

Use of Data Mining Techniques to Determine Customer Loyalty by Performing Market Basket Analysis on E-commerce Database

Ms. Pratik Kiran Sayanak¹, Prof. Keerti Naregal²

Student, Dept of Computer Science & Engg, KLE DR M S Sheshgiri College of Engg & Tech, Belagavi, India¹

Professor, Dept of Computer Science & Engg, KLE DR M S Sheshgiri College of Engg & Tech, Belagavi, India²

Abstract: Lately with the quick development of e-commerce, a lot of information is gathered through operational transactions, data mining techniques are used to find and comprehend unknown customer purchase patterns. Earlier, data mining has been utilized to discover which products are connected as far as having high deals and furthermore discover which customers merit credit services. There has not been much work done in the utilization of data mining to guarantee customer loyalty in the e-commerce industry and furthermore have methodologies of expanding retail organizations to use e-commerce as a beneficial method of doing business. Customer-Product relationship has consistently been the most essential relationship for development of business. The better this relationship better is income generation and better is the development of organization. Results obtained in the piece of work are based on the concept of classification and association rule mining to provide better understanding of customer and product relationship and thus encourage organizations to develop the use e-commerce for conducting profitable business.

Keywords: Data Mining, E-Commerce, Classification using decision trees, Association rule mining.

I. INTRODUCTION

Data mining is a process that involves requirements gathering, information accumulation, database storage, business intelligence, and deployment. The whole data mining procedure can be quickly summed up in straightforward layman term as: Large volume of data – value-based, operational and other non conventional types of data is altogether gathered and put away in expansive databases normally called as database warehouses. This information should be pre-processed, cleaned and changed before it can be mined. Mining includes disclosure of novel, insignificant and valuable patterns. The patterns so found offer ascent to particular area learning which users can follow up on. Most of the commonly used data mining techniques are classification, clustering, association analysis and regression analysis.

The E-Commerce (e-Commerce) industry is one of the most recent to utilize data mining technology. E-commerce is the utilization of data and communication technology through the Internet platform to share business data, keep business connections, and lead business transactions. In e-commerce, diverse data mining methods can be utilized for several reasons. For instance, in sales advancement the promoting staff might need to discover which items their customers will probably purchase together. This data will enable them to put these things in a business package keeping in mind the end goal to expand income. The utilization of Web log data gives licenses to comprehend customers conduct. This information contains data about consumer and may indicate potential patterns in their purchase, and recognize potential customers of e-commerce organizations. This learning is valuable to: change marketing strategies; recognize division of customers, enhance customer retention, anticipate customers use and market patterns, give customized services to customers, examine shopping basket, forecast better sale deals; overhaul the website to give a superior services or potentially settle on better business choices.

Thus in this digital world with the rapid development of e-commerce lot of information is gathered through operational transactions, data mining techniques are considered to be more valuable to find and comprehend obscure customer patterns.

Customers are the integral part of any business venture. The increase in the cost of drawing in new customers on the Internet and the relative trouble in holding them make customer loyalty a fundamental resource for many e-commerce merchants. In the traditional (non-internet) marketplace, customer loyalty is fundamentally the result of exceptional service quality and the trust that is established for a particular business organization. Establishing online customer loyalty and maintaining existing customers is the necessity for many e-commerce vendors.

Customer-Product relation has consistently been the most vital relationship for development of business. The better this relationship better is income generation and better is the development of organization. This present work focuses on the use of data mining techniques such as classification and association rule mining to understand the customer-product

relationship, determine customer loyalty from the considered e-commerce database and hence help in the overall development of e-commerce business.

II. RELATED WORK

This section surveys the work done in utilization of data mining techniques in e-commerce and to derive hidden patterns from the collected information.

In 2012 [2] the authors, determined the client behavior model by using semi-directed learning in Customer Relationship Management. Presently organizations are directed from item oriented to customer driven associations. For which it's by and large trusted that holding existing customers is more productive than pulling in new ones. Customer information is quickly developing, subsequently, organizations are confronted by the issue to examine esteemed client for affiliation and holding. Data mining advancements remove concealed data and learning from expansive datasets put away in databases. The authors have introduced semi-directed learning system to enhance CRM forms and their productivity. The proposed semi regulated strategy is useful to foresee the class of potential or/and obscure consumers. The proposed method is feed-forward neural framework prepared by using back propagation calculation. Authors have analysed the proposed system on Bank and Insurance datasets.

In 2013 [3], the authors determined a ski resort's effect on deals for which they introduced promotional and publicizing techniques utilizing decision tree. This ski resort provides facilities for winter games, skiing and snowboarding. They utilize diverse specialized strategies like advertising, publicizing and deals advancement to interact with its key market section. They are also utilizing another advanced correspondence channels for M-business, due to the development of digital gadgets and e-commerce. The authors introduced the use decision tree framework to profile the administration and technology. These innovations and administrations consolidate smaller scale blogging services, resort sites and online coupon administrations. Ski resorts divides client in two noteworthy classifications i.e. millennial or Generation X (not exactly or equivalent to age 35) and non-millennial (more prominent than age 35) and use advance promoting strategies for them. This ski resort enormously utilizes Social Media Networks, Websites and Micro blogging advances for advancing and promoting its administrations. As indicated by obtained outcomes the providing of advanced services have extraordinary effect on resort deals. The effect is empowering and constant in nature.

In 2014 [4], the authors has proposed a system based on data mining to distinguish customers purchasing patterns. The success of any business relies upon the capacity to comprehend their clients' needs. To know the why purchasers always go to their favoured stores is assuming a vital part in accomplishing upper hand and holding their marketplace. In today's world, Business Analytics is useful to investigate the tremendous measure of information keeping in mind the end goal to pick up clients experiences and enhance client connections. The authors introduced the data mining framework which can be utilized to find designs in clients' visits to a store and distinguish their shopping pattern. The utilization of this proposed system is implemented in purpose of offer information of eight delegate stores for a Greek retailer. It can be utilized to help a few choices in the retail area and enhance the connections amongst retailers and purchasers.

In 2015 [5], the authors have introduced an effective CRM based data mining system for deriving the patterns to analyse the behaviour of customers. CRM information mining system is useful for overseeing relationship amongst associations and clients. The model enhances the basic decision making process for holding esteemed clients. Information mining methods like classification are the most widely used as a part of customer relationship management (CRM). The authors proposed comparison between following classification methods; Naive Bayes and Neural Networks and results demonstrate that Neural Networks execution is superior to Naïve Bayes. The authors implemented the proposed system on the bank promoting dataset that is standard UCI datasets.

Thus, from the above literature survey, we can observe that different data mining techniques are used to predict the customer purchase behaviour in e-commerce. However in this project the emphasis is on the use of data mining techniques such as association analysis and classification based on decision trees to understand the customer-product relationship and thus determine customer loyalty from the considered e-commerce database.

III. PROBLEM DEFINITION

Diverse issues confronted by e-commerce ventures are as following:

- Lack of comprehension of customer characteristics.
- Lack of comprehension of product traits.
- Lack of comprehension of customer-product relationship.

Reasons for the failure of e-commerce ventures:

- The principle issue with e-commerce based organizations is that occasionally it goes for pitching incorrect and unintended product to the customers. This makes customers unsatisfied and unhappy. So it must be avoided.



- If one needs to acquire higher income then he/she needs to ensure that the business site is obliging destitute customers at time of their need with the products that they require.
- The e-commerce has its existence only on web. Accordingly one should make ideal utilization of this stage for promoting.

Thus to overcome the above stated facts this proposed work uses the data mining concepts such as classification and association analysis to better understand the customer-product relation and hence determine customer loyalty from e-commerce database.

IV. SYSTEM DESIGN

System design is the development of decomposing a generous complicated structure into substructures. It contains all the procedure of “Data Mining” process. The structure is depicted in Figure 1.

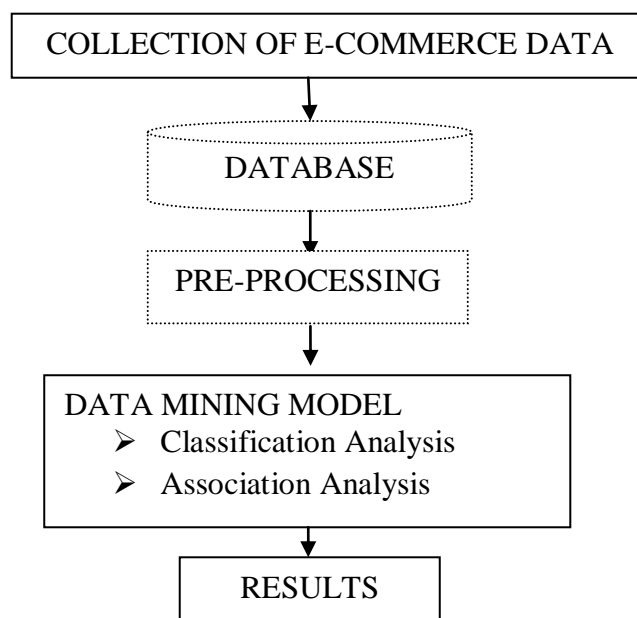


Figure 1: System Design.

The complete structure is composed of 3 modules; they are collecting e-commerce data, pre-processing the collected data and data mining module.

1. Collecting E-commerce data: This step comprises of collecting all accessible data from the e-commerce database. To do this, an arrangement of components that can help e-commerce organizations were distinguished and gathered from the diverse information sources accessible. The information was accumulated from e-commerce data storehouse, which comprised of MySQL databases. All the data was incorporated into a dataset, utilizing Microsoft Excel.

2. Pre-processing of the collected e-commerce data: In this step the dataset is groomed to apply the data mining strategies. To do this, conventional information pre-processing strategies, for example, data cleaning, change of variables and division of data. Different methods, for example, the choice of attributes and the re-adjusting of information were additionally applied so as to take care of the issues of high dimensionality and imbalanced information that are commonly introduced in these datasets.

The pre-processing of e-commerce data consists of the following steps: Data cleaning: This includes information pre-processing to expel information irregularity and commotion.

- Data incorporation: this includes consolidating information from numerous information sources. This incorporated the distinguishing of the important tables in the database which were the Product Table and Order Details Table.
- Data choice: This includes recovering those records and credits pertinent to our analysis.
- Transformation of data: This includes the change of the information into proper forms for mining by performing outline operations. Reduction of data was additionally performed in this stage to get a little portrayal of the original information.



3. Data mining model: The pre-processed e-commerce dataset is then loaded into this model and the data mining concepts such as classification using decision trees and association analysis using apriori algorithm are used to understand the customer product relations and hence determine customer loyalty from e-commerce database. Finally all the results of the analysis are displayed using a web application.

V. IMPLEMENTATION

A. Data Collection and Pre-processing module.

“Data collection” is essential step used to represent a procedure of assembling and gathering data. It is a precise collection of data being used for appropriate reason from different origin, which has been precisely recognized, registered and classified. For any “decision making” technique, data is the necessary input. The scope of collecting the data is to retrieve information, to preserve data, to form decisions regarding crucial problems and transmission of data to the residuals.

The considered e-commerce dataset is collected from “UCI Machine Learning Repository”. This e-commerce dataset has approximately about 8000 entries consisting of product, customer and order details using Microsoft Excel format.

The data is then pre-processed to remove missing and inconsistent values that are not required for our analysis. The pre-processed dataset is stored in the server called as “WAMP Server”. It stands for “Windows, Apache, MySQL and PHP”. It is regularly utilized for testing and developing the web interface, however may likewise be utilized to serve live sites. The data is stored in “MySQL” database and it consists of certain attributes.

B. Data Mining Techniques

In an organization based on e-commerce it is important to communicate and interact with future and present customers since there won't be any physical and genuine communication with the customers. The communication with customers will be in type of requesting that the customers visit the site page and pick the items they need to buy. This process will include many stages. Every one of these stages depends on utilizing business information and data to transform it into learning base. This change of data to information requires data mining methods such as clustering, sequential pattern analysis, regression analysis, classification and association rule analysis.

In this project the emphasis is on the use of the following two data mining techniques so as to understand the customer-product relation and thus determine customer loyalty from the considered e-commerce dataset, they are as follows:

- Classification based on usage of decision tree analysis.
- Association analysis using “Apriori Algorithm”.

1. Classification using Decision Tree Analysis:

In applications of data mining one regularly accepts that the data is now in some kind of advanced frame (something like a major spreadsheet). Here, one might need to anticipate the estimation of a specific property (a specific segment in the spreadsheet). At the point when this trait, also referred to as the class property, incorporates a limited number of discrete components, we facing a classification issue. In this sort of issue, we construct a scientific model from the accessible information. This model gets the data of a novel occasion whose class is obscure and creates an estimation of the classification which it has a place with. Our assignment is to play out this estimation as precisely as possible. Thus classification is said to be a data mining technique that assigns categories to a collection of data in order to aide in more accurate predictions and analysis.

The aim is to generate a set of classification rules that will answer a question, make a decision, or predict behaviour. To begin with, a set of training data is developed that contains a certain set of attributes as well as the likely outcome. The work of the classification algorithm is to discover how that set of attributes reaches its conclusion.

One of the most commonly utilized classification methods is based on decision tree analysis. Decision tree is the learning based on class of labelled training tuples. A decision tree is a flowchart-like tree structure consisting of the following:

- Internal nodes(non leaf node) denotes a test on an attribute.
- Branches represent outcomes of tests.
- Leaf nodes (terminal nodes) hold class labels.
- Root node is the topmost node.

Use of decision tree for classification:

- Verify or test the attributes of a tuple against the decision tree.
- A path is followed from the root to a leaf hub which holds the prediction for that tuple.

Basic Decision tree algorithm:

- Tree is constructed in a top-down recursive divide-and-conquer manner.



- At start, all the training examples are at the root.
- Attributes are categorical.
- Examples are partitioned recursively based on selected attributes.
- Test attributes are selected on the basis of a heuristic or statistical measure (e.g., information gain).

Conditions for stopping partitioning:

- All samples for a given node belong to the same class.
- There are no remaining attributes for further partitioning.
- There are no samples left.

Example: Consider the below given tuple with the following attributes

Product_id	Click_count	Customer	Date_difference	Class?
1	0	Yes	<250 to 550	NO

Here for a particular product with ID 1 we have:

- ✓ Test on Click_count: 0
- ✓ Test of Customer: yes
- ✓ Date Different: Less than 250 and equal to 550.
- ✓ Class NO: the customer is unlikely to buy a product.

Below is the decision tree constructed for the above data indicating whether the customer will buy a particular product:

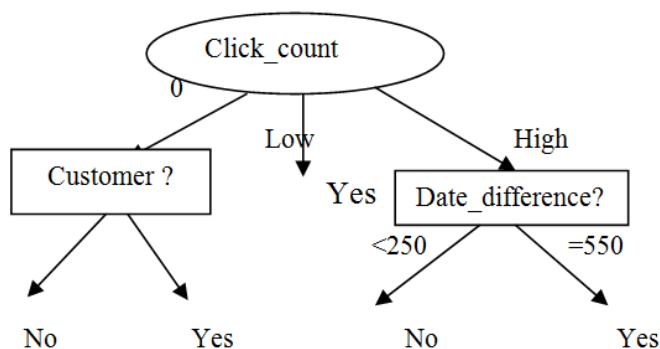


Figure 2: Decision tree

Association Analysis using Apriori Algorithm:

Another real issue examined in data mining is the with respect to association analysis. In this specific situation, our information is acclimated from the transactions carried out, e.g., the bill that incorporates a rundown of items that you purchased in a market, in case of this project the e-commerce website. The aspects of data are unique: things purchased don't really rehash in two bills, however as a rule people have a tendency to act likewise in their purchasing patterns. Association analysis endeavours to find those patterns. In this specific circumstance, one might be keen on knowing which designs are more regular. A well known illustration is the connection amongst bread and butter in supermarket bills. Data like this gives helpful information at the time a supermarket is outlined: on the off chance if you realize that individuals will purchase bread and butter you can assemble them, or place them in inverse corners, expanding the likelihood that the clients will see different items they may be keen on.

Thus association rule mining is usually required for finding frequent patterns, association, correlation or casual structures among a set of item or objects in the database consisting of transactions, relational database and other information repositories. Some of the application of association analysis are cross-marketing, catalog design, sale campaign analysis, web log (click stream) analysis, DNA sequence analysis and shopping cart data analysis also called as "market basket data" analysis. The set of products a customer purchases during one visit to supermarket, with respect to this project, the visit or browsing of the e-commerce website is known as market-basket data. Itemset is generally used to indicate the set of products purchased by the customer.

An association rule is generally represented in the form $Q \Rightarrow R$, where $Q = \{q_1, q_2, q_3, \dots, q_n\}$ and $R = \{r_1, r_2, r_3, \dots, r_n\}$ are set of items, with q_i and r_i being distinct items for all i and j . For an association rule to be of interest, it must satisfy a minimum support and confidence.

The basic concepts used in association analysis are as follows:



- **Frequent pattern:** it refers to a pattern (a set of items, subsequences, substructures, etc.) that often occurs frequently in a given data set.
- Consider an itemset $Q = \{q_1, \dots, q_k\}$
 - Find all the rules $Q \cap R$ with minimum support and confidence.
 - Support, s , $P(Q \cap R) / \text{Total no. of transactions}$, it means probability that a transaction contains both Q and R .
 - Confidence, c , $P(Q \cap R) / P(R)$, measures the conditional probability that how often item in R appears in a transaction having the item Q .

Example: The information indicating that customers who purchase scanner also tend to buy printer at the same time is represented by association rule as follows:

- Scanner \Rightarrow Printer [support = 3%, confidence = 70%]
- A support of % indicates that 3% of all the transactions under analysis show that scanner and printer equipments are purchased together.
- A confidence of 70% indicates that 70% of the customers who purchased a scanner also bought the printer equipment.

Association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Usually, association rule mining can be seen as a two-stage process:

1. Find all itemsets that are frequent: By definition, each of these itemsets will appear as frequent as predetermined minimum support count i.e. min sup.
2. Generate solid association rules from the itemsets that occur frequently: By definition, these standards must fulfill minimum support and minimum confidence.

If an itemset is frequent, at that point each of its subsets is also frequent; a noteworthy test in mining frequent itemsets from a vast informational index is that it produces countless itemsets fulfilling the minimum support (min sup) limit, particularly when min sup is set low.

Apriori Algorithm:

Apriori Algorithm is a standout amongst the most famous and powerful algorithms in association rule mining. It is a algorithm for mining itemsets that occur frequently and learning association rule over value-based databases. The Apriori Algorithm was first presented by Agarwal and Srikant (1994) which produces frequent itemsets in light of a threshold called 'Minimum Support'.

The **Apriori property** of frequent patterns states that "Any subset of a frequent itemset must be frequent".

Example: If { computer ,scanner ,printer } is frequent, then so is {scanner, printer} i.e., every transaction having { computer ,scanner ,printer } also contains { scanner ,printer }. If a set cannot pass the test, all of its supersets will fail the same test as well.

The following are the steps for Apriori Algorithm:

Method:

Let C_k - Candidate itemset of size k
 L_k - Frequent itemset of size k

1. Initially, scan DB once to get frequent 1-itemset.
2. Generate length $(k+1)$ candidate itemsets from length k frequent itemsets, C_k .
3. Prune C_k using the Apriori property.
4. Test the candidates against DB to generate the list of frequent $(k+1)$ itemsets satisfying the min sup, L_k .
5. Terminate when no frequent itemset, L_k , or candidate set, C_k can be generated.

Association rules used for determining customer loyalty from the considered e-commerce dataset where the support threshold is set to 15% and the confidence threshold is set to 70% are given below. Note that for the purpose of comparison, the support confidence values are attached in brackets:

1. If click_count is zero and [s=28%, c=66%] then the date_difference is minimum.
2. If date_difference is a mid value between say 250-550 and [s=38%, c=92%] then cart_number is zero.
3. If click_count is zero and [s=39%, c=94%] then the cart_number is zero.
4. If click_count is a low value say 19 and [s=24%, c=83%] then the cart_number is zero.
5. If click_count is a average value between say 19 to 49 and [s=12%, c=78%] then cart_number is zero
6. If click_count is a high value between say 49 to 500 and [s=11%, c=82%] then cart_number is zero.
7. If unit_price is ≤ 25 and order_count=10 [S=16%, c=100%] then the customer might buy this product and might also buy other products along with the current product.



VI. RESULTS AND ANALYSIS

The data collection is represented in the preceding stage and through examining these collected data; the subsequent results can be procured.

E-commerce dataset:

The below Figure.3 shows details in the considered e-commerce dataset, which is obtained from UCI machine learning repository. Here we can see the product details such as its name, category and subcategory. It also contains customer and order details which are not shown in figure. This e-commerce dataset is in Microsoft Excel format and has approximately about 8000 transaction entries.

Product Category	Product Sub-Category	Product Name
Office Supplies	Storage & Organization	Eldon Base for stackable storage shell, platinum
Office Supplies	Appliances	1.7 Cubic Foot Compact "Cube" Office Refrigerators
Office Supplies	Binders and Binder Accessories	Cardinal Slant-D® Ring Binder, Heavy Gauge Vinyl
Technology	Telephones and Communication	R380
Office Supplies	Appliances	Holmes HEPA Air Purifier
Furniture	Office Furnishings	G.E. Longer-Life Indoor Recessed Floodlight Bulbs
Office Supplies	Binders and Binder Accessories	Angle-D Binders with Locking Rings, Label Holders
Office Supplies	Storage & Organization	SAFECO Mobile Desk Side File, Wire Frame
Office Supplies	Storage & Organization	SAFECO Commercial Wire Shelving, Black
Office Supplies	Paper	Xerox 198
Office Supplies	Paper	Xerox 1980
Office Supplies	Rubber Bands	Advantus Map Permanent Flags and Round Head Tacks
Office Supplies	Appliances	Holmes HEPA Air Purifier
Technology	Computer Peripherals	DS/HD IBM Formatted Diskettes, 200/Pack - Staples
Office Supplies	Binders and Binder Accessories	Wilson Jones 1" Hanging DUBL Lock® Ring Binders
Furniture	Office Furnishings	Ultra Commercial Grade Dual Valve Door Closer
Office Supplies	Envelopes	#10-4 1/8" x 9 1/2" Premium Diagonal Seam Envelopes
Furniture	Bookcases	Hon 4-Shelf Metal Bookcases
Furniture	Tables	Lesro Sheffield Collection Coffee Table, End Table, Center Table, Corner
Technology	Telephones and Communication	g520
Technology	Telephones and Communication	LX 788
Office Supplies	Labels	Avery 52
Office Supplies	Rubber Bands	Plymouth Boxed Rubber Bands by Plymouth
Office Supplies	Binders and Binder Accessories	GBC Pre-Punched Binding Paper, Plastic, White, 8-1/2" x 11"
Technology	Computer Peripherals	Maxell 35" DS/HD IBM-Formatted Diskettes, 10/Pack
Office Supplies	Pens & Art Supplies	Newell 335
Office Supplies	Pens & Art Supplies	SANFORD Liquid Accent Tank-Style Highlighters
Technology	Copiers and Fax	Canon FC340 Copier
Technology	Copiers and Fax	Canon FC340 Copier

Figure 3: E-commerce dataset.

Data pre-processing module:

The below Figure.4 shows the "JAVA" admin/user interface that is used by the concerned person for processing the considered e-commerce dataset. The user interface is composed of few java combo box and buttons as shown in the above figure. Another feature provided is the analysis of the product attributes which is carried out after the pre-processing of e-commerce dataset.

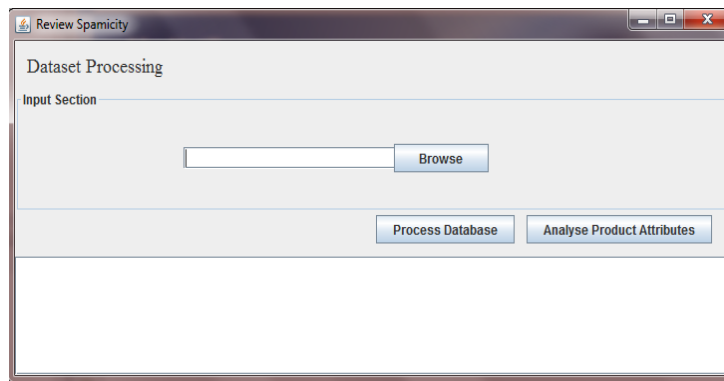


Figure 4: Admin/User Interface.



Figure 5: Output window and the alert message box.

The above Figure.5 shows the output window and alert message box indicating that the pre-processing of the e-commerce dataset is completed.

Product attributes analysis:

Below Figure.6 shows the output window shows the analysis of product attributes, the support and confidence threshold calculation for a particular product. The alert message box indicates the completion of analysis.

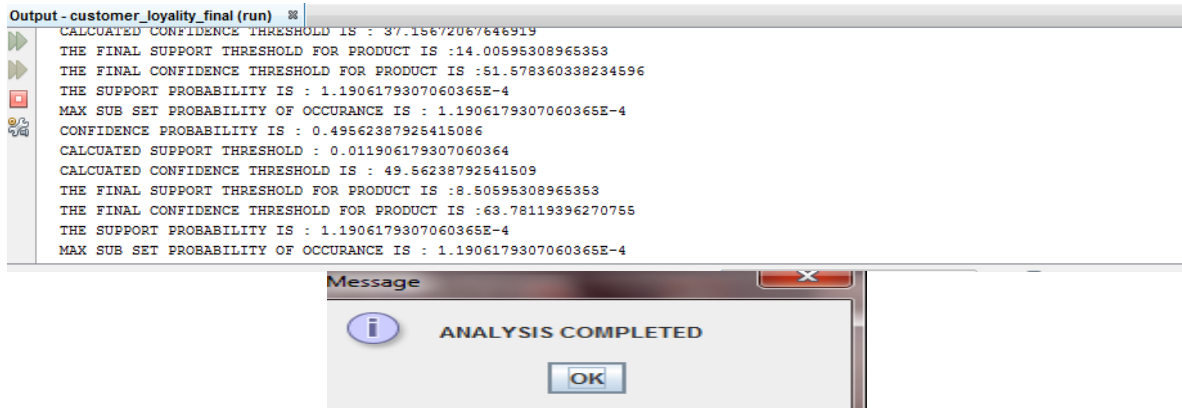


Figure 6: The output window for the analysis of product attributes and alert message box for completion of analysis.

Web Part:

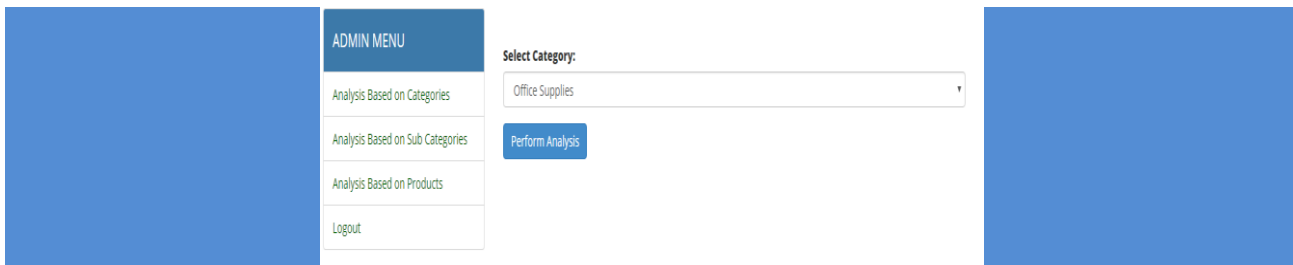


Figure 7: Admin menu page.

The above Figure.7 shows the admin menu page, where in the admin can carry out analysis based product categories and subcategories. Here in this figure on selection the analysis based on categories option we select the particular product category from the drop down and then perform analysis.

Category Analysis:

The below Figure.8 shows the graph for analysis of products based on categories. Here the y-axis indicates the products available in the furniture category and the x-axis indicates the number of products frequently bought in that particular categories.

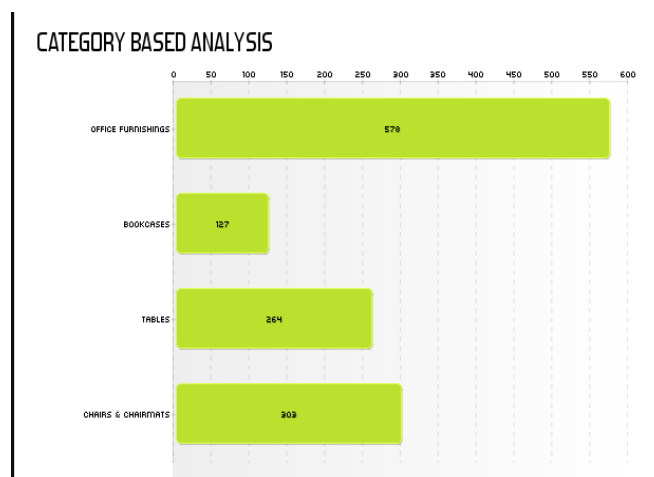


Figure 8: Bar graph representation for analysis of product categories.



Subcategory analysis:

Figure.9 shows the graph for analysis of products based on subcategories. Here the y-axis indicates the purchase count of the products frequently bought and the x-axis indicates the product ID of the particular product in that subcategory.

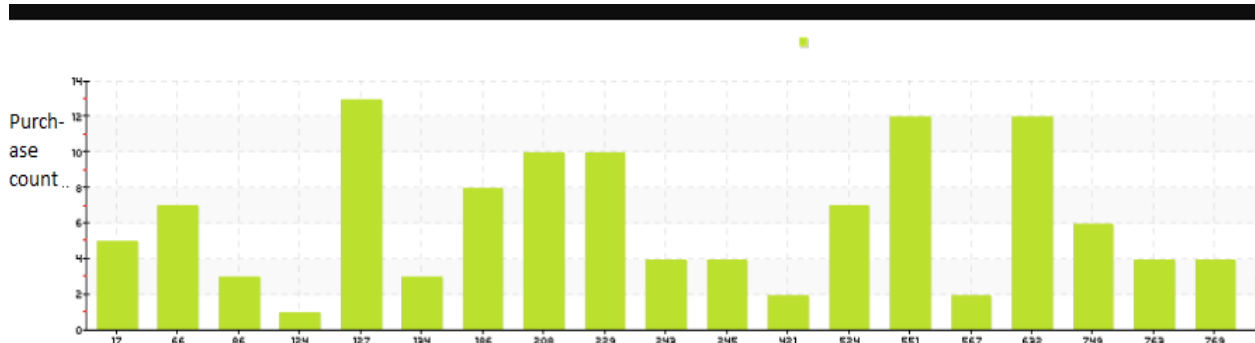


Figure 9: Bar graph representation for analysis of product subcategories.

Analysis based on products and order:

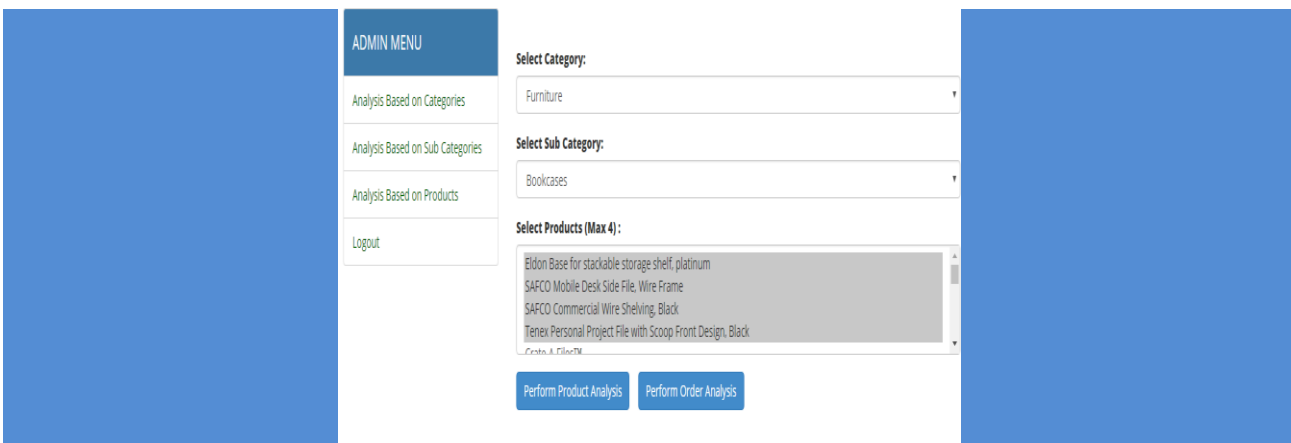


Figure 10: Analysis based on products.

The above Figure.10 shows that the admin can perform analysis based on products and order. Here the admin can select a particular category, subcategory and any number of products available in that subcategory from the considered e-commerce dataset.

Output for product and order based analysis:

The below Figure.11 shows the product attributes determined and upon clicking the prediction button the output of order analysis is obtained. Here the row with cream color indicates that the sale of that particular product is low, the row with blue color indicates average sale of products, and the row with green color shows high sale of products which is not seen in the below snapshot.

Product Name	Category Name	Sub Category Name	Click Counts	Date Difference	Sold Count	Options
Eldon Base for stackable storage shelf, platinum	Office Supplies	Storage Organization	& 11 (LOW)	362 (250 To 550)	3 (1 To 9)	Prediction
SAFCO Mobile Desk Side File, Wire Frame	Office Supplies	Storage Organization	& 28 (Average)	715 (>550)	6 (1 To 9)	Prediction
SAFCO Commercial Wire Shelving, Black	Office Supplies	Storage Organization	& 29 (Average)	728 (>550)	5 (1 To 9)	Prediction
Tenex Personal Project File with Scoop Front Design, Black	Office Supplies	Storage Organization	& 63 (HIGH)	751 (>550)	15 (10 To 29)	Prediction
Carina Double Wide Media Storage Towers in Natural & Black	Office Supplies	Storage Organization	& 53 (HIGH)	742 (>550)	12 (10 To 29)	Prediction
Tennsco Snap-Together Open Shelving Units, Starter Sets and Add-On Units	Office Supplies	Storage Organization	& 23 (Average)	725 (>550)	9 (1 To 9)	Prediction
Fellowes Recycled Storage Drawers	Office Supplies	Storage Organization	& 41 (Average)	758 (>550)	8 (1 To 9)	Prediction
Eldon Base for stackable storage shelf, platinum	Office Supplies	Storage Organization	& 11 (LOW)	362 (250 To 550)	3 (1 To 9)	Prediction
SAFCO Mobile Desk Side File, Wire Frame	Office Supplies	Storage	& 28 (Average)	715 (>550)	6 (1 To 9)	Prediction

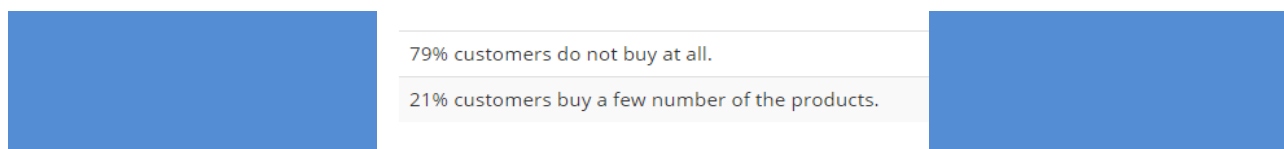


Figure 11: Output for order analysis.

VII. CONCLUSION

Data mining assumes an exceptionally vital part in the e-commerce industry and furthermore in the advancement of applications based on e-commerce which helps us to derive useful and significant information from the transaction operation carried out by the e-commerce organization. The aim of this presented work was to influence the e-commerce organization's customers to purchase propelled product bundles with a specific end goal to help their deals by utilizing the mined outcomes from the unique product bundles to build up an association outline that will obviously promote the distinction in sales returns. Having model customers will likewise help pull in potential customers who are not utilizing the organization's e-commerce site. The organization may now have instruments to not only exclusively influence potential customers to utilize the organization's products but in addition to direct them in choosing the appropriate modules for leading their business procedures and their customers can utilize them to augment purchased product satisfaction. Results acquired from one of this work uncover that there is an unavoidable requirement for a change in the sales of products provided by the e-commerce site. Either a change in the outline of the site to make it more appealing to customers or, on the other hand a presentation of advancements, for example, "get one get one free" which won't just expand the rate at which products are sold but additionally broadens the customer base for the e-commerce organization.

REFERENCES

- [1] DATA MINING TUTORIAL Easy Learning by tutorialspoint.com
- [2] S. Emtiyaz, and M. Keyvanpour, "Customers behaviour modeling by semi-supervised learning in customer relationship management," arXiv preprint arXiv:1201.1670, 2012.
- [3] P. Duchessi, and E. J. Lauria, "Decision tree models for profiling ski resorts' promotional and advertising strategies and the impact on sales," *Expert Systems with Applications*, 40(15), 5822-5829, 2013.
- [4] Griva, C. Bardaki, S. Panagiotis, and D. Papakiriakopoulos, "A Data Mining Based Framework to Identify Shopping Missions," 2014.
- [5] T. F. Bahari, and M. S. Elayidom, "An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour," *Procedia Computer Science*, 46, 725-731, 2015.