A PRELIMENERY REPORT ON

PRODUCT RANKING USING PRODUCT RANK ALGORITHM ALONG WITH SENTIMENT ANALYSIS ON REVIEWS ON E-COMMERCE WEBSITE

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

SUYASH LOKHANDE	**71814623K**
SARVESH DHAKUNKAR	**71814354L**
SUYASH CHOPDA	**71926361B**
VEDANT SHARMA	**71815045H**

DEPARTMENT OF COMPUTER ENGINEERING

DR. D. Y. PATIL INSTITUTE OF TECHNOLOGY

PIMPRI, PUNE 411018

SAVITRIBAI PHULE PUNE UNIVERSITY

2020-2021

CERTIFICATE

This is to certify the project report entitles

"PRODUCT RANKING USING PRODUCT RANK ALGORITHM ALONG WITH SENTIMENT ANALYSIS ON REVIEWS ON E-COMMERCE WEBSITE"

Submitted by

SUYASH LOKHANDE	**71814623K**
SARVESH DHAKUNKAR	**71814354L**
SUYASH CHOPDA	**71926361B**
VEDANT SHARMA	**71815045H**

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Prof. B. S. Satpute** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

(Prof. B. S. Satpute)
Guide,
Department of Computer Engineering

(Dr. S.V. Chobe)
Head,
Department of Computer Engineering

(Dr. Pramod Patil)
Principal,
Dr. D. Y. Patil Institute of Technology Pune -18

Place: Pune
Date: 25-February-2021

ACKNOWLEDGEMENT

It is a matter of great pleasure by getting the opportunity of highlighting. A fraction of knowledge, we acquired during our technical education through this Project.

This would not have been possible without the guidance and help of many people. This is the only page where we have the opportunity of expressing our emotions and gratitude from the core of our heart to them. This project would not have been success without enlightened ideas, timely suggestion and interest of our most respected guide **"Prof. B. S. Satpute"**.

We would like to thank **"Dr. S. V. Chobe"** Head of Department of our institution for providing necessary facility during the period of working on this project work.

We would also like to thank our principal **"Dr. Pramod Patil"** who encouraged us and created healthy environment for all of us to learn in best possible way. Finally, we would like to pay our respect and love to our parents and all other family members as well as friends for their love and encouragement throughout our career.

NAME OF THE STUDENTS

SUYASH LOKHANDE

SARVESH DHAKUNKAR

SUYASH CHOPDA

VEDANT SHARMA

ABSTRACT

In today's world, e-commerce is the strongest and the fastest growing platform. People inclined to buy products from websites now and then. They are spending their lot of time in searching best product and applying filters. Up to the present time price comparisons of the same product on different websites are available but now, moving a step ahead. The purpose of this study is to develop a system tends to serve as a smart web spider for all e-commerce websites, which crawls the whole website and find out the cheapest yet best quality products in real time. Moreover, this assures the union of all e-commerce website in form of one rigorously filtered system. Using a product ranking algorithm this system analyzed cheapest yet best quality products. It is not only ranking products but also considering reviews of customers on each product and present a sentiment analysis approach on these reviews which involves quantification of the sentiments of thousands of reviews. And finally, providing the end user summarized data about the expressed sentiments in way of intuitive and easy to understand visualization technique. But this system is not stopping here, as with increasing expenditure people are also concerned about the prices that how cheap they can get the products. As every time the prices keep on fluctuating and the prices are not stable for any product so keeping a track to them is also important. So, this system helps the user to keep a track on the price of the product and it will provide the notification to the user that when the price of a particular product has been dropped. Due to this user does not need to check the e-commerce website continuously and thus it will save a lot of time of the user.

TABLE OF CONTENTS

LIST OF ABBREVATIONS

ABBREVIATION	ILLUSTRATION
SDLC	Software Development Life Cycle
SQL	Structured Query Language

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

E-Commerce is the world's fastest growing platform where nearly billions of products are purchased every day. Companies like Amazon, Flipkart, Myntra and Snapdeal are offering customers with online grocery, apparel and entertainment shopping and makes trillions of profits every year. Most of them are during festive seasons. At each website, there are millions of products are available, and customers are not able to find products which are offered at low prices, with more discount and best quality. Even filters fail to sort the products from low-high rate due to large and dynamic set of databases. Therefore, it is necessary to filter out unnecessary and overpriced products and provide customers, especially the late adopters with few quality products available on websites. Not only millions of products but also a lot of reviews are there for each product. Today, the average consumer relies heavily on user reviews and opinions expressed on the product before deciding to buy. It may be either positive or negative towards the product. So, reading all those reviews and then deciding to buy a product is very tedious and time-consuming.

This time can be saved by developing a system which automatically gives the best product of choice and saves a lot of time. It's time to make a machine to do what humans are doing now. In this age of increasing machine learning-based algorithms reading thousands of reviews to understand a product is rather time-consuming where we can polarize a discussion on a particular category to understand its popularity among the buyers all over the world.

1.1 MOTIVATION

Currently there is a sudden growth in the number of people switching to online shopping but due to large availability of product it is difficult for customers to decide which product to buy without an "information at your fingertips" type of application.

A well-designed application that compares all product under one platform is the need of the hour as it provides large improvement over the current methods of manually checking different sites and wasting both time and effort of the customer.

The current method would benefit from a platform where all products can be compared along with the recommendation for the optimal product. Additionally, the system is going to perform sentiment analysis on the reviews and will tell the customers the polarity of product. Further, it will improve user experience as it would notify the user whenever there is a price drop in a product of their choice by tracking the price of the product.

1.2 PROBLEM DEFINITION

The system tends to serve as a smart web spider for e-commerce websites, which crawls the whole website and find out the cheapest yet best quality products with good reviews in real time.

2. LITERATURE SURVEY

Customers prefer online shopping rather than offline shopping because of better price and product availability. Due to an increase in healthy competition in e-commerce market, it is also important to know which e-commerce website is most ideal, in terms of quality and price, amongst all. [1]

Products on E-commerce websites are growing rapidly day by day. Many comparison websites are available on the internet which compares the price of a product with different websites. Most of them are operated manually thus it becomes very tedious to update each product manually. Thus, it is very much important to have a robust automated search engine available in which products can be compared with an automated engine with minimal human intervention. [2]

Up to the present time price comparisons of the same product on different websites are available but now, moving a step ahead. To develop a system tends to serve as a smart web spider for all e-commerce websites, which crawls the whole website and find out the cheapest yet best quality products in real time. [3]

Research aims to automate the process of gathering online, end user reviews for any given product or service and analyzing those reviews in terms of the sentiments expressed about specific features. [4]

Analyzes the reviews of customers about a restaurant and will predict based on these reviews whether the customer has liked the product of the restaurant or not. That means whether it is a positive review or negative review, based on the available text review. The prediction will be carried out by classification models and will find out which classification model is best in this task of prediction.[5]

Work concentrates on mining reviews from the websites like Amazon, which allows user to freely write the view. It automatically extracts the reviews from the website. It also uses algorithm such as Naïve Bayes classifier, Logistic Regression and SentiWordNet algorithm to classify the review as positive and negative review. [6]

In this, focuses tools and techniques used in opinion mining. The process of opinion summarization has three main steps, such as “Opinion Retrieval, Opinion Classification and Opinion Summarization.” User comments are retrieved from review websites. These comments contain subjective information and they are classified as positive or negative review. Depending upon the frequency of occurrences of features opinion summary is created. [7]

3. SOFTWARE REQUIREMENTS SPECIFICATIONS

3.1 INTRODUCTION

E-commerce in India is the fastest growing platform where nearly more than thousands of products are purchased in every second. Indian e-commerce is growing at an annual rate of 51%, the highest in the world, and is expected to jump from $3.535 trillion in 2019 to $6.542 trillion in 2023. Companies like Amazon, Flipkart, Myntra and Snapdeal makes trillions of profits every year and most of them are during festive seasons. At each website, there are millions of products are available and customers are not able to find products which are really offered at cheap prices, with more discount and best quality. Even filters fail to sort the products from low-high rate due to large and dynamic set of databases. Therefore, it is necessary to filter out unnecessary and overpriced products and provide customers with few quality products available on websites. Not only millions of products but also lot of reviews are there for each product. Today, the average consumer relies heavily on user reviews and opinions expressed on product before making a decision to buy. It may be either positive or negative towards the product. So, reading all those reviews and then deciding to buy a product is very tedious and time consuming.

This time can be saved by developing a system which automatically gives a best product of choice and save lot of time. It's time to make machine to do what humans are doing now. In this age of increasing machine learning based algorithms reading thousands of reviews to understand a product is rather time consuming where we can polarize a review on particular category to understand its popularity among the buyers all over the world. As the growing expenditure people are also concerned about the prices as there is no-one from whom they can negotiate for the prices. Due to this people keep on adding products in their wish list and waiting till the price is dropped is only the option.

The proposed system merges all the products available at different websites and filters out the best product for the customer. Using web crawlers, only those products are available to a customer who comes under predefined parameters like price, brand, product discount and popularity of the product. System is also developed to keep a real time check on the

prices of the products. This system will help the user to buy products when the prices are dropped and it will save a lot of time as user doesn't need to continuously check the website as this system will keep a real time track of the price and it will provide notifications to the user when the price of a product is changed.

3.1.1 Project Scope

The project scope is to find best quality and cheapest product desired by user. Also give some additional features like word cloud analysis or reviews to user.

3.2 FUNCTIONAL REQUIREMENTS

3.2.1 User Should be Able to Enter Product Name

User will be provided with search box in which user will enter product name, user will also be having option of inputting range of price(optional) to get more desired and expected product.

3.2.2 Ranked Product list will be Generated

After searching user will get ranked product list, where he/she find most suitable product at top.

3.2.3 Option of Word Cloud and Buy

User will have option of word cloud for each product to analyze reviews of products. User also can buy products from respective site directly.

3.3 EXTERNAL INTERFACE REQUIREMENTS

3.3.1 User Interfaces

The interface will meet the following requirements to conform to the user's needs.

1. It will be simple and easy to understand.
2. Controls which allow the user to interact with the application will be clear and imply their functionality within the application.
3. The interface will include user inputs as well as graphics.
4. The graphics displayed to the user will provide a visual representation of the output produced.

3.3.2 Hardware Interfaces

The application is intended to be a stand-alone, single-user system. The application will run on a Laptop. No further hardware devices or interfaces will be required.

3.3.3 Software Interfaces

1. Inputs: The software will receive input from One source. First, the user interface.
2. The user interface will supply the input method.
3. Outputs: The output will be the best product among all search result of user search.
4. Operating System: Windows.

3.4 NONFUNCTIONAL REQUIREMENTS

3.4.1 Performance Requirement

Performance of the functions and every module must be well. The overall performance of the software will enable the users to work efficiently.

3.4.2 Safety Requirements

The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

3.4.3 Security Requirements

Users are authenticated using many security phases so reliable security is provided.

3.4.4 Software Quality Attributes

Our software has many quality attributes that are given below: -

1. Adaptability: This software is adaptable by all users.
2. Availability: This software is freely available to all users. The availability of the software is easy for everyone.
3. Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.
4. Reliability: The performance of the software is better which will increase the reliability of the Software.

5. User Friendliness: Since, the software is a GUI application; the output generated is much user friendly in its behavior.
6. Integrity: Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled.

3.5 SYSTEM REQUIREMENTS

3.5.1 Database Requirement

1. SQL
2. PhpMyAdmin
3. Xammp Server

3.5.2 Software Requirement

1. Operating System: Windows 10
2. IDE: PyCharm
3. Programming Language: Python

3.5.3 Hardware Requirement

1. Hardware: intel core
2. Speed: 2.80 GHz
3. RAM: 4GB
4. Key Board: Standard Windows Keyboard

3.6 ANALYSIS MODELS: SDLC MODEL TO BE APPLIED

An effective System Development Life Cycle (SDLC) should result in a high-quality system that meets customer expectations, reaches completion within time and cost evaluations, and works effectively and efficiently in the current and planned Information Technology infrastructure. System Development Life Cycle (SDLC) is a conceptual model which includes policies and procedures for developing or altering systems throughout their life cycles. SDLC is used by analysts to develop an information system.

SDLC includes the following activities:

• Requirements

• Design

• Implementation

• Testing

• Deployment

• Operations

• Maintenance

Phases of SDLC: Systems Development Life Cycle is a systematic approach which explicitly breaks down the work into phases that are required to implement either new or modified Information System.

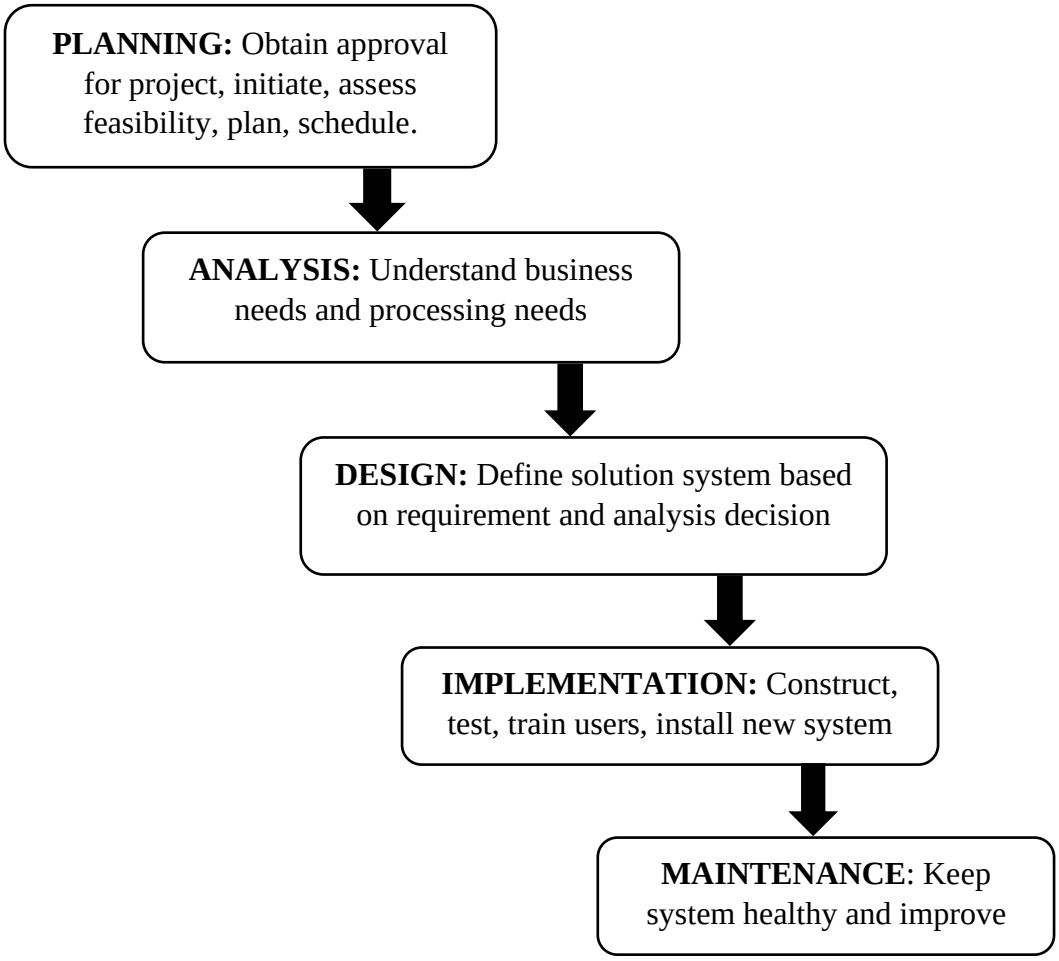

Fig No. 3.1 Phases of SDLC

3.7 SYSTEM IMPLEMENTATION PLAN

Sr. No.	Name/Title	Start Date	End Date
1	Preliminary Survey	14-08-2020	20-08-2020
2	Introduction and Problem Statement	21-08-2020	26-08-2020
3	Literature Survey	27-08-2020	25-09-2020
4	Project Statement	26-09-2020	29-09-2020
5	Software Requirement and Specification	30-09-2020	14-10-2020
6	System Design	14-10-2020	11-11-2020
7	Partial Report Submission	12-11-2020	20-11-2020
8	Architecture Design	21-11-2020	27-11-2020
9	Implementation	30-11-2020	13-12-2020
10	Testing	14-12-2020	01-01-2021
11	Report Submission	20-02-2021	25-02-2021

4. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

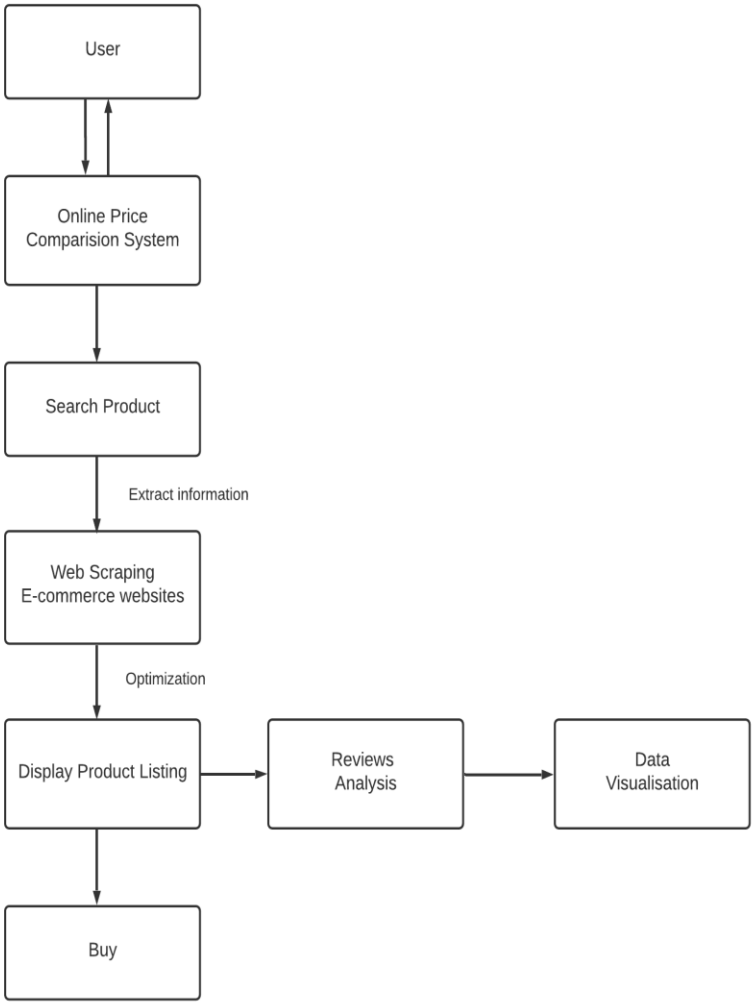

Fig 4.1 System Architecture

4.2 DATA FLOW DIAGRAMS

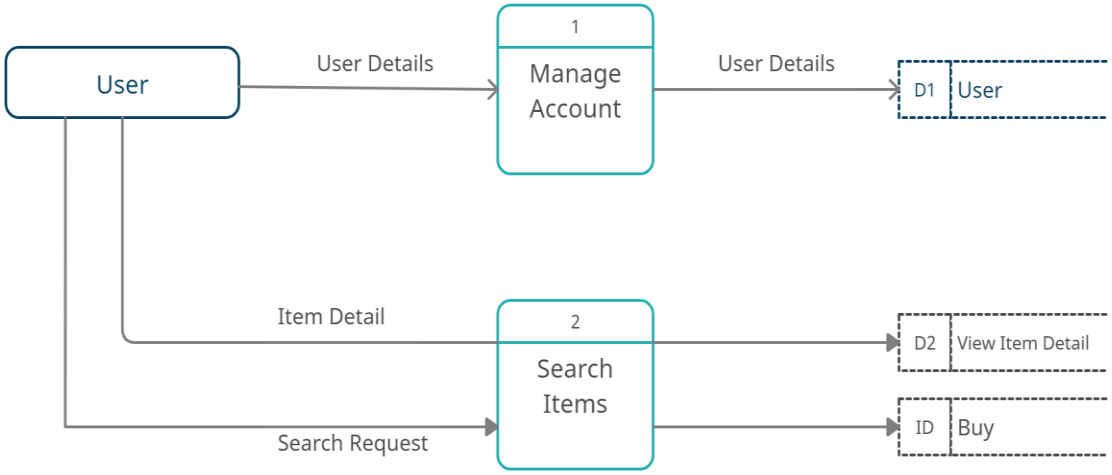

Fig No. 4.2 DFD Level-0

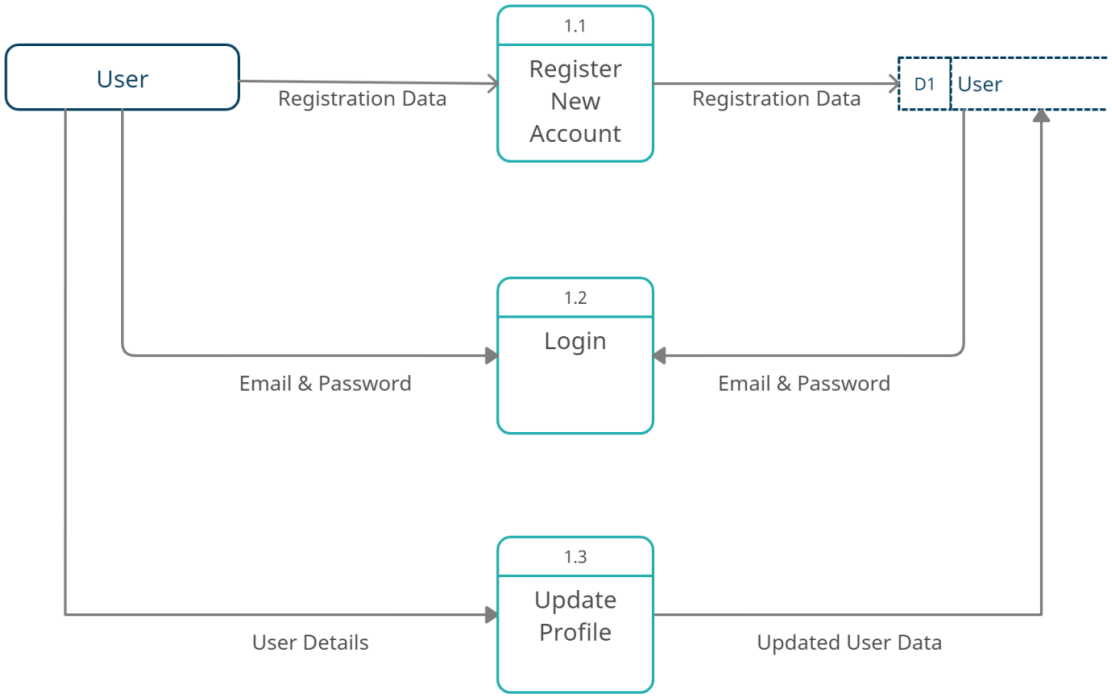

Fig No. 4.3 DFD Level-1: Manage Account

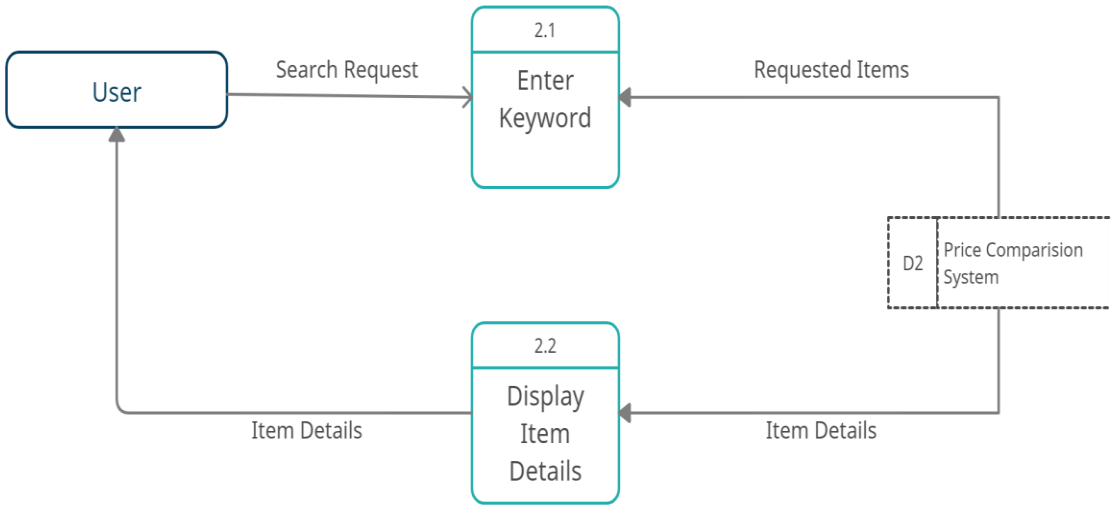

Fig No. 4.4 DFD Level-1: Search Item

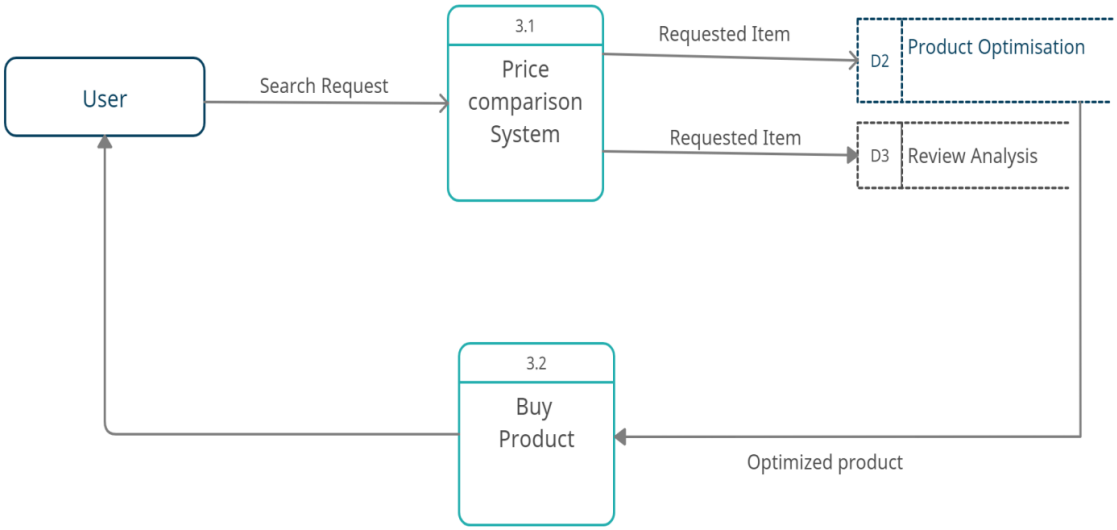

Fig No. 4.5 DFD Level-2: Price Comparison System

4.3 Entity Relationship Diagrams

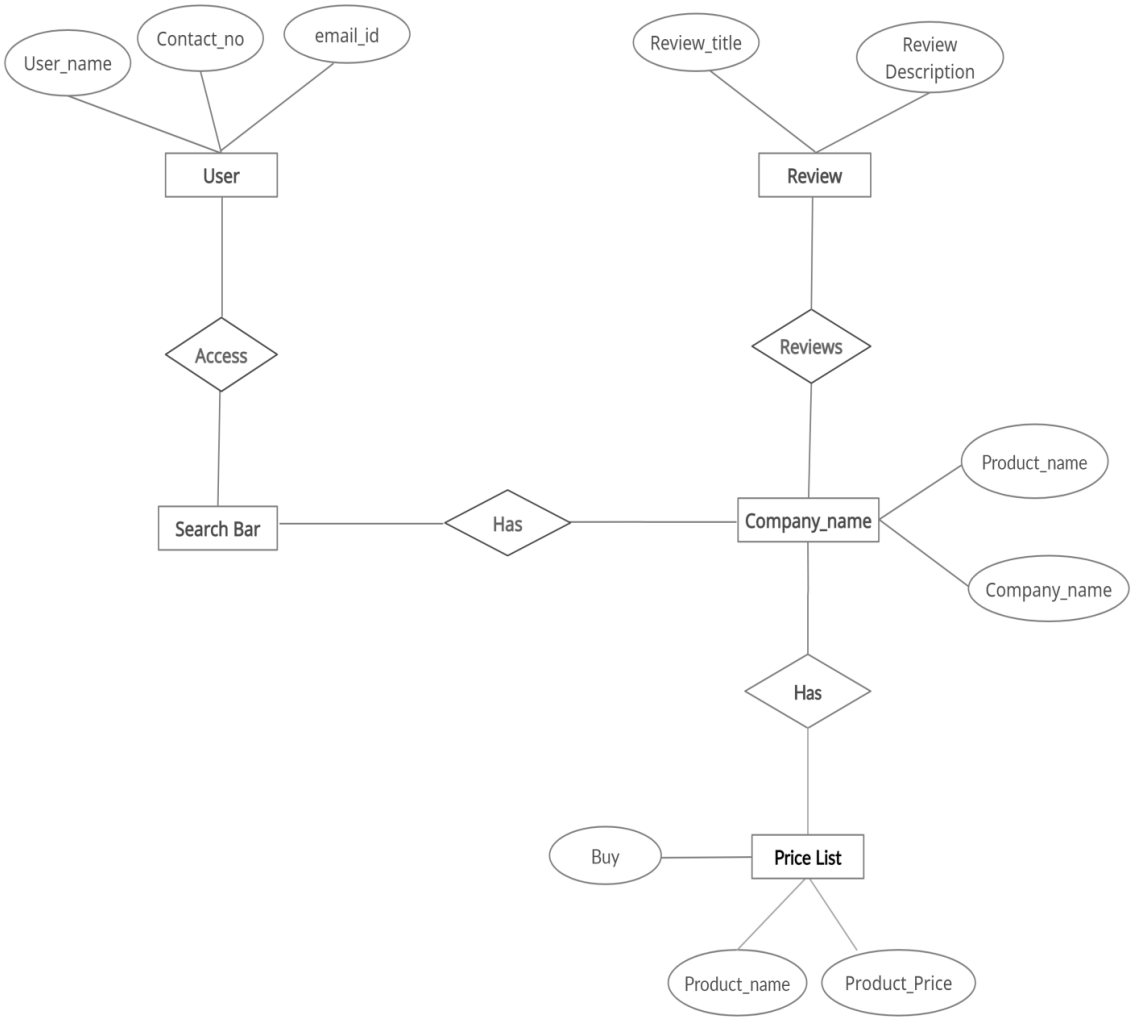

Fig No. 4.6 ER Diagram

5. OTHER SPECIFICATION

5.1 ADVANTAGES

1. System provides fast searching platform with minimum inputs.
2. System will suggest optimum product according to user's requirement by considering price, popularity, ratings, reviews and features
3. System will actively track and notify price of products across multiple e-commerce sites.
4. System will provide sentiment analysis on reviews of products with proper visualization.
5. Most importantly, system will save lots of time and efforts of user as manual work done by them will be automated.

5.2 LIMITATIONS

1. Internet connection is mandatory to use the system.
2. User can see only top 10 ranked products (not all products).

5.3 APPLICATIONS

1. No need to waste time finding the lowest prices online & best deals.
2. Prompt help in online shopping
3. Smooth and intuitive user experience.
4. One Click Online Price Comparison feature.
5. User can see the negative and positive reviews on WordCloud

6. CONCLUSION & FUTURE WORK

Due to the increase of numerous products in e-commerce, customers are unable to find proper product or in many of the cases customers buy overpriced product which is non beneficial. In order to increase customers, experience at e-commerce, filtration of unnecessary products is needed and listings of quality products are needed and analysing review to give one shot result is also needed. This system will provide customers the crystal-clear view of the searched product and provide cheapest product yet best quality product in one click. System also provides facility to track product price. This system will save human's time and money.

With the inrush of the big data era, the facts on the web have become way more dynamic, leading to better quality in Ranking and Sentiment Analysis. Therefore, as future work, we need to either design or use new product ranking algorithms that meet global standards of ranking challenges with efficiency and will provide users with both user specific as well as relevant information quickly and swiftly. Also, we can include local market sites along with retail sites like Flipkart, Amazon, Myntra etc. to get better accuracy and usability. In this study, we used Word Cloud and Naïve Bayes for sentiment analysis without further modification. Still, in future work, it will be necessary to adjust tools and approaches to the sentiment analysis, particularly for mixed emotions, since our manual observations indicate that users actively discuss the pros and cons of a product. So, adding additional features, we can get the exact perception of the product by the user. And lastly, we will try to continue this research until we generalize it to all kinds of text-based reviews and comments.

APPENDIX A

Np-hard Np-Complete:

What is P?

• P is set of all decision problems which can be solved in polynomial time by a deterministic.

• Since it can be solved in polynomial time, it can be verified in polynomial time.

• Therefore, P is a subset of NP.

P: The increasing number of Dump and Deaf People Problem that language not knowing properly.

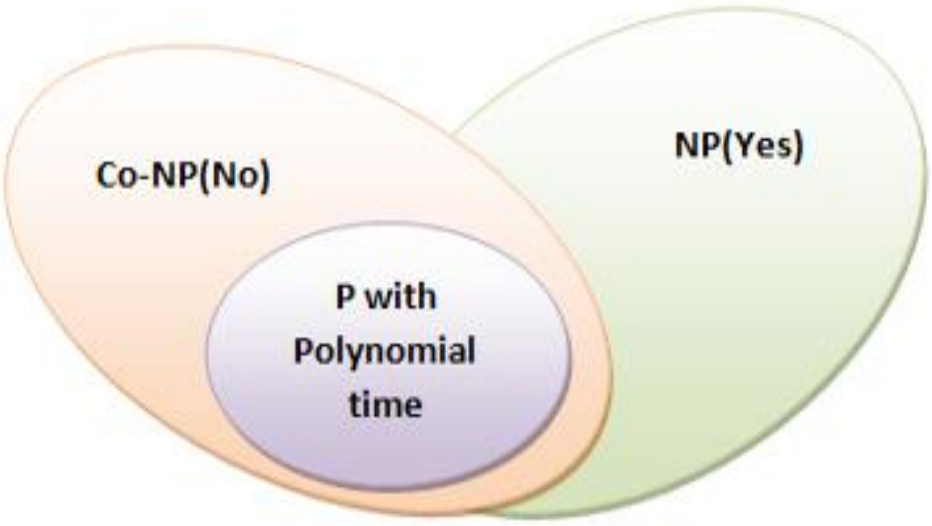

Fig No. 6.1. P

What is NP?

NP means we can solve it in polynomial time if we can break the normal rules of step-by step computing". What is NP Hard? A problem is NP-hard if an algorithm for solving it can be translated into one for solving any NP-problem (nondeterministic polynomial time) problem. NP-hard therefore means at least as hard as any NP-problem," although it might, in fact, be harder.

NP-Hard:

So here in this case the ‘P’ problem is NP hard i.e., P=NP-Hard

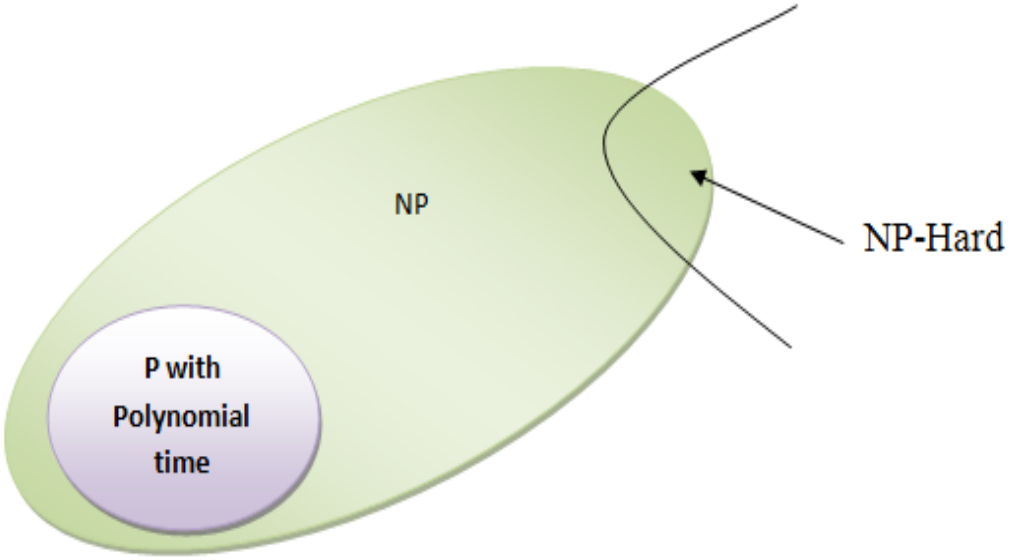

Fig No. 6.2 NP-Hard

APPENDIX B

1. *Sergey Brin and Lawrence Page, The Anatomy of a Large- Scale Hypertextual Web Search Engine. Computer Science Department, Stanford University, Stanford, CA 94305, USA, pp.1-20 1999.*

   Products on E-commerce websites are growing rapidly day by day. Many comparison websites are available on the internet which compares the price of a product with different websites. Most of them are operated manually thus it becomes very tedious to update each product manually. Thus, it is very much important to have a robust automated search engine available in which products can be compared with an automated engine with minimal human intervention.

2. *Urvesh Rathod, Prof. Aruna Pavate, Vaibhav Patil. Product Rank Based Search Engine for E-Commerce. 3rd International Conference for Convergence in Technology (I2CT), IEEE, Apr 2018 DOI: 10.1109/I2CT.2018.8529503*

   Up to the present time price comparisons of the same product on different websites are available but now, moving a step ahead. To develop a system tends to serve as a smart web spider for all e-commerce websites, which crawls the whole website and find out the cheapest yet best quality products in real time.

3. *Prashast Kumar Singh, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Amit Sharma, an approach towards features specific opinion mining and sentimental analysis across e-commerce websites, 5th International Conference- Confluence the Next Generation Information Technology Summit (Confluenc),2014*

   Research aims to automate the process of gathering online, end user reviews for any given product or service and analyzing those reviews in terms of the sentiments expressed about specific features.

4. *Santhosh Kumar K L, Jayanti Desai, Jharna Majumdar, Opinion Mining and Sentiment Analysis on Online Customer Review, IEEE International Conference on Computational Intelligence and Computing Research, 2016*

   Work concentrates on mining reviews from the websites like Amazon, which allows user to freely write the view. It automatically extracts the reviews from the website. It also uses algorithm such as Naïve Bayes classifier, Logistic Regression and SentiWordNet algorithm to classify the review as positive and negative review.

APPENDIX C

Plagiarism Checker X - Report

Originality Assessment

Overall Similarity: **5%**

Date: Feb 25, 2021

Statistics: 133 words Plagiarized / 2665 Total words

Remarks: Moderate similarity detected; you better improve the document (if required).

REFERENCES

1. https://www.ibef.org/industry/ecommerce.aspx 07th December 2017
2. Sergey Brin and Lawrence Page, The Anatomy of a Large- Scale Hypertextual Web Search Engine. Computer Science Department, Stanford University, Stanford, CA 94305, USA, pp.1-20 1999.
3. Urvesh Rathod, Prof. Aruna Pavate, Vaibhav Patil. Product Rank Based Search Engine for E-Commerce. 3rd International Conference for Convergence in Technology (I2CT), IEEE, Apr 2018 DOI: 10.1109/I2CT.2018.8529503
4. Prashast Kumar Singh, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Amit Sharma, an approach towards features specific opinion mining and sentimental analysis across e-commerce websites, 5th International Conference- Confluence the Next Generation Information Technology Summit (Confluenc),2014
5. https://analyticsindiamag.com/step-by-step-guide-to-reviews-classification-using-svc-naive-bayes-random-forest/
6. Santhosh Kumar K L, Jayanti Desai, Jharna Majumdar, Opinion Mining and Sentiment Analysis on Online Customer Review, IEEE International Conference on Computational Intelligence and Computing Research, 2016
7. G.Angulakshmi, Dr.R.ManickaChezian, An Analysis on Opinion Mining: Techniques and Tools. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 7, July 2014.