Product Pitching using Data Mining

Malla Jyotsna[1], Preetham Lekkala[2], Sai Surya Varshith[3], Ragunthar T[*]

[1]School of Computer Science and Engineering,Vellore Institute of Technology
Vellore, Tamilnadu, India
[2]School of Computer Science and Engineering,Vellore Institute of Technology
Vellore, Tamilnadu, India
[3]School of Computer Science and Engineering,Vellore Institute of Technology
Vellore, Tamilnadu, India
[*]Assistant Professor Senior, School of Computer Science and Engineering,Vellore Institute of Technology
Vellore, Tamilnadu, India

[*]Corresponding author: ragunthar.t@vit.ac.in

Abstract

Ecommerce websites have gained immense popularity in recent years. These days consumers prefer to buy products and goods online through e-commerce websites instead of the physical retail shops owing to the standardization of prices across all websites. It results in increased competition among various e-commerce businesses. These businesses compete for drawing the attention of the consumers to their websites and the products associated with it. Ecommerce businesses have to keep in mind to price their products reasonably and also to pitch the features well to the people before launching a new product in the e-commerce websites. The customers' inclination is more towards how valuable the product is. So the price to features ratio needs to be kept high to assure the customers that the product is highly valuable.

This paper focuses on helping e-commerce businesses to pitch their products more effectively by understanding consumer behavior from the feedback gathered from them. This data is mined using data mining techniques that result in a better understanding of the features of the product that needs to be pitched by the businesses to increase the conversion rates.

Keywords: e-commerce, pitch, data mining, conversion rates

1. Introduction

A product that is to launch shortly needs to be pitched extraordinarily to persuade the customers to purchase it. The e-commerce businesses should aim at making an interactive website that pitches the features well and leaves a long-lasting mark on the customers. Strong pitching creates an impact on customers who were not interested in buying the product earlier and persuades them to own it. It increases its popularity among the people. E-commerce websites reduce the search costs relative to visiting physical stores and comparison sites reduce search costs still further. Businesses should understand how pricing impacts consumer buying behavior. Ecommerce websites lead to better standardization of prices and this increases the competition among them. Profits can increase with better pitching that results in increased conversion costs. The necessity of innovative and appealing motion design has reached its peak where numerous competitors are competing to gain a second or minute of the customers' attention.

The conversion rate is the successful completion of the payment process of a product. The most important metric for improving the conversion rate is sales. Numerous challenges are observed in the context of increasing sales. These include inefficient handling logistics, warehouse management, and poor distribution, payment security issues in e-commerce, inefficient, lower sales conversion On-Site.Various website analytic tools are used to analyze the patterns from various e-commerce websites. Heatmaps are used to predict the exact user navigation patterns on a website. Analytics is done to search the high ranking and popular websites, the traffic sources, and also to find out the pages on the website that have high bounce frequency. User Tests are done to find congestion on nodes and to check for any suspicious activity while surfing the websites. Survey forms are used to gain feedback about the user experience while surfing the website, rating, and any recommendations and suggestions that would make the user experience on the websites better. Based on the results obtained through these tools solutions to the challenges have been identified. These include optimizing the checkout process, increasing visibility of the system status of the website, providing a locus of control to the customer on the website, and increasing user trust (Saleem et al, 2019; S. Hernández et al., 2017).

Recently semantic search and data tracking is being done by companies to target ads and products to the user; it becomes all the more necessary for digital marketers to make the user interface and design to stand out among its competitors. The issue of confidentiality also arises; so it is crucial to give assurance to the customers about the proper usage of the data collected from their activity on the website.

Online reviews of the products play a huge role in determining its sales. Customers go through reviews to understand the risk factor associated with the purchase of the product. This paper examines information from the website of a particular brand of phone. The data obtained from the online reviews are processed and then some suggestions are given to improve the product performance and thereby increase the conversion rate. This study resulted in showing that a strong correlation existed between the change in the degree of feature satisfaction and phone improvement. It will help the organizations to determine the direction and contents of the phone improvements based on the information gathered from the online reviews of the preceding models. It will also help the organizations to understand consumer behavior and to make innovations in the product according to the feedbacks and recommendations obtained from the previous models(Zhang et al., 2018).

Data mining technologies and their applications are applied in e-commerce activities. Data mining is done using methods like correlational analysis that aims to bring out hidden relationships in data. Classification analysis is used to classify a part of data into the category it fits in most. Cluster analysis is done to divide the database into a set of clusters according to definite rules. These methods can be implemented in e-commerce websites data for mining customer buying behavior, analyzing customer access to the site, gender positioning customer network, path analysis. Information gathered from the data mining processes in the form of rules, concepts, patterns, plots, maps can help the e-commerce websites to make comparisons between historical and current data, discover hidden relationships and patterns between data, to predict the possible future behavior of customers(Li Yan et al., 2015; Sudhamathy et al., 2012).

E-Marketing must help the consumers in their purchase on e-commerce websites. So the owners of e-shops must know the consumer decision journey. The analysis uses the digital footprint that a customer leaves on an e-commerce website. This is used to predict average customer behavior, requirements, desires, suggestions to improve the web presence using Data Mining. A survey of 86 e-shops has shown the importance of SMEs to become active members of web mining solutions. It can be analyzed that price analysis is not relevant to them since the majority of it is established by the product providers. Almost all the companies which were surveyed had deployed web analytics over their website but an exhaustive data analysis was not followed by the majority of them. The study shows that in real sales, the process of selling in e-commerce shops should be as per the customers' requirements before and after the sale. It requires applying data mining in the phase to satisfy the necessity of enriching the data and add value according to the better information about the client(Jon et al.; 2015).

We focus on a product that would be launched shortly and gain feedback about the price and the features that the customers are more likely to pay for using an interactive website. The website provides users with a feedback module that gathers data and information about the expectations and requirements of the people from the product. Feedbacks help the business grow; it clears the minds of developers and helps in creating products that would help to increase conversion costs. The primary objective of this study is to help increase conversion costs. This persuasion is surmountable considering into account the fact that the business goes one step closer where they ask the customers about how much they want the products to be priced. It improves the bond between companies and customers. It helps to keep customers in the circle and they would already be into the product which also serves our purpose of getting them to feel attracted to the product such that they are ready to purchase it.

The data gathered from the feedback responses are used to gather insights about the features that require to be pitched to increase sales of the product. We used Market Basket Analysis as a data mining tool used to understand the behavior of customers. It uncovers the items that are frequently bought together and leads to increased sales and better marketing.

Data mining helps e-commerce businesses to help understand customer behavior. Log data of websites visitor traffic on the e-commerce website is analyzed to predict the visitors' behavior. Using a data analysis approach backtracks can be found by the sequences produced during the stage of preprocessing. Backtrack pages are undesirable for the users and products which are backtracked would be automatically pushed down across a range of products. The navigation of the user can also be used to predict the next page to be viewed based on the previous page, previously ordered sequence, previous unordered sequence. Based on the analysis of sequences association rule mining is performed to find the correlation between items which are webpages and the aim is to analyze how the visitor visit to the website is related by considering the weblog data according to (Adnan et al.; 2011).

A data mining system at Blue Martini Software is developed for e-commerce businesses. This data mining system includes data collection, data warehouse creation, data transformation, and associated business intelligence systems that include reporting, visualizations, and data mining. The data mining lifecycle stages have been considered in the natural order:

requirements gathering, data collection, data warehouse construction, business intelligence, and deployment. To develop the data mining system; each part of the lifecycle should be carried out. The lessons learned and the challenges observed during the process requires further investigation. This data mining system helps e-commerce businesses to make their analysis(Ron Kahavi et al.; 2004). This paper focuses on a new approach to retrieve information from a given e-commerce website; a collection of data from the sites' structure, retrieving the semantic information in predefined locations, and analysis of user's access logs. It enables the development of accurate models for predicting customers' future behavior. It is achieved with the help of a web mining process analyzing the sites' structure, content and, usage in a pipeline that results in a web graph of the website and categorization of each page and website's user profiles. This study results in the development of an all-in-one process to collect and structure the data obtained from an e-commerce websites' structure and users. The process involves crossing the data collected from various sources to find the explicit relations. Finally, the informational model proposed contains the collected and structured information obtained from an e-commerce website(Dias et al.; 2005).

This design of a Semantic and Neural based Ecommerce page ranking algorithm called SNEC page ranking algorithm that can be used as a website ranking tool. It helps customers to find the best-ranked websites for their requirements and the businesses can also compare themselves with their competitors and hence find ways to improve their strengths. SNEC algorithm leads to the implementation of the Website Priority Determination Tool. This tool shows a comparison of the top 6 websites about a particular product or topic. This ranking shown by the Tool can help the designers of websites to optimize the structure of the website according to the rank. This tool can be improved using various tabs on the interface such as page loading speed comparison, ease of navigation comparison, and security comparison. Further cloud computing frameworks such as Hadoop Distributed File Systems can be introduced into the framework to help mine easily the Big Data produced by e-commerce websites(Vinoth Kumar S et al.; 2009).

2. Materials and Methods

2.1 Market Basket Analysis

Market Basket Analysis commonly known as Association Rule Mining is a method used for data analysis in retail and marketing purposes. This technique is used in many fields like bioinformatics, education, marketing, nuclear science. It is used for analyzing the purchasing behavior of the customers(Kaur et al.; 2016). It helps to discover hidden relationships and patterns in the data and gives an understanding of the frequent preferences and choices of the customers. It takes input as categorical data in the form of transaction records and gives the output as the association rules based on the choices of the customers. This helps e-commerce businesses to achieve effective sales and marketing. Association rule mining is a two-stage process:

1. Frequent item sets generation.
2. Rules generation from frequent item sets(Kulkarni et al.; 2012).

Association rules(X-> Y) are implication also known as if-else statements where X is the antecedent and Y is the consequent. The strength and association between the antecedent and consequent of the rules is determined with the support, confidence, and lift values. Support indicates frequent item sets. Frequent itemset occurs when the itemset has a support value greater than the minimum support threshold. Support values help to generate the association rules from the frequent item sets. Association rules(X-> Y) are implication also known as if-else statements where X is the antecedent and Y is the consequent. The strength and association between the antecedent and consequent of the rules are determined with the support, confidence, and lift values. Support indicates frequent item sets. Frequent itemset occurs when the itemset has a support value more than the minimum support threshold. Support values help to generate the association rules from the frequent item sets. Given that the antecedent has already occurred the probability of the consequent to occur is known as confidence. It only takes in the importance of the occurrence of the antecedent, thus it is not significant.

The correlation between the antecedent and the consequent is called lift. A lift value greater than one indicates a positive correlation i.e strong association. Lift value less than 1 indicates a negative correlation i.e weak association. The Lift value as zero indicates that the antecedent and consequent are independent(Zhang et al.; 2011; Haorianto et al.; 2005).

2.2 Apriori Algorithm

Apriori algorithm is a data mining method. It is used for determining the frequent item sets and then generating the association rules. It operates on a record consisting of transactions. Apriori algorithm is simple and easy to understand and can be used with large datasets. The demerits of the apriori algorithm are that it needs to find a large number of candidate rules many times that would be expensive computationally. Since the algorithm needs to scan the entire database calculating the support value is expensive(Harun et al.; 2017).

2.3 FP-Growth Algorithm:

FP- Growth overcomes the shortcomings of the apriori algorithm. Frequent patterns are generated without the candidate generation.

The database is structured in the form of a tree called a frequent pattern tree. This maintains the association between the item sets. The tree is used for determining the most frequent pattern. The lower nodes of the FP tree represent an item of an itemset and the root node is considered null. FP-Growth algorithm follows the divide-and-conquer approach. Fragmentation of the database is based on the most frequent item called the "pattern fragment". By analyzing the item sets of fragmented patterns finding the frequent item sets is comparatively less expensive(Santhosh et al.; 2009).

3. Results and Discussion

Our dataset contains 100 records collected over a week.V1, V2, V3, V4, V5, V6, V7, V8, V9 correspond to the different features offered to the customer. Item-Frequency Plots determine the top N items with the highest item frequency. Figure 1 and Figure 2 represents the Bar-plot and Tree-map show the top 5 features which are selected frequently by the customers. It shows that the top five features are V9, V1, V10, V7, V4. These features correspond to curved edges, face-unlock, smartwatch, dual-front Camera, Shatterproof display.

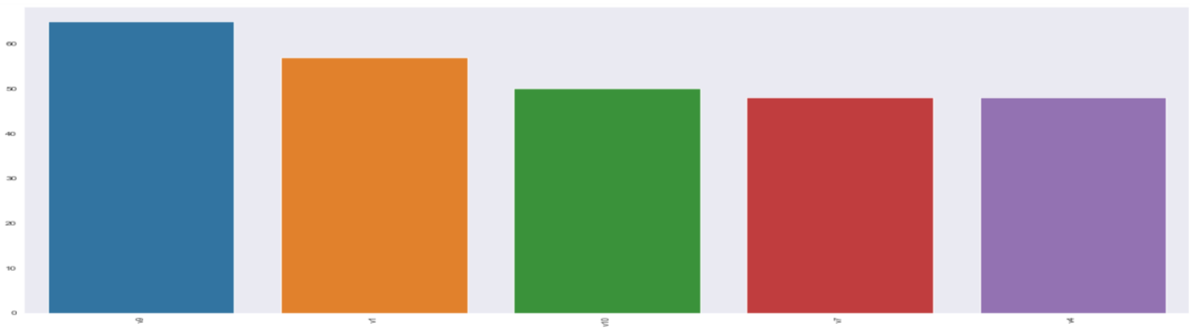

Figure 1: Item Frequency Bar plot

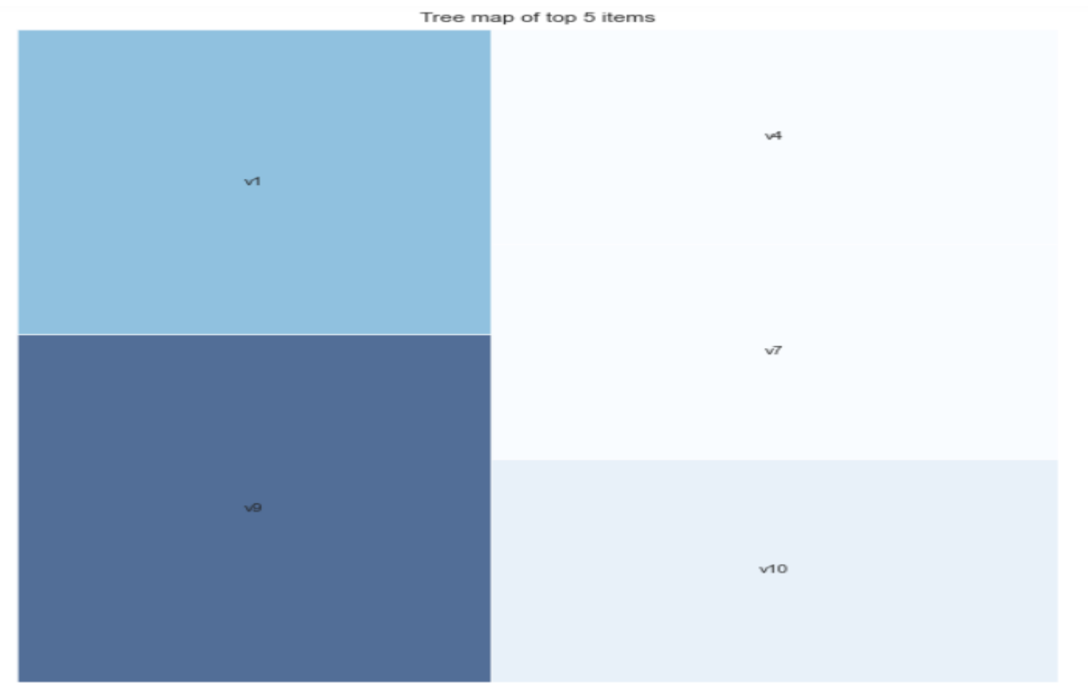

Figure 2: Item Frequency Tree Map

Given the dataset containing 100 records collected over a week and the criteria that the features which are chosen at least 5 times a day are taken as frequent item sets; the min_sup =0.35. Table 3 shows the 11 frequent item sets of varying.

Table 1: Frequent Item sets

No	Support	Item sets
0	0.57	(V1)
1	0.50	(V10)
2	0.43	(V3)
3	0.48	(V4)
4	0.44	(V5)
5	0.35	(V6)
6	0.48	(V7)
7	0.40	(V8)
8	0.65	(V9)

9	0.38	(V1,V9)
10	0.37	(V10,V9)

Now we generate the association rules.

Taking confidence like 20%; the minimum threshold value is 0.57. Figure 3 shows that 4 association rules are generated.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(v1)	(v9)	0.57	0.65	0.38	0.666667	1.025641	0.0095	1.050000
1	(v9)	(v1)	0.65	0.57	0.38	0.584615	1.025641	0.0095	1.035185
2	(v10)	(v9)	0.50	0.65	0.37	0.740000	1.138462	0.0450	1.346154
3	(v9)	(v10)	0.65	0.50	0.37	0.569231	1.138462	0.0450	1.160714

Figure 3: Association rules based on lift measure

Based on the rules generated it can be inferred that 66% of the customers' who choose face-unlock also chose curved edges. 58% of the customers' who chose curved edges also chose face-unlock. 74% of customers' who chose smart-watch also chose curved edges. 56% of customers' who chose curved edges also choose smartwatch. There is also more chance that smartwatch and curved edges are chosen together than curved edges and face-unlock. 2 frequent item sets are generated for minimum support threshold is 0.35. Table 2 shows that 2 frequent item sets are generated. It indicates that face-unlock, curved-edges and smart-watch, curved edges are the most frequent item sets.

Table 2: Frequent item sets of length 2

No	Support	Itemset	Length
9	0.38	(V1, V9)	2

10	0.37	(V10, V9)	2

Dataset is also trained using the FP-Growth algorithm.

Considering the minimum support value to be 0.35; 11 frequent item sets of varying lengths are generated. The frequent item sets are the same as those generated using the Apriori algorithm. Similarly. 4 association rules are generated using the minimum lift threshold value as 0.57. The analysis obtained from the association rules is the same as that obtained from the Apriori algorithm. The FP-Growth algorithm is comparatively 5 times faster than the Apriori algorithm since it does not require creating candidate sets explicitly. Figure 4 shows a line graph plotted to compare the run times of the two algorithms against the minimum support threshold. It shows that the run time of FP-Growth is comparatively much faster than the Apriori algorithm.

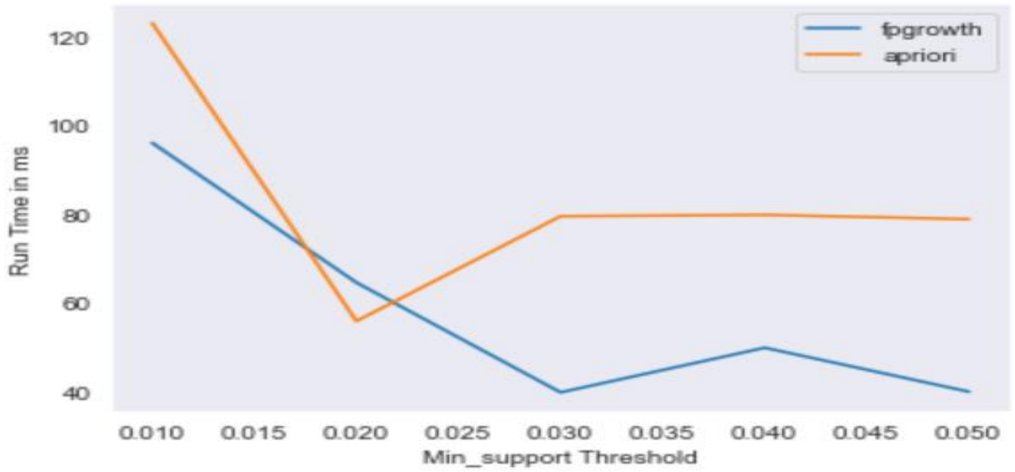

Figure 4: Line graph to compare run-times

4. Conclusion

It is observed from this analysis that data mining methods can help to gather patterns and understand the behavior of the customers in the e-commerce websites. Using Market Basket

Analysis the features of the product that are chosen frequently by the customers are analyzed. FP-Growth algorithm is used for faster and more efficient implementation of the Market Basket Analysis Technique. Our analysis would help the e-commerce businesses and the product developers to understand the expectations of the product which will launch shortly. The patterns discovered from mining the data helps the businesses to properly pitch the key features of the product to attract the customers. This will help to increase the conversion rates of the product and would enable the e-commerce businesses to make better strategic decisions. Thus it brings out the importance of incorporating data mining analysis in e-commerce businesses to help them compete with other businesses and increase their conversion rates.

Acknowledgment

This research would like to thank the support of Vellore Institute of Technology, Vellore, Tamilnadu.

References

Adnan, Muhaimenul & Nagi, Mohamad & Kianmehr, Keivan & Tahboub, Radwan & Ridley, Mick & Rokne, Jon., 2011."Promoting where, when and what? An analysis of web logs by integrating data mining and social network techniques to guide ecommerce business promotions", Social Network Analysis and Mining, Volume 1, pp 173-185.

Dias, João & Ferreira, Hugo., 2009. "Automating the Extraction of Static Content and Dynamic Behaviour from e-Commerce Websites", Procedia Computer Science, Volume 109, pp 297-304.

Jon Kepa Gerrikagoitia, Iñigo Castander, Fidel Rebón, Aurkene Alzua-Sorzabal., 2015. "New trends of Intelligent E-Marketing based on Web Mining for e-shops", Procedia - Social and Behavioral Sciences, Volume 175, pp 75-83.

Harun, Nur & Makhtar, Mokhairi & Aziz, Azwa & Zakaria, Zahrahtul & Syed Abdullah, Fadzli & Jusoh, Julaily., 2017. “The Application of Apriori Algorithm in Predicting Flood Areas”, International Journal on Advanced Science, Engineering and Information Technology, Volume 7(3).

S. Hernández, P. Álvarez, J. Fabra and J. Ezpeleta., 2017. “Analysis of Users’ Behavior in Structured e-Commerce Websites” , in *IEEE Access*, vol. 5, pp. 11941-11958.

Kaur, Manpreet & Kang, Shivani., 2016. “Market Basket Analysis: Identify the changing trends of market data using association rule mining”, Procedia Computer Science, Volume 85, pp 78-85.

Ron Kohavi and Llew Mason and Rajesh Parekh and Zijian Zheng.,2004. “Lessons and challenges from mining retail e-commerce data”, Machine Learning, Volume 57, pp 83-113.

Kulkarni, Raja., 2012. “Association Rule – Extracting Knowledge Using Market Basket Analysis”, Research Journal of Recent Sciences, Volume 1, pp 19-27.

Dr. Santhosh Kumar, Dr. K.V Rukmani., 2009. “Implementation of web usage mining using APRIORI and FP growth algorithms”, Int. J. of Advanced Networking and Applications, Volume 1,Issue 6, pp 400-404.

Vinoth Kumar S., Shaheen H., Sreenivasulu T., 2008 “An Improved Scheme for Organizing E-Commerce-Based Websites Using Semantic Web Mining, Intelligent Computing and Application”, Advances in Intelligent Systems and Computing, vol 1172.

Saleem, Hussain & Uddin, M Khawaja & Habib-ur-Rehman, Syed & Saleem, Samina & Aslam, Ali., 2019 .“Strategic Data Driven Approach to Improve Conversion Rates and Sales Performance of E-Commerce Websites”, International Journal of Scientific and Engineering Research, Volume 10, pp 588-593.

Sudhamathy, G. & Venkateswaran, Jothi., 2012. “Fuzzy Temporal Clustering Approach for E-Commerce Websites”, International Journal of Engineering and Technology, Volume 4, pp 119-132.

Haorianto Cokrowijoyo Tjioe & David Taniar., 2005. “Mining Association Rules in Data Warehouses”, International Journal of Data Warehousing and Mining (IJDWM), Volume 1(3), pp 28-62.

Li Yan., 2015. “The application of data mining technology in e-commerce”, 5th International Conference on Education, Management, Information and Medicine .

Zhang, H., Rao, H. & Feng, J., 2018. “Product innovation based on online review data mining: a case study of Huawei phones”, *Electron Commer Res* 18, 3–22 .

Zhang, Shichao & Wu, Xindong., 2011. “Fundamentals of association rules in data mining and knowledge discovery”, Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery, Volume 1, pp 97-116.