

PREDICTION OF CUSTOMER PURCHASE INTENTION USING LINEAR SUPPORT VECTOR MACHINE IN DIGITAL MARKETING

G. Saranya^{1,*}, N. Gopinath², G.Geetha³,K.Meenakshi⁴ and M. Nithya⁵

¹Department of Information Technology, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India 603203.

²Department of computer Science and Engineering, Sri Sairam Engineering College, Tamil Nadu.

³Department of Information Technology, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India 603203.

⁴Department of Information Technology, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India 603203.

⁵Department of computer Science and Engineering, Sri Sairam Engineering College, Tamil Nadu.

*Corresponding author e-mail: saranyag3@srmist.edu.in

Abstract. Digital marketing is taken into account the well-liked method comparing to traditional marketing. It can be used by both researchers and academicians for social media marketing and to predict the customers purchase intention. The Proposed work revolves around some valuable information and processes in accordance to the behavior of customer during the online purchase. Business owners, scientists, researchers all post their ads, details on the Web so that they can be linked to owners quickly and easily by web scrap searching on searchable product websites to gain a lot of data from websites. Details on websites are stored in an unstructured manner. To avoid this issue, Web Scraping helps to gather unstructured content and turn it into a structured type that can be used for further study. Hence, customer price and rating of product evaluation and prediction has become an important research area. The analysis is done by Support Vector Machine (SVM- Linear) to gather several information and provide variation analysis. The major goal remains to investigate and analyze the extracted dataset using ML oriented algorithms with best accuracy possible. The analysis has a proper path to sentimental analysis of parameters in accordance to the ratings and price of the product to find proper accurate calculations.

1. INTRODUCTION

Machine learning predict larger knowledge from past. Machine learning focuses on the incidence of change programs once interacting with new knowledge, using straightforward machine learning algorithms. Special algorithms area unit used in the employment and prediction method. It provide training rules with the algorithm, that the rule uses this training knowledge to build predictions concerning various take a look at knowledge. Machine learning is usually divided into three classes.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd

This includes supervised learning, unsupervised learning and reinforcement learning. The computer file is allotted to a monitored learning program, thus affiliate labeling should be labelled before a personal to search out the information [1] [2]. Unsupervised provides a training rule. This rule finds the agglomeration of the computer file. Finally, reinforcement learning interacts dynamically with its atmosphere and receives a positive or feedback to enhance its performance.

Scientists use a spread of machine learning algorithms to get Python models that cause purposeful statistics. At the best level, these totally different algorithms area unit usually divided into 2 groups, that support the method of "studying" knowledge to formulate predictions: supervised and unsupervised . Classification is that the method of estimating a given information class, categories area unit referred to as goals / labels or classes. Classification Prediction Modeling is that the task of hard the mapping operate from input variables (X) to distinct output variables (Y).

2. RELATED WORK

In a survey conducted by choudhary and Rishab the neural influences of a product appearance and performance are tested and studied. When the resources are limited, the customer appears to be the supreme and requires much time to finalize a product. Comparing with Han Et Al's consumers affirm ratio this is justified and concluded that they have to work on something to ease the decision. Some conditions eventually had higher purchase rate than others but had very minimal difference between them. Due to limited and one-way resources customers made decisions based on appearance than focusing on the performance. In order to resolve this issue 20 subjects were called and asked to remember a 20-digit number in 20s. The subjects were asked to buy a range of products which were displayed with their appearance and performance information and all their actions were recorded [1].

The response of the subjects was divided into few categories or conditions. The conditions namely being more attractive with high performance, more attractive with less performance, less attractive and high performance and less attractive and less performance. Once the purchase was completed the recorded results were evaluated and concluded that the products with the conditions of more attractive with high performance and more attractive with less performance had high purchase rates than other conditions. This tells us about the manner in which the customer approaches on purchasing a product [2].

In a survey conducted in Vietnam, there was huge competition among the coffee suppliers with the booming of Coffeehouse chains. In order to meet the demands of the coffee drinker i.e. Customer and also to expand the shares in market, eventually the suppliers shifted heads on promotional activities and improving the quality. Moreover, it was found that coffee beverages with better taste and meagre price satisfy the customer immensely and are more evident to be purchased next time. This aimed to utilize the questionnaire-based approach to interact with customer and then the grey model GM (O, N) is implemented and the impact of marketing is evaluated [3].

As the approach was implemented, around 220 responses were recorded. Out of the recorded total data nearly 176 questionnaires reflected that five elements of the marketing mix impacted to the customer decision on purchasing the coffee. Among all the responses it was widely weighted that promotions and products were evaluated as the most crucial and important factors. This study concluded that the customers more over are highly attracted to better promotional activities and taste while purchasing the coffee beverages. Thus, this can be implemented to gain immense profits and make the customer more intended to buy your product [4].

[5] In the widespread of technology and ecommerce customers are more exposed towards online purchases. Online marketing has emerged as a huge industry especially for fashion products and related apparels. This study talks about that the customers are looking for new information such as comparison among products and live and real customer feedbacks. We are more likely to be impressed with the experiences of other consumers that have bought and used the same product than by believing in the official product details given by distributors or manufacturers. Customers are often drawn more by visual data than by textual content, such as images of clothing fitted to mannequins.

[6] This research also addresses a variety of issues about the Engel-Kollat-Blackwell Customer Purchase model theory. The questions centered on the effect of the customer's actions on the collection of information when they were engaged in shopping. Several of the studies indicated that, if consumers are happy with external details, they may end up purchasing the product without even trying it out. It revolves around the impacts of personality of the customer finding that they are majorly interfered by the involving risk and customer's inability to attain the information. This research used sample which consists around 350 experiences of customers buying product online. Based on the experiences, it concludes that the behavior is related to making decisions with the help of information like visual, feedbacks, promotions, etc. other than the performance and personal risk.

[7] This study focused on repeated purchasing of a product by a customer. The acceptance value and satisfaction are the determining factors for repeated purchase. The customers trust and quality are likely to impact a lot. These revelations were consistent with past researches and studies. As per this study, this value helps to increase sales for the managers with repeated purchase. The model introduced ways to promote this value and also suggested the need of an attractive dynamic website. So, the manufacturers and retailers started focusing on websites with added designs, security and worked on to provide a customer friendly user interface to attract customers. They also focused to improve the modes of communication to build that trust among customers.

One of the surveys talks about the challenges and exposure that digitalization presents to the market. It researches about the factors which influence the customer purchase intention. The study is conducted on different varieties and channels of products. The results bring out three factors that influence the most namely, social involvement, performance expectations and security understanding. Social influence being the strongest. Social influence plays a key role to add the value of a particular product involving family, friends and other trusted person. The introduction of well-designed mobile applications and social media revolutionized the business models and its behaviors. It makes the handling difficult with several new challenges [8].

This revolution brings in new ideas for companies to store and manage their products. They also look forward to expanding to other channels creating a multi-level company. The companies need work personnel to maintain as each product has different characteristics. The customers are skeptical about their own security and for trustable, authorized transactions. The study focuses on improved security and assurance of customers when developing to a multi-product industry. As in the conclusion it highlights the immense impact of social influence for customer purchase intention behavior especially for multi-product or multi-channel companies [9].

A theoretical model was made and survey conducted in China to learn about the factors that influence the customer about remanufactured electronic products. The study enlarged the factors of predicting with addition of perceived performance quality, benefits and knowledge to the original model. It aimed to benefit academicians and researchers by presenting a guideline about customer acceptance about the remanufactured goods. Remanufacturing is an important step as this component encourages circular economy and boosts as an engine of the supply chain process. The survey collected data by asking customers and asked them to test their model via a questionnaire. The results reflect that attitude, rules, control and knowledge of the customer plays the key role and provides insights about their purchase. The efficient quality and knowledge bring in positive attitude and impacts the sales. Risk can be negative and positive and impact attitude and the intention. The customers are scared to invest on remanufactured good and are tensed about their quality and performance hence reluctant to pay high price [10] [11].

The benefits always have an honest and good impact on the purchase intention but risk always had bad or negative influence on the purchase intention. The decision tree states that risk has negative influence whereas quality and performance having positive influence proves the hypothesis. Risk includes several components like risks that include service, price benefits, experimental risks, price hike risks, etc. The reviews that are generated by the customers already purchased the product influences a lot. In this study the reviews were obtained from Tianmao and choice tree approach was used to analyze the reviews. The collected data produced results which reflected that risk serves as a

negative tendency to purchase whereas the benefits acts as a positive tendency and also has a wide influence. The benefits like quality, performance, ratings and price have altogether different influence on the prediction of customer intention. It concludes to instruct the retailers and companies to increase customer satisfaction and eventually expand their sales and purchase rate [12] [13].

3. PROPOSED METHODOLOGY

We used data miner tool for web crawling or web scraping product details from various e-commerce sites like Amazon, Flipkart, Jabong, Myntra etc. Web scraping is an advanced tool to collect a vast number of data from webs. The data on the websites is unstructured. Web scraping aims to identify and store these unstructured data in a standardized way [14] [15].

When we execute the web scraping application, a query will be sent to the URL you listed. The server sends the data as a response to the request and enables you to access the HTML tab. The application then parses the data from the HTML file and removes it. Prediction of consumer intensity by supervised machine learning as a decision tree algorithm prediction is useful for the following purposes: it produces better performance in a classification task.

The following papers discusses how the machine learning algorithms are used in healthcare applications [15] [17] [19] [20]. This results from a bag estimation error that has been shown to be objective in many experiments and is fairly easy to figure with. Figure 1 show the workflow of customer purchase intension model.

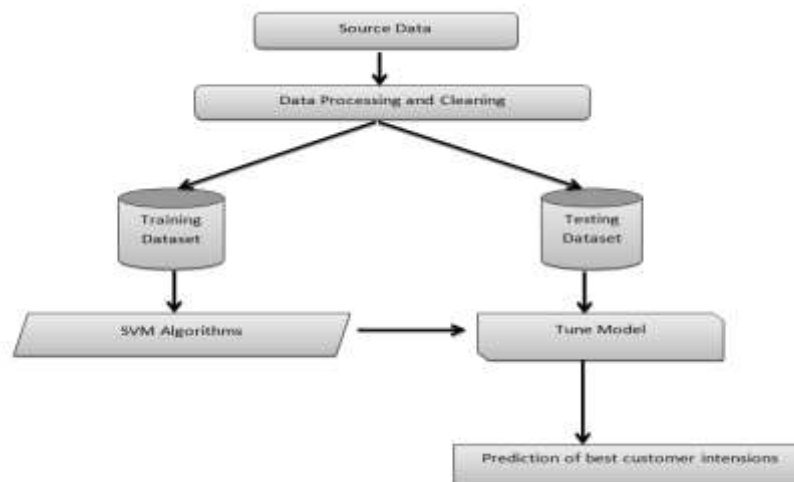


Figure 1: Workflow of customer purchase intention model

3.1. Scraping Data

We have to scrape Flipkart website using data miner extension for distillation of Price, Name, Rating and etc. So, scrutinize the page, under identification of info we would like to scrape is accumulated. To scrutinize the page distillate the Product's Price, Product Name, and the product rating which is accumulated in the ID identification. After distilling the information, you might want to save in a certain way it will be saved in CSV format. Figure 2 shows the conversion of unstructured data to structure from web pages using Web scrapping.



Figure 2: Web scrapping of data from webpages.

3.2. Data Validation and Preprocessing Technique

Data Validation/ Preparing Process: The dataset which has been extracted is imported to the packages included in the library. The libraries analyses the dataset and helps to remove delicacy, fix the missing values and shape the data in order to perform operations and yield accurate results[21] [22].

Data Pre-processing: The extracted data needs to be fixed before the algorithms are applied to it. This is an important step and is called cleaning or preprocessing of the data. It is a systematic process that helps to convert the outlier or data to a clean set of data that can be analyses. The extracted raw data is not feasible for analysis and this step helps to enable the process with an ease.

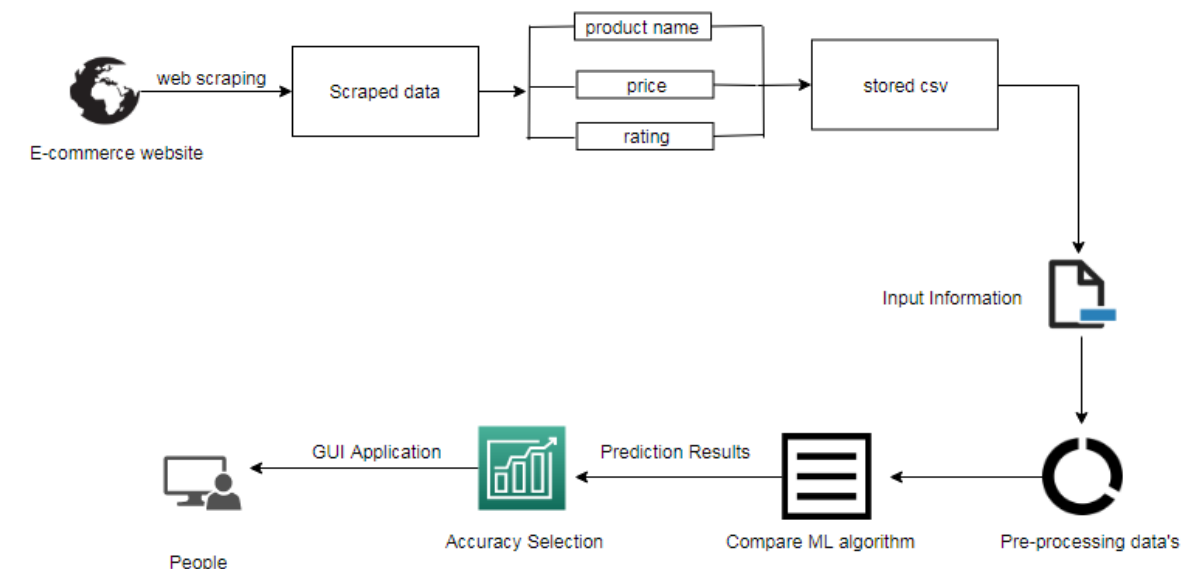


Figure 3. Proposed architecture for enhanced customer purchase retention Model

3.3. Data Visualization and Training Dataset

To train a model by given attribute with perception. Statistics requires undoubtedly specialize in quantitative elaboration of estimations of knowledge. Data visualization provision a momentous suite of tools for gathering a conditional understanding. This can be important when inspecting and operating to know a dataset and may help with knowing patterns, bad data, and many more [24][25].

Once the extracted data is cleaned and fit for analysis the classifying and regressing processes take place. The process yields the results and helps the user to present in a statistical table or in graphs. The visualized results helps in better understanding of the results. This process of presentation is called Data Visualization. It is also used to get the summary of the entire data in a numerical form.

3.4. Performance Measurement of' SVM

Support Vector Machines (SVM): A Semantic class that order informational index by setting a perfect hyperplane between orientations. We used this technology since this is perfectly adaptable inside this quantity of different kernelling capacities which will be applied and this model can yield a high consistency rate.

3.5. Web Based Application

Web application allow businesses to streamline their processes, increase productivity and reduce costs. In order to simplify the results, we use a user interface as a web application to display the results and also classify the user reviews into positive, negative and neutral reviews.

4. RESULTS

The findings are used to predict the purpose of the customer that is graphically portrayed with various products used by customers and its product rating. Retailers, as end-users, must produce outcomes in a concise way so that they can be interpreted by all. In order to visually view the results in easier way, we use the web application, to show the results and also to identify the user feedback as positive, negative and neutral feedback. We have used Linear Support vector machine for assessing of customer products with ratings and also compared the results of linear SVM with all other Machine learning models. Figure 4 shows the customer intension of products with ratings. Figure 5 shows the detailed product information page and figure 6 shows the Comparison of SVM-Linear with other algorithms in customer purchase intension where SVM-linear gives more accuracy compared to other ML algorithms.



Figure 4. Customer Intension Products by Ratings



Figure 5. Detailed Product Information Page

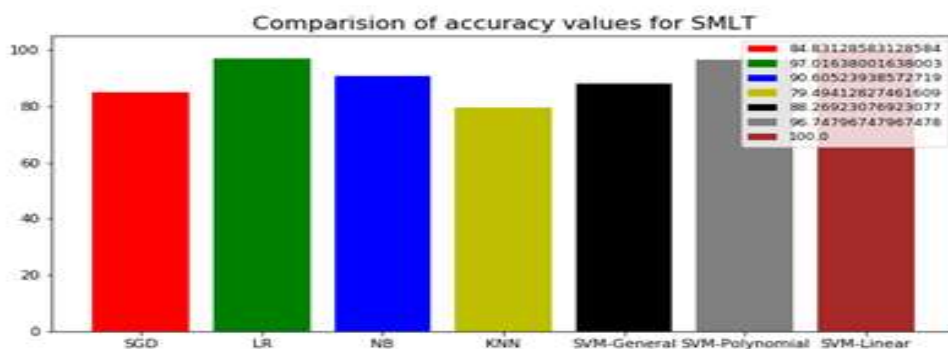


Figure 6. Comparison of SVM-Linear with other algorithms in customer purchase intension

5. CONCLUSION AND FUTURE ENHANCEMENT

With the growth of the e-commerce industry, the consumer plays a critical part in the purchasing of such a product. There are a lot of data available on the internet that can be used for research. The data is collected from the internet so as to go through the basic steps of machine learning algorithms. Data preprocessing is followed by data cleaning and data analysis so that the algorithms run on the provided data to produce results. The results are used to predict the customer intention which is represented graphically. The retailers being the end users, the results have to be given in a simplified way so that it can be understood by everyone. In order to simplify the results, we use a user interface i.e. a webpage to display the results and also classify the user reviews into positive, negative and neutral reviews. The proposed work is completely based on the data available from the websites and the training data that is produced during the algorithms. With rise in ecommerce and customer oriented companies and demands this is more likely to be effective in the future.

The advancement in technology, cloud computing and e-commerce will lead to the rise in customer generated results. This process can be further automated to show the result via a desktop application. The work can also be optimized to be implemented in Artificial Intelligence Environment. This

process can also help in future technologies like sales force and supply chain management which depends on the feedbacks or results of the customer.

REFERENCES

- [1] Choudhary, R. and Gianey, H.K., 2017, December. Comprehensive review on supervised machine learning algorithms. In *2017 International Conference on Machine Learning and Data Science (MLDS)* (pp. 37-43). IEEE.
- [2] Stamp, M., 2017. *Introduction to machine learning with applications in information security*. CRC Press.
- [3] Dasgupta, A. and Nath, A., 2016. Classification of Machine Learning Algorithms. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, 3(3), pp.6-11.
- [4] Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J.O., Olakanmi, O. and Akinjobi, J., 2017. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), pp.128-138.
- [5] Morwitz, V., 2014. Consumers' purchase intentions and their behavior. *Foundations and Trends® in Marketing*, 7(3), pp.181-230.
- [6] Han, W. and Bai, H., 2018, July. Influence of Cognitive Resource Limitation on Consumer Purchasing Decision: An Event-related Potentials Perspective. In *2018 15th International Conference on Service Systems and Service Management (ICSSSM)* (pp. 1-4). IEEE.
- [7] Chai, Y.C., Huang, Y.F. and Dang, H.S., 2017, July. Determining the effects of marketing mix on customers' purchase decision using the grey model GM (0, N)-case study of the western style coffeehouse chains in Vietnam. In *2017 International Conference on System Science and Engineering (ICSSE)* (pp. 575-578). IEEE.
- [8] Chen, Y.C., Wu, H.C., Lee, Y.H., Sung, Y.C. and Chen, H.Y., Online Apparel shopping Behavior.
- [9] Chuang, L.W. and Chiu, S.P., 2018, May. Exploring Consumers' Repeat Purchase Intentions in Green Economy. In *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)* (pp. 1-2). IEEE.
- [10] Susanto, H., Suchahyo, Y.G., Ruldeviyani, Y. and Gandhi, A., 2018, October. Analysis of factors that influence purchase intention on omni-channel services. In *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 151-155). IEEE.
- [11] Ma, L., Su, X., Wang, C., Lin, K. and Lin, M., 2017, June. Consumers' intention to purchase remanufactured electronic products: An empirical study in China. In *2017 International Conference on Service Systems and Service Management* (pp. 1-6). IEEE.
- [12] Liu, K., Xu, L. and Zhao, J., 2014. Co-extracting opinion targets and opinion words from online reviews based on the word alignment model. *IEEE Transactions on knowledge and data engineering*, 27(3), pp.636-650.
- [13] Denecke, K., 2008, April. Using sentiwordnet for multilingual sentiment analysis. In *2008 IEEE 24th international conference on data engineering workshop* (pp. 507-512). IEEE.
- [14] Sun, F., Belatreche, A., Coleman, S., McGinnity, T.M. and Li, Y., 2014, March. Pre-processing online financial text for sentiment classification: A natural language processing approach. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)* (pp. 122-129). IEEE.
- [15]
- [16] Saranya, G., & Pravin, A. (2020). A comprehensive study on disease risk predictions in machine learning. *International Journal of Electrical and Computer Engineering (IJECE)*, 10(4), 4217.
- [17] Gupte, A., Joshi, S., Gadgul, P., Kadam, A. and Gupte, A., 2014. Comparative study of classification algorithms used in sentiment analysis. *International Journal of Computer Science and Information Technologies*, 5(5), pp.6261-6264.

- [18] Saranya, G., Geetha, G., & Safa, M. (2017). E-antenatal assistance care using decision tree analytics and cluster analytics based supervised machine learning. 2017 International Conference on IoT and Application (ICIOT).
- [19] Cambridge, U.P., 2009. Introduction to information retrieval.
- [20] Meenakshi K, Safa.M, Krthick T, Sivaranjani N.” A novel study of machine learning algorithms for classifying health care data”, Research Journal of Pharmacy and Technolgy, Research Journal of Phamacy and Technology, 2017.
- [21] Meenakshi K, Maragatham, G, Agarwal, N., Ghosh, I, “ A Data mining Technique for analysing and predicting the success of movie”, Journal of Physics: Conference Series, Vol 1000, Issue 1, 2018.
- [22] Gamallo, P. and Garcia, M., 2014, August. Citius: A NaiveBayes Strategy for Sentiment Analysis on English Tweets. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*.
- [23] Saranya, G., Lavanya, S., & Sivasankari, S. (2018). An efficient power saving technique based location alarm for smart phones. *Journal of Physics: Conference Series*, 1000, 012121.
- [24] Nithya, M., & Sheela, T. (2018). Predictive delimiter for multiple sensitive attribute publishing. *Cluster Computing*, 22(S5), 12297-12304
- [25] G Geetha, M Safa, C Fancy, D Saranya “A hybrid approach using collaborative filtering and content based filtering for recommender system”, Journal of Physics: Conference Series Vol 1000, Issue 1, 2018.
- [26] M Srivastava, S Pallavi, S Chandra, G Geetha,”Comparison of optimizers implemented in Generative Adversarial Network (GAN)” International Journal of Pure and Applied Mathematics Volume 119 Issue 12,2018

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.