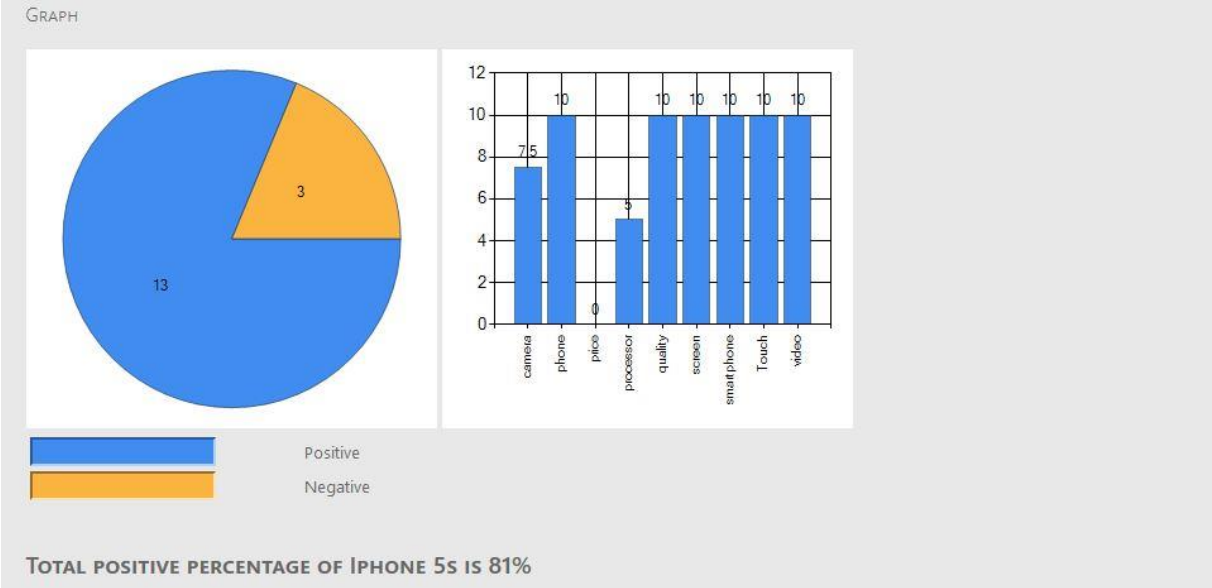


Fig 6.6: Feature Extraction Module

Chapter 7

Development of Test Cases

Test cases are used to test the system's performance and output under various circumstances. Here we test the different modules for all possible test cases, like incorrect password etc. to see whether the system fared well under those circumstances or not.

System testing is a critical phase implementation. Testing of the system involves hardware devise and debugging of the computer programs and testing information processing procedures. Testing can be done with text data, which attempts to stimulate all possible conditions that may arise during processing. If structured programming Methodologies have been adopted during coding the testing proceeds from higher level to lower level of program module until the entire program is tested as unit. The testing methods adopted during the testing of the system were unit testing and integrated testing.

UNIT TESTING:

Unit testing focuses first on the modules, independently of one another, to locate errors. This enables the tester to detect errors in coding and logical errors that is contained within that module alone. Those resulting from the interaction between modules are initially avoided.

Table 7.1: Test Cases

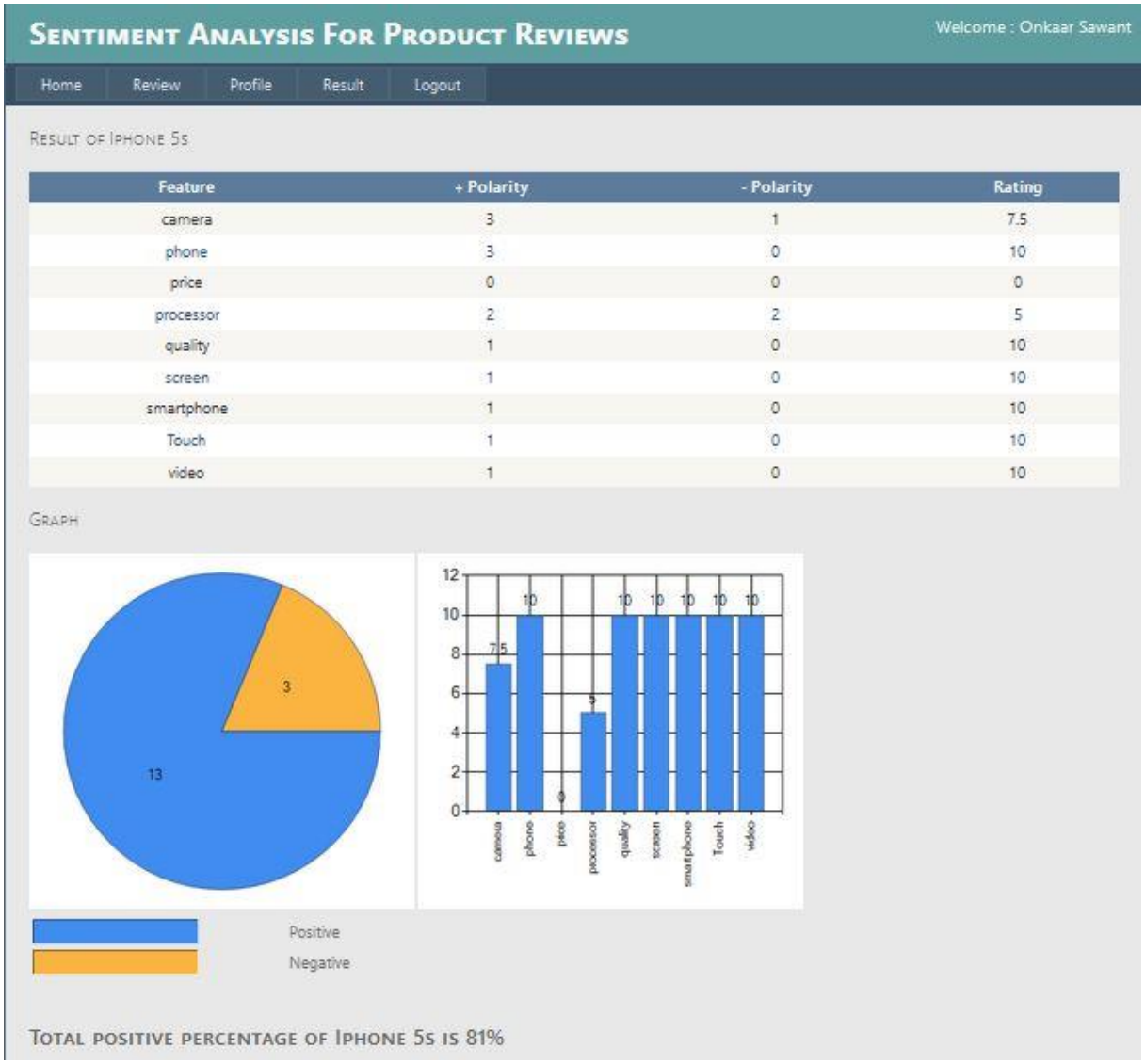
Test case id	Objective	Steps/Description	Input	Expected output	Actual output	Result	Remark
001	To test login credentials	Give null and wrong credentials to check if appropriate response is given or not.	Null values or incorrect credentials	Error message.	Error message and re-enter option.	Pass	Test passed.
002	To test Retrieval of product information	Select a product from available categories to check reviews are retrieved from website correctly	A product as a selection parameter from categories	Reviews of the particular product are retrieved from website	Reviews of the particular product are retrieved from website	Pass	Test passed.
003	To test POS tagger Component	Preprocessed reviews are tagged using POS tagger.	Preprocesse d review text	The text in the review is tagged appropriately	The text in the review is tagged appropriately	Pass	Test passed.
004	To test feature score generation	Tagged reviews are used to generate feature score and feature opinion table	Tagged reviews and feature words	Feature scored is calculated and feature opinion table is generated	Feature scored is calculated and feature opinion table is generated	Pass	Test passed.
005	To test creation of feature opinion graph	Contents of Feature-opinion table are represented graphically.	Feature Opinion Table	Pie chart with sentiment distribution.	Pie chart with proper representation of positivity and negativity.	Pass	Test passed.

Chapter 8

Results and Analysis

8.1 Results

In this project, we have successfully implemented a E-commerce product review management system. The application has been developed for gaining and understanding the overall sentiment of a users's product review. The application successfully validates the users. The application is also able to fetch reviews online and present it to the user. The final output of the application is the overall sentiment of the review, showing the keywords which affected the decisions, and present a graphical representation of the same. A bar graph of the rating of various features of the product is also displayed. The bar graph helps determine how good a certain feature of the product was liked by the user.



8.2 Analysis

In this project, the first step for the user is to create/login to his/her account. After logging in to the application, the user can select the product of his choice by browsing through the drop-down menus. After selecting the choice, the application then fetches the reviews, displays it to the user in a text field view. The user then should select the “POS Tagging” option so that POS tagging can be performed on the reviews.

POS tagging is the process of marking up a word in a text (corpus) as corresponding to a part of speech, based on both its definition and its context. POS tagging identifies the words as nouns, verbs, adjectives, adverbs, etc.

After this, the user should select “Feature Score” option which gives a sentiment score for the nouns and adjectives tagged by the POS tagger. In order to determine the sentiment polarity, we make use of SentiWordNet. Hence, according to these scores, a feature opinion table as well as a graph representing positive to negative score are created. Also, the total percentage of positive reviews for the selected product is also displayed. This information help the user to determine whether the product is worth to buy or not.

Chapter 9

Conclusions and Future Scope

Conclusion:

Classifying and summarizing opinions of reviewers has several interesting and commercially significant applications. However, this task is much more difficult than classifying regular text and requires intensive Preprocessing. The success of the opinion mining task is mainly dependent on the efficiency and sophistication of the Preprocessing and feature extraction steps.

Results indicate that the multistep feature based semi supervised opinion mining approach used in this project can successfully identify opinionated sentences from unstructured user reviews and classify their orientation with acceptable accuracy. This enables reliable review opinion summarization which has several commercially important applications.

Future Scope:

In the future, we want to perform opinion mining on larger and more varied blog data sets. We would also like to extend our work to fuzzy opinion classification with support for fuzzy user querying. We intend to do this by learning the strength of various adjective descriptors along with corresponding linguistic hedges and include them in the feature-orientation table generated during the mining process.

The classification technique proposed in the paper can be naturally extended to support fuzzy classification.

References

[1] L. Zhao and C. Li, "Ontology based opinion mining for movie reviews," in Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management, pp. 204–214, 2009.

[2] A. Balahur, Z. Kozareva, and A. Montoyo, "Determining the polarity and source of opinions expressed in political debates," in Proceedings of the 10th International Conference on Intelligent Text Processing and Computational Linguistics, vol. 5449 of Lecture Notes in Computer Science, pp. 468–480, Springer, 2009.

[3] Y. H. Gu and S. J. Yoo, "Mining popular menu items of a restaurant from web reviews," in Proceedings of the International Conference on Web Information Systems and Mining (WISM'11), vol. 6988 of Lecture Notes in Computer Science, pp. 242–250, Springer, 2011.

[4] M.Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the 10th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04), pp. 168–177, August 2004.

[5] M. A. Jahiruddin, M. N. Doja, and T. Ahmad, "Feature and opinion mining for customer review summarization," in Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence (PReMI '09), vol. 5909 of Lecture Notes in Computer Science, pp. 219–224, 2009.

[6] S. Shi and Y. Wang, "A product features mining method based on association rules and the degree of property co-occurrence," in Proceedings of the International Conference on Computer Science and Network Technology (ICCSNT '11), pp. 1190–1194, December 2011.

[7] S. Huang, X. Liu, X. Peng, and Z. Niu, "Fine-grained product features extraction and categorization in reviews opinion mining," in Proceedings of the 12th IEEE International Conference on Data Mining Workshops (ICDMW'12), pp. 680–686, 2012.

[8] C.-P.Wei,Y.-M.Chen,C.-S.Yang, and C. C.Yang, "Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews," Information Systems and e-Business Management, vol. 8, no. 2, pp. 149–167, 2010.

Acknowledgement

We are personally indebted to a number of people who gave us their useful insights to aid in our overall progress for this project. A complete acknowledgement would therefore be encyclopedic. First of all, we would like to give our deepest gratitude to **our parents** for permitting us to take up this course.

Our sincere thanks and heartfelt sense of gratitude goes to our respected Principal, **Dr. Bhavesh Patel** for all his efforts and administration in educating us in his premiere institution.

We take this opportunity to convey our sincere thanks to our Head of the Department, **Prof. Uday Bhave** and our project mentor, **Prof. Rupali Kale** for her constant encouragement and invaluable suggestions throughout the project and for his technical support rendered during the course of our project.