ABSTRACT

Data mining has emerged from just an offshoot of simple database querying techniques to a full blown information science major in Computer science field. The primary purpose of data mining is to find useful information from a huge amount of data in order to apply the same in various applications. Web data mining is a fast growing field of data analytics and finds its usage in various applications. Web data mining involves secure collection of information from online web activity and putting it into implementation for business purposes. It involves navigating through thousands of web pages on the internet and gaining intelligence on the web URL patterns observed pertaining to desired business logic. Previously, the business trends were observed manually from the collection of data in the records. As the amount of data increases, the ability to discern useful data from the unnecessary data becomes difficult. Hence, Web Data Mining is of importance for business in today’s world.

In this project, a prototype of an e-commerce web application has been implemented which makes the usage of web mining to improve the user experience of an e-commerce website. Using the application server log information, it enables the admin to determine the products they need to sell in order to maximize the profit ratio. They can also use these statistical information to determine various business performance indices such as the most popular product, the most viewed product, products viewed by category etc. In short, today business logic runs on catching user’s attention. This project makes an attempt to show how a business strategy can be converted into a business logic with the help of web data mining which can be used to reap benefits in the e-commerce domain.
# TABLE OF CONTENTS

Abstract ........................................................................................................................................... ii
Table of Contents .......................................................................................................................... iii
List of Figures .................................................................................................................................... vi
1. Background and Rationale ........................................................................................................... 1
   1.1 Need for Web Data Mining .................................................................................................... 1
   1.2 What is Data Mining ............................................................................................................. 1
   1.3 What is Web Data Mining .................................................................................................. 2
   1.4 Web Usage Mining .............................................................................................................. 2
   1.5 Web Structure Mining ....................................................................................................... 3
   1.6 Web Content Mining ......................................................................................................... 4
2. Narrative ...................................................................................................................................... 5
   2.1 Problem Statement ................................................................................................................. 5
   2.2 Motivation ............................................................................................................................. 5
   2.3 Project Objective .................................................................................................................. 6
   2.4 System Requirements .......................................................................................................... 6
   2.5 Hardware Requirements ...................................................................................................... 6
   2.6 Software Requirements ....................................................................................................... 6
3. Proposed System Design .............................................................................................................. 7
   3.1 System Flowchart .................................................................................................................. 8
      3.1.1 Hit Count Tracking ......................................................................................................... 9
      3.1.2 Page Rank Algorithm .................................................................................................. 10
   3.2 Framework ............................................................................................................................ 12
   3.3 Steps to Project Development ............................................................................................. 12
Evaluation ........................................................................................................30
  5.1 User Interface .................................................................................................31
    5.1.1 User View .................................................................................................31
      5.1.1.1 User Registration .................................................................................32
      5.1.1.2 User Login ..........................................................................................32
      5.1.1.3 Searching Products .............................................................................33
      5.1.1.4 Purchasing Products ............................................................................33
      5.1.1.5 Purchase Summary ..............................................................................34
    5.1.2 Admin View ...............................................................................................35
      5.1.2.1 Pending Orders ....................................................................................35
      5.1.2.2 Approved Orders ................................................................................35
      5.1.2.3 Delivered Orders ................................................................................36
    5.1.3 Add/Edit Product Information ..................................................................37
    5.1.4 Reports ......................................................................................................38
      5.1.4.1 Top viewed Products ..........................................................................38
      5.1.4.2 Top Sold Products ..............................................................................38
      5.1.4.3 Monthly Activity Report ......................................................................39
    5.1.5 Test Cases ..................................................................................................40
      5.1.5.1 Test Case 1 ..........................................................................................40
      5.1.5.2 Test Case 2 ..........................................................................................40
      5.1.5.3 Test Case 3 ..........................................................................................41
      5.1.5.4 Test Case 4 ..........................................................................................42
  6. Results ............................................................................................................44
LIST OF FIGURES

Figure 3.1 System Model ................................................................. 7
Figure 3.2 System Flowchart ............................................................ 9
Figure 3.3 PageRank Algorithm Illustration ..................................... 12
Figure 3.4 Admin Use Case Diagram .............................................. 16
Figure 3.5 Customer Use Case Diagram .......................................... 17
Figure 3.6 Login Use Case Diagram ............................................... 18
Figure 3.7 Order Use Case Diagram ................................................ 19
Figure 3.8 Class Diagram ............................................................... 20
Figure 3.9 Sequence Diagram ........................................................ 21
Figure 4.1 Product Information Query ............................................. 23
Figure 4.2 Hit Count Query ............................................................. 23
Figure 4.3 Hits in Database .............................................................. 24
Figure 4.4 Top Most Viewed Products Query .................................. 24
Figure 4.5 Display Top Products .................................................... 25
Figure 4.6 Top Most Viewed Products ............................................. 25
Figure 4.7 Page Rank Implementation……………………………………26

Figure 4.8 Page Rank Values…………………………………………………27

Figure 4.9 Product Search using Page Rank Implementation………………28

Figure 4.10 Product Search using Page Rank………………………………28

Figure 4.11 Google API libraries……………………………………………29

Figure 4.12 Draw Chart………………………………………………………29

Figure 5.1 Registration………………………………………………………32

Figure 5.2 Login………………………………………………………………32

Figure 5.3 Search products…………………………………………………33

Figure 5.4 Product Cart………………………………………………………34

Figure 5.5 Purchase Summary……………………………………………34

Figure 5.6 Pending Orders…………………………………………………35

Figure 5.7 Approved Orders………………………………………………36

Figure 5.8 Delivered Orders………………………………………………36

Figure 5.9 Add Product Information………………………………………37
Figure 5.10 Edit Product Information…………………………………………………………37
Figure 5.11 Top Products Viewed…………………………………………………………38
Figure 5.12 Top Products Sold……………………………………………………………39
Figure 5.13 Monthly Activity……………………………………………………………..39
Figure 5.14 Test Case 1_1…………………………………………………………………40
Figure 5.15 Test Case 1_2…………………………………………………………………40
Figure 5.16 Test Case 2_1…………………………………………………………………40
Figure 5.17 Test Case 2_2…………………………………………………………………41
Figure 5.18 Test Case 3_1…………………………………………………………………41
Figure 5.19 Test Case 3_2…………………………………………………………………42
Figure 5.20 Test Case 4_1…………………………………………………………………42
Figure 5.21 Test Case 4_2…………………………………………………………………43
Figure 5.22 Test Case 4_3…………………………………………………………………43
1. BACKGROUND AND RATIONALE

1.1 Need for Web Data Mining in E-Commerce

These days, internet has become a powerful tool for consumer oriented businesses. With less investment, these businesses are raking in great profits. In this world of intense competition among corporates, particularly in the consumer oriented market, it is of prime importance to catch a user’s attention. Hence, every statistical information of great use in order to increase a company’s market share in a specific market domain. Here comes the role of business intelligence.

A business requirement becomes the necessity for a business strategy and a business strategy in turn is converted into a business logic. This business flow is a common scenario in almost all the business powerhouses. But for building a business logic, we need data pertaining to the requirements of the business.

In this case, we focus on the online web consumers who rely on different e-commerce websites for their shopping needs. Hence, we need data pertaining to the online transactions or activities to study different patterns and design our business model in a way which can match perfectly with the consumer purchase trends. The data mining done on the web is called Web data mining [1].

1.2 What is Data Mining?

Data Mining is a process of collection of important or useful data from a huge amount of sparsely distributed data [2]. Previously, manual records were used as references for any information required for business process improvements. The records
included Excel sheets, documents etc. which were used to store business transactions for future reference.

But, as the years passed by it was becoming problematic to store huge amount of data and the manual records were eventually replaced by databases. These databases are used to store information in the form of digital data. Database querying differs from data mining in the sense that in querying we know what we want from the large pool of data. In data mining, it is all about finding patterns in the data so as to use such data for some business interests.

1.3 What is Web Data Mining?

Web data mining employs data mining techniques to extract information from the internet webpages. The difference being that web data mining deals with unstructured data compared to structured data dealt with in data mining. The data on the internet is very much unstructured and so web data mining has to be implemented in unique algorithms to deal with the task of collecting the data and filtering it so as to find patterns from the same.

There are different categories in Web data mining [3] as given below:

1. Web Usage Mining
2. Web Structure Mining
3. Web Content Mining
1.4 Web Usage Mining

Web usage mining enforces techniques in discovering the web browsing habits of users on a particular website or a web portal. It is an efficient technique to design the websites in a way so as to enable the users with easy access to web resources and thus garnering more customer attention in the process. The aim of Web usage mining is to analyze the browsing patterns of the users surfing a website.

In the current context, web usage mining has been implemented on the product pages so as to find the popularity of the products. This has been done so as to display the results on the users’ interface which can help them to find the best products available.

1.5 Web Structure Mining

Web structure mining is a technique which connects the webpages as a network of nodes and edges. Its purpose is to find relation between different webpages which are linked by some information or hyperlinks. Here, the webpages are represented as nodes and hyperlinks are represented as edges. The edges are directed from one webpage to another in a random fashion.

In the current context, the web structure mining has been implemented on the product pages so as to find relative rankings of the product pages. When a user types a keyword in the search bar, the results displayed will be relevant to the highly ranked products.
1.6 Web Content Mining

Web content mining is the process in which images, text and hyperlinks etc. are extracted in order to find a relevance to the search query. It is the second process in web data mining [12]. Also called text mining, it is usually done after related webpages are grouped together using web structure mining. Text mining helps the search engines with finding search results with high degree of accuracy. It also helps web designers align the web pages of their websites accordingly with respect to their degree of information relevance.
2. NARRATIVE

In this project, we focus only on web usage mining and web structure mining implementation in the e-commerce domain. The web usage mining focuses on the statistics of users visiting the website, whereas web structure mining focuses on finding relationship between different webpages of the website.

2.1 Problem Statement

With the growth of e-commerce industry, there is a constant tussle among different product based companies to tap the vast potential of internet in attracting consumers to view their respective products and purchase them. Strategies such as exhaustive manual analysis are implemented in studying the customer purchase pattern.

In this scenario, there is a need to implement business intelligence so as to achieve maximum customer attraction and subsequently increase the company’s market share. Also, there is a need to optimize the website user experience which can help increase the product sales on the e-commerce website.

2.2 Motivation

Many of the e-commerce websites follow different strategies to attract customers. Web data mining plays an important role in doing so. The web designers also make constant efforts to streamline the web user experience so as to give the users a feel-good factor about using their website.

An important factor which drives the online shopping business is trust. The trust
which users have over the e-commerce company that it is user-centric in its approach. For this to be achieved, the web designers employ different strategies to design the webpages which display products as per the users’ interests.

2.3 Project Objective

The objective of this project is to build a prototype of an e-shopping website which implements web data mining. Web usage analysis and web structure analysis has been employed to help users find the popular products. The admin of the website can also pull up different reports and statistics in order to analyze customer purchase patterns.

2.4 System Requirements

The following are the requirements of the system used to build the project and execute the same.

2.4.1 Hardware Requirements

<table>
<thead>
<tr>
<th>Component</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Core i3 2.1 GHz</td>
</tr>
<tr>
<td>RAM</td>
<td>4GB</td>
</tr>
<tr>
<td>Hard Disk</td>
<td>500GB HDD SATA</td>
</tr>
</tbody>
</table>

2.4.2 Software Requirements

<table>
<thead>
<tr>
<th>Component</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating System</td>
<td>Windows 8/10</td>
</tr>
<tr>
<td>Programming Languages</td>
<td>Java/JSP</td>
</tr>
<tr>
<td>IDE</td>
<td>Eclipse Luna</td>
</tr>
<tr>
<td>Web Server</td>
<td>Tomcat v7.0</td>
</tr>
<tr>
<td>Database</td>
<td>MySQL</td>
</tr>
</tbody>
</table>
3. PROPOSED SYSTEM DESIGN

A user logs into an e-commerce web site and views different products and purchases a product depending on his/her choice. The Figure 3.1 shows how a user generated data is collected and processed through web data mining and used as a feed for generating reports and used for improving the user interface. Generally, the web application logs contain the information regarding the user activity.

In this case, the information related to the product web page visited is stored. Since all the products in a web site have their own web page, the view is unique to a product. Similarly, when a user purchases a product, the transaction is counted towards the number of sales for the product.

![Figure 3.1 System Model](image-url)
For each user click, the hit count of a product gets incremented and stored in the database. This feature comes under web usage mining. The data collected from the user interaction is stored in a database. In web structure mining context, the product ranks are calculated based on the number of references each product webpage receives.

At the admin side, reports pertaining to top most selling product, top most viewed product etc. are generated based on the statistical data available. Besides this, we can also use the statistics for improving the user interface by providing customer recommendations etc. This business model helps the company in recognizing those products in which the customers are interested.

The assumptions are that the hit count tracking mechanism is enabled for all the product pages and the product ranks of all the products are assigned a constant value initially. Also, there is no cap on the number of users visiting the website. Currently, there are 44 products available in the website and hence there are 44 product webpages as well. But there is a provision for addition of any number of products.

3.1 System Flowchart

The Figure 3.2 shows the process flow in the system. The information such as hit count of a product, product purchase details is collected from the user activity. Also, the page rank of the product webpages is decided based on the number of references a product webpage gets from other webpages. The information is then stored in the database and used to generate reports and used to provide feedback to the user interface.
3.1.1 Hit Count Tracking

Each product webpage in the e-commerce website has a URL which contains Product ID pertaining to that specific product. Each time a user accesses a product webpage, the web server generates a query which registers a ‘hit’ against that particular product and the hit count for that product increases by 1 in the database.

From the stored data in the database, the most viewed products are selected and displayed on the users screen. Also, most sold products are selected according to the number of sales of that product. This information is displayed to the admin for the sake of statistical analysis.

At the administration side of the website, the admin can generate reports related to
monthly activity like most viewed products, most sold products and statuses of the different orders. This information helps them decide various strategies for different products as per the current market trends.

3.1.2 Page Rank Algorithm

The Page rank algorithm helps in determining the popularity of a web page from among all other web pages in a website or a web portal or the internet. Basically, it works on the principle that a page is ranked on quality compared against other web pages. The key idea is a page is ranked high if it is pointed by many highly ranked pages [6].

The Page rank algorithm was used to order the webpages based on the quality of the content of the webpages. Whenever a user searches using a keyword, the search engine responds with results ordered according to the Page rank. However, from realistic point of view this is not the only criteria used. We make an effort to use this algorithm in our application to sort the product in the search results which can help users assume the product quality. Consider the internet as a graph. Then the webpages are ‘nodes’ and the hyperlinks are the ‘edges’.

Given below is the Page Rank algorithm [6],

“Page Rank of page p,

\[ PR(p) = \frac{d}{n} + (1 - d) \sum_{(q,p) \in G} \frac{PR(q)}{\text{Outdegree}(q)} \]
Where,

d = probability that a user arrives at a webpage by typing the URL or from a bookmark
(‘d’ value is usually set between 0.1 and 0.2)

n= number of product web pages.

(1-d)= probability that a user arrives at the webpage by traversing a link.

PR (q) = Page rank of the page ‘q’

PR (p) = Page rank of the page ‘p’

G = graph containing the webpages ‘p’ and ‘q’

Out degree (q) = number of hyperlinks in the webpage.

The first term “(d/n)” indicates the probability of a user visiting a webpage by typing a URL instead of traversing a link. The second term “((1 - d) \sum_{(q,p)\in G} \frac{PR(q)}{\text{Out degree}(q)})” on the right hand side indicates the probability of a user arriving on a webpage by traversing a link.”

In the Figure 3.3, each node represents a webpage and each edge represents a hyperlink connecting one webpage to the other. In here, each product’s rank affects the rank of the other product to which it references to. For example, when a product 1 references product 2 then the product rank of the product 2 increments by a value equal to the “\left(\frac{\text{product rank of product 1}}{\text{No.of hyperlinks on the product 1 webpage}}\right)”. In this way, the product with most number of references from higher ranked products gets the highest product rank.
3.2 Framework

Web usage mining is carried out by tracking the user clicks and the web structure mining is implemented using the PageRank algorithm. The project implements the technologies such as JSP, servlets, JSTL and Java Beans. For the user Interface of the web application, HTML5, JQuery and CSS3 have been used. The reports have been generated using the Google charts API.

3.3 Steps to Project development

The following are the steps followed in the process of the e-commerce web application development.

- Development of the User Interface for the User and the admin.
Database creation pertaining to different users, products, categories etc.

- Implementing Web Usage Mining.
- Implementing Web Structure Mining.
- Implementing the functionality for the generation of different reports.

### 3.4 Environment

The project has been implemented using Java language and the environment used is Eclipse IDE.

#### 3.4.1 Eclipse IDE

Eclipse is an Integrated Development Environment (IDE). It includes a syntax highlighting editor, Incremental code compilation, project manager and a thread aware source-level debugger [7]. The application was introduced in November 2001. The developers can develop applications in this IDE using various languages like C, C++, Java, Perl, PHP, Python and Ruby. Eclipse is registered under public license and is an open source software.

#### 3.4.2 J2EE (Java 2 Platform Enterprise Edition)

It is a platform independent environment for developing web applications. It depends on Java Server Pages and servlets to generate HTML and other data format to the client. The components of J2EE are EJB (Enterprise JavaBean) and JDBC (Java Database Connectivity). JavaBean is a class which encapsulate more than one object into one single object. It is the location where the application’s logic is stored. JDBC is used to connect to the java databases in order to access the application data [8].
3.4.3 JSP (Java Server Pages)

JSP technology allows us to create static and dynamic content on the web. It’s a simple text based language which describes how to render web pages based on a request and constructing an appropriate response to it. The JSTL (Java Standard Tag Library) is a component of JSP technology which provides a collection of useful JSP Tags [9]. There are various types of tags available which perform various tasks as per the requirement.

3.4.4 MySQL

It is a relational database management system based on SQL (structured query language). It runs on several platforms like Linux, Windows and UNIX. It is an open-source software and is registered under public license. Although it can be used as a backend for a variety of applications, it is mainly used as a backend for web applications. It was originally created by a Swedish company called MySQL AB, but later it was acquired by Oracle in 2008 [11].

3.5 UML Diagrams

The following are the UML diagrams related to the e-commerce web application. The UML diagrams are used to specify the interactions between different objects diagrammatically.

The UML diagrams can be categorized into two types.

- Static
- Dynamic
The Static UML diagrams are used to focus on the relations between different objects and the different operations each object can perform. Whereas the Dynamic UML diagrams focus on the behavior of the system as whole in different contexts and the interactions between different objects during a process.

### 3.5.1 Use Case Diagrams

The Use Case Diagrams demonstrate the operations which can be performed by different objects in the diagram. Here, in this case, the actors are the Admin and the User. The following are the Use Case Diagrams for the e-commerce web application.

#### 3.5.1.1 Admin Use Case Diagram

The Figure 3.4 demonstrates the interaction between different actors and the functions handled between them. In this case, there is only one actor – Admin. The use case diagram for the actor ‘Admin’ demonstrates the functions performed by him as an administrator of an e-commerce web application. The admin logs into the website. The admin performs functions like adding/deleting products, modifying the product details. If any product is out of stock, the Admin can add more products to the inventory depending on purchase order from other wholesale dealer. When a User places an Order, the Admin has permissions to approve or reject the Order depending on the Customer credentials. The Admin can also add a User as an Administrator of the website. The Admin then logs out of the e-commerce website.
3.5.1.2 Customer Use Case Diagram

The actor in this case is ‘Customer’. The Figure 3.5 given below illustrates the functions performed by the customer when he/she logs into the e-commerce web application. The customer registers into the e-commerce website and logs into it. User searches for products. If the User intends to purchase a product, he/she can add that product to the Cart and proceed to Checkout or he/she can continue shopping. Once the desired products have been added to the Cart, the User can proceed to Checkout and add the delivery details on the Order and place the Order. The User then logs out of the e-commerce website.
3.5.1.3 Login Use Case Diagram

The Login use case diagram given in Figure 3.6 illustrates the interactions between the Admin and the Customer at the login interface. The User registers on the website and then logs into it. The User can maintain his/her account details and manage his/her password as well. The Admin can help the User in performing these tasks and can also add a User as an Admin of the website.
Figure 3.6 Login Use Case Diagram

3.5.1.4 Order Use Case Diagram

The Order use case diagram as shown in Figure 3.7 illustrates the interactions between the Admin and the User when the User is in the process of purchasing a product. The User logs into the website and purchases a product of his/her choice. The User adds the details of the Order such as Name, Delivery address etc. After the Order is placed, the Admin can approve/reject the Order and can notify the User of the same. The Admin can also change the status of the Order on verification of the payment credentials of the User.
3.5.2 Class Diagram

The Class diagram as shown in Figure 3.8 illustrates the interactions between various classes in the application. There are three sections in a Class. The first sections contains the Class name. The second section contains the class members or attributes. The third sections contains the Class methods. The arrows represent the interaction between two classes. The DB_Conn class provides the login to the Administrator and the User. The Product class is used to add/update products. The Cart class is used to add items to a transaction. The Order class is used to store the purchase orders with respect to a particular user.
3.5.3 Sequence Diagram

The sequence diagram as shown in Figure 3.9 is a visual combination of parallel vertical lines and horizontal arrows. The parallel lines represent the lifetime of the process and the horizontal lines represent the messages exchanged between the processes. Here, the Customer orders a new product. The order request is then processed by the Admin. The Admin can either approve or cancel the order request. Any such action will be notified to the Customer. In case the Admin approves the order request, then the Customer receives the bill.
Figure 3.9 Sequence Diagram
4. IMPLEMENTATION

This section deals with the implementation of different modules of the web application such as web usage mining, web structure mining and the generation of different reports pertaining to customer activity on the website and customer purchase patterns.

4.1 Web Usage Mining Implementation

Web usage mining is realized with the usage of ‘Hit Count’ as the criteria to determine the number of ‘Hits’ each product receives in the process of Customers surfing through the website.

The users visit the shopping website and surf through the product webpages while shopping for the products. In the process, they click on a product URL. Clicking on a product URL triggers an event, whereby the product ID of the product is extracted from the product URL and the corresponding ‘Hit Count’ of that product is incremented by ‘1’.

In the given code snippet as shown in Figure 4.1, the ‘request.getParameter’ method extracts the product ID from the requested webpage hyperlink of the product. Then, the DB_Conn() method implements a connection to the database. Then we create a statement from the connection and pass a query named ‘getProductQuery’ to it. The result set obtained contains the information related to all the products.
Clicking the product hyperlinks invokes a script which extracts the product ID from the hyper link and registers a ‘hit’ against that particular product ID in the database. In the given code snippet as shown in Figure 4.2, it is a database execution statement which increments the product_hits by ‘1’ in the product table against that particular product ID when the user clicks the any product hyperlink on a webpage.

```java
String id = request.getParameter("id");
if (request.getParameter("id") == null) {
    response.sendRedirect("viewProducts.jsp");
} else {
    DB_2Conn c = new DB_2Conn();
    Connection con = c.getConnection();
    Statement st = con.createStatement();
    String getProductQuery = "SELECT * FROM `products` p INNER JOIN `images` i USING ( `product-name` ) WHERE `product_id` =" +id+" GROUP BY `product-name` ";
    ResultSet rs = st.executeQuery(getProductQuery);
    rs.next();
    String product_id = rs.getString("product_id");
    int product_hits = rs.getInt("hits");
    String product_name = rs.getString("product-name");
    String sub_category_name = rs.getString("sub-category-name");
    String category_name = rs.getString("category-name");
    String company_name = rs.getString("company-name");
    String price = rs.getString("price");
    String summary = rs.getString("summary");
    String image_name = rs.getString("image-name");
    st.close();
```
The hits are registered in the database shown in Figure 4.3.

![Figure 4.3 Hits in Database](image)

After the hits have been registered, the same information is used to display the popular products on the user interface. The code snippet as shown in Figure 4.4 is used to retrieve the top most products information.

```java
String sqlTopMostProds = "SELECT * FROM `products` p
    INNER JOIN `images` i
    USING ( `product-name` )
    WHERE `product_qty` > 5
    GROUP BY `product-name`
    ORDER BY `hits` DESC
    LIMIT 0,15;"
```

![Figure 4.4 Top Most Viewed Products Query](image)

After the top most products information has been retrieved, the same is displayed onto the users’ screen using the following code snippet as shown in Figure 4.5.
The output of the above given code snippet is as shown in Figure 4.6. The products are arranged according to the hit count of each product.
4.2 Web Structure Mining Implementation

The web structure mining is all about finding a relation between different webpages and allotting them a rank as per the number of references each web page has. The page ranks are decided according to the page rank algorithm.

Initially, all the products are equally likely to be visited by the user by directly typing the URL in the browser. Hence, the products are given a default product rank value which is equal to the probability of the product page being visited by a user from among all other product web pages (i.e., \( \frac{1}{\text{number of total products}} \)). Then in the second iteration, the product ranks are determined as per the page rank algorithm [6].

The page rank algorithm code implementation is as shown in Figure 4.7.

```java
if(product_rank==0)
{
    String getProductdetails = "SELECT * FROM 'products' ";
    ResultSet rs1 = st.executeQuery(getProductdetails);
    ArrayList<Integer> links=new ArrayList<Integer>();
    ArrayList<Float> pr=new ArrayList<Float>();
    float prtotal=0;
    int rowcount = 0;
    if (rs1.last()) {
        rowcount = rs1.getRow();
        rs1.beforeFirst(); // not rs.first() because the rs.next() below will move on, missing the first element
    }
    if(rs1.isBeforeFirst())
    {
        while(rs1.next())
        {
            if(rs1.getInt("product_id") != Integer.parseInt(product_id))
                links.add(rs1.getInt("Nooflinks"));
            if(rs1.getFloat("productrank")==0)
            {
                pr.add((float)1/rowcount);
            }
            else
            {
                pr.add(rs1.getFloat("productrank"));
            }
        }
    }
    for(int i=0;i<(rowcount-1);i++)
    {
        prtotal+=(float)pr.get(i)/links.get(i);
    }
    prtotal=(float)(0.2+(0.8*prtotal));
    st.executeUpdate("UPDATE 'products' 
    SET 'productrank'='"+prtotal+"'
    WHERE 'products'.'product_id'='"+product_id+"'");
}
```

Figure 4.7 Page Rank Implementation
The variable ‘prtotal’ is declared as the rank of a product. Initially, it is set to ‘0’. Then for each iteration of the ‘for’ loop execution, the ‘prtotal’ value is incremented by 

\[ \text{pr.get(i)} \times \text{links.get(i)} \]

where ‘pr.get(i)’ is the product rank of other individual product webpages and ‘links.get(i)’ is the number of hyperlinks in that corresponding product webpage. After the execution of the page rank algorithm, the product ranks of the products are updated in the database as shown in Figure 4.8.

![Figure 4.8 Page Rank Values](image)

After the product ranks have been determined, the product ranks are added as a search criteria for the product search functionality. The code snippet as shown in Figure 4.9 illustrates the usage of ‘product rank’ as search criteria for product search. The string ‘sqlsearch’ is used to provide the database with the search query entered by the user on the user Interface. The minimum search length to be given by the user is ‘3’. Then the ‘sqlsearch’ string is concatenated with keywords such as ‘productrank’, ‘product-name’ and ‘product_qty’. This search query is then passed to the database through a database connection using the ‘jdbc’ connector library. The results given by the database are then loaded in a dataset and displayed to the user.
Figure 4.9 Product Search using Page Rank Implementation

The user inputs a product name or any key word while searching for a product. Then the results are sorted according to their respective product ranks. The search results are displayed as shown in Figure 4.10 for some random key input given.

Figure 4.10 Product Search using Page Rank
4.3 Reports Generation

The different reports are generated using the Google API charts. The following are the steps to convert a dataset into visual form or a graph or a chart.

- The first is loading the JSAPI library loader.
- Then load the visualization library.
- Preparing the Dataset
- Customize the Chart
- Draw the Chart

The code snippet as shown in Figure 4.11 represents the set of JavaScript tags used to call the Google API (‘jsapi’) and Visualization libraries (‘core.js’ and ‘core1.js’).

```html
<!--Loading the AJAX API-->
<script type="text/javascript" src="js/gclibrary/jsapi.js"></script>
<script type="text/javascript" src="js/gclibrary/core.js"></script>
<script type="text/javascript" src="js/gclibrary/core1.js"></script>
```

Figure 4.11 Google API libraries

Then to prepare the data, we need to create a data table and populate it with appropriate content as per the requirements. The final step being the visualization of the data set. This can be implemented using the code snippet as shown in Figure 4.12. The draw() method accepts two attributes i.e., ‘data’ and ‘options’. The ‘data’ attribute is used to input the data table content and the ‘options’ attribute is used to specify the title, X-axis parameter and the Y-axis parameter.

```javascript
var chart = new google.visualization.LineChart(document.getElementById('chart_div_lineCombo'));
chart.draw(data, options);
```

Figure 4.12 Draw Chart
Testing is a stage of the software process where thorough verification of the software application is carried out. This ensures that the software is bug free and is ready to be deployed in real time. The software has to go through various levels of testing before it can be certified bug free. Given below are few of the different levels of testing:

1. Unit Testing
2. Black box Testing
3. White box Testing
4. Integration Testing
5. System Testing

**Unit Testing**

It’s a module of Software testing where a particular unit of the code is tested in independent conditions. Here, an input is provided to the code unit and output is verified. If any failure occurs, the code segment is revisited and fixed.

**Black box Testing**

In this type of testing, the software tester just gives the input to the software application and verifies the output. Here, the tester is not concerned with what happens inside the system. No attempt is made to go through the code.
White Box Testing

In this kind of testing, the code undergoes thorough testing although not all at once. The code is split into individual modules and reviewed. Each code segment is provided an input and the output is observed. If any error occurs, then the error is fixed.

Integration Testing

In this testing methodology, two or more related modules of the program code are integrated and tested as a unit. The aim of this kind of testing is to verify if the modules work together well or not.

System Testing

After all the different types of testing has been done, the system is tested as a whole to verify if all its functionalities work as expected. This includes giving input and observing output for boundary conditions.

5.1 User Interface

The user interface of the shopping website has two views. The first one is the ‘admin view’ and the second one is ‘user view’.

5.1.1 User View

The user view provides the users with an interface where they can register themselves, purchase products and manage their account.
5.1.1.1 User Registration

First time a user visits the shopping website, the user has to register himself. The user interface for the same is as shown in Figure 5.1.

![Figure 5.1 Registration](image1.jpg)

5.1.1.2 User Login

After successful registration, the user is allowed to login to the website. The user interface for the same is as shown in Figure 5.2.

![Figure 5.2 Login](image2.jpg)
5.1.1.3 Searching Products

The logged in user can search the products using the Search bar or going directly to the products page. Here, the minimum length for the search keyword to be entered is set to ‘3’. The user interface for the same is as shown in Figure 5.3.

![Figure 5.3 Search products](image)

5.1.1.4 Purchasing products

User can purchase the products and add to the cart and later checkout the same at the end when he/she is done with shopping. The user interface for the same is as shown in Figure 5.4.
5.1.1.5 Purchase Summary

The user is displayed the following order summary when he/she is done with the payment of the transaction and submission of the shipment details. The user interface for the same is as shown in Figure 5.5.
5.1.2 Admin View

The admin view provides the administrators of the websites to manage the website’s different functionalities such as add/remove/update the products, view/approve/cancel orders, generating reports etc.

5.1.2.1 Pending Orders

After the user purchases a product, the purchase order is placed on the admin dashboard to be approved. The admin has a choice to either approve or reject the order. The user interface for the same is as shown in Figure 5.6.

![Pending Orders](image)

Figure 5.6 Pending Orders

5.1.2.2 Approved Orders

The pending orders are approved by the administrator if the Customer credentials satisfy the Admin expectations. Such orders move into the category of ‘Approved Orders’. The user interface for the same is as shown in Figure 5.7.
5.1.2.3 Delivered Orders

After the goods have been delivered to the customer, the order status is changed to ‘delivered’. Such orders move into the category of ‘Delivered Orders’. The user interface for the same is as shown in Figure 5.8.

Figure 5.7 Approved Orders

Figure 5.8 Delivered Orders
5.1.3 Add/Edit Product Information

The admin has the privileges to add a product information in the database using the admin view. The user interface for the same is as shown in Figure 5.9.

![Add Product Information](image1)

Figure 5.9 Add Product Information

The admin also has the privileges to edit a product information. The user interface for the same is as shown in Figure 5.10.

![Edit Product Information](image2)

Figure 5.10 Edit Product Information
5.1.4 Reports

The reports are generated by the Admin using the data present in the database in order to get information regarding popular products and customer purchase patterns.

5.1.4.1 Top viewed products

The top most viewed products report displays the products that have received the maximum hits. The X-axis displays the name of the products and the Y-axis displays the number of hits the corresponding product received. The same is as shown in Figure 5.11.

![Top 10 products viewed currently](image1)

Figure 5.11 Top Products Viewed

5.1.4.2 Top Products Sold

This report indicates the top most products sold. The X-axis indicates the name of the products and the Y-axis displays the number of units sold for each product. The same is as shown in Figure 5.12.
5.1.4.3 Monthly Activity Report

The monthly report displays the expenses and sales occurred during a particular month. It also displays if an order is in pending, approved or delivered state. The same is as shown in Figure 5.13.
5.1.5 Test Cases

5.1.5.1 Test Case 1

In this test case, we verify if the hit count of the product is getting updated when the user clicks the product hyperlink.

Before clicking the product link, the hit count of the product is 5. The same is as shown in Figure 5.14.

![Figure 5.14 Test Case1_1](image1)

After clicking the product link, the hit count is incremented by ‘1’. The same is as shown in Figure 5.15.

![Figure 5.15 Test Case 1_2](image2)

5.1.5.2 Test Case 2

In this test case, we verify if the products are displayed as per the hit count of the products.

The Figure 5.16 shows the order of the products according to the hit count ‘hits’.

![Figure 5.16 Test Case 2_1](image3)
The Figure 5.17 shows the arrangement of the Top Most Viewed products as per the hit count of each product.

<table>
<thead>
<tr>
<th>Top Most Viewed Products Currently</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Assassin's Creed III" /></td>
</tr>
<tr>
<td>Assassins Creed III</td>
</tr>
<tr>
<td>By Ubisoft Games</td>
</tr>
<tr>
<td>$600.00</td>
</tr>
</tbody>
</table>

Figure 5.17 Test Case 2_2

**5.1.5.3 Test Case 3**

In this test, verify if the search results are getting displayed according to the Page Rank values of the products.

The Figure 5.18 shows the order of the products according to the product rank.

<table>
<thead>
<tr>
<th>product_id</th>
<th>product-name</th>
<th>sub-category</th>
<th>category</th>
<th>company-rev</th>
<th>price</th>
<th>sur</th>
<th>tags</th>
<th>proc</th>
<th>last U. hits</th>
<th>product rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>Angry Birds Oil Pastels</td>
<td>Colors</td>
<td>Stars...</td>
<td>Rovio</td>
<td>60.00</td>
<td>&lt;..</td>
<td>Angry Birds Oil Pastels</td>
<td>19</td>
<td>20...</td>
<td>3</td>
</tr>
<tr>
<td>36</td>
<td>Cars Of Pastels</td>
<td>Colors</td>
<td>Stars...</td>
<td>Disney</td>
<td>100.00</td>
<td>&lt;..</td>
<td>Cars Of Pastels</td>
<td>18</td>
<td>20...</td>
<td>18</td>
</tr>
<tr>
<td>37</td>
<td>Sakura Pastels</td>
<td>Colors</td>
<td>Stars...</td>
<td>Disney</td>
<td>110.00</td>
<td>&lt;..</td>
<td>Sakura Oil Pastels</td>
<td>20</td>
<td>20...</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 5.18 Test Case 3_1

At the users’ end, the products should be displayed in the order of the Page rank values when the search key matches the product name. The same is as shown in the Figure 5.19.
5.1.5.4 Test Case 4

In this test case, we verify if the product rank of a product webpage is being generated when it is loaded onto the webserver for the first time.

To prove this, log into the e-commerce website as an Admin using the credentials of an administrator. Then when a new product is added then the product rank of the new product is ‘0’ as shown in Figure 5.20.

Then the new product page is loaded on the web server for the first time as shown in the Figure 5.21.
When the new product’s webpage is loaded for the first time on the web server, then its product rank is calculated and updated as shown in Figure 5.22.

Figure 5.22 Product Rank Update
6. RESULTS

The web application enables the customers to have a clear view of the popular products in the website through the usage of web usage mining. It also enables the customers to search for quality products through the usage of web structure mining. At the administration end, the admin can have a clear information of the products which are visited and purchased by the most customers. The admin can use the information in recommending the popular products to the users. The web application also displays the monthly activity report to the admin which helps in the system audit. It will also enable the companies to design customized strategies for marketing their products so as to maximize their product share in the market by attracting the customers. This will also spare the admin from performing the tedious work of analyzing the product market trends using the product sales data manually.
7. CONCLUSION AND FUTURE WORK

In this project, the web usage mining has been implemented to take advantage of the user clicks on the product webpage hyperlinks and organize the product popularity to show the same on the users’ screen. In advancement to this, the usage mining can be extended to study the user behavior like time spent on a specific webpage or the securing the traffic load on a website. Similarly, the web structure mining can be coupled with text mining in order to give highly desirable results in the search query. The text mining along with structure mining can also be used to track users’ movements on the website which can help the analysts in finding relation between different product purchase patterns. In the current context, the Page Rank algorithm has been implemented in the intra-site perspective in order to calculate product ranks. In order to realize this in the real time, it needs to be implemented in inter-site perspective. Hence, the web application needs to be hosted onto the internet wherein the product webpages in the current web application are registered onto the global web server and their corresponding product ranks are calculated based on the references each product receives from webpages all over the internet.
BIBLIOGRAPHY AND REFERENCES


10. http://patft.uspto.gov/netacgi/nph-

APPENDIX

The following code snippet implements the Page Rank algorithm in order to calculate the product ranks.

```java
if(product_rank==0)
    
String getProductdetails = "SELECT * FROM `products` ";
ResultSet rsl = st.executeQuery(getProductdetails);
ArrayList<Integer> links=new ArrayList<Integer>();
ArrayList<Float> pr=new ArrayList<Float>();
float prtotal=0;
int rowcount = 0;
if (rsl.last())
    
rowcount = rsl.getRow();
    rsl.beforeFirst(); // not rs.first() because the rs.next() below will move on, missing the first element

if(rsl.isBeforeFirst())
    
while(rsl.next())
    
if[rsl.getInt("product_id") != Integer.parseInt(product_id)]
    
    links.add(rsl.getInt("Wooflinks"));
    
if[rsl.getFloat("productrank")==0]
    
pr.add((float)1/rowcount) ;
    
else
    
pr.add(rsl.getFloat("productrank"));
    
}

for(int i=0;i<(rowcount-1);i++)
    
prtotal+=((float)pr.get(i)/links.get(i);
}

prtotal=(float)(0.2+(0.8*prtotal));
st.execute("UPDATE `products` 
+" SET `productrank`="+prtotal+""
+" WHERE `products`.`product_id` ="+product_id="");
```