

A Review on Consumer Behavior towards Online Shopping using Machine Learning

Waqas Haider Bangyal^{1*}, Adnan Ashraf², Rabia Shakir³, Najeeb Ur Rehman⁴

^{1*}Dept. of Computer Science, Kohsar University Murree, Punjab, Pakistan

²School of Computer Science, Beijing Institute of Technology, Beijing, 100081, China

³Dept. of Computer Science, Federal Urdu University of Science and Technology, Islamabad, Pakistan

⁴Dept. of Computer Science, University of Gujrat, Pakistan

*Corresponding author

Abstract

Study of consumer behavior in online shopping, as a rule, manages identification of consumers and their purchasing behavior. The purpose of such studies is to verify who purchases where, what, when, and how. The analysis of such consumer behavior is useful to get the buyer's prerequisites and requirements for their future aims towards the product. Through this review, E-commerce organizations can follow the utilization and sentiments appended to their items and adopt suitable promoting strategies to give a customized shopping experience to their buyers, consequently expanding their hierarchical benefit. This paper purpose to utilize information-driven promoting models, for example, information perception, natural language processing, and AI models that help in getting the demographics of an association. Additionally, make recommender frameworks through cooperative filtering, sentiment analysis, and neural networks.

Keywords: Sentiment Analysis; Machine Learning; Natural Language Processing; Data Visualization; Market Strategies.

1. Introduction

The use of E-commerce has immeasurably expanded in this modern innovative era. People like to shop online as opposed to shopping in business sectors. In online business, the information created by consumers as reviews and surveys of a specific item can be utilized for authenticity and exposure of the

item. Customers settle on whether to purchase a given item by checking out these evaluations and surveys. Such happiness can be positive or negative reviews made by customers/consumers who have recently utilized the item. An exact investigation of this consumer-produced content can be useful to the online business association to acquire bits of knowledge and get their consumer's expectations and necessities (T.Yoshida,M.Hasegawa,T.Gotoh,H.Iguchi,K.Sugiokaand K. Ikeda,, 2020).

ML Algorithms [27-32] can assist us with plotting exact visual portrayals of such customer conduct. Sentimental analysis is regularly used to assemble a social conduct diagram on human internet-based conduct to observe the connection among exchanging and volume costs of stocks and products. Sentimental analysis was a help to perform on the information removed from SentiWordNet utilizing a hybrid selection model to show what market patterns mean for item notoriety and rate (Valecha, 2018). Different RS learning models have been created for cooperative sifting. A model is the wide and profound learning model that consolidates the strength of element remembrance of the straight model and element speculation of the profound learning model. It is proposed for suggesting applications for the Google Play store. In this way, it is vital to comprehend the reason why, when, how, and what different variables that impact purchasing choice of the customers. For anticipating buyer conduct keen strategies are required other than figuring out what they had bought before. Purchasers can group as indicated by their genuine way of behaving (A.H.Kumar, F.S.John, & S. Senith, 2015).

These days, large undertakings like Siri Apple, Eco Amazon, Google, Facebook, and Microsoft utilize AI methods to naturally examine their customer's shopping conduct and offer suitable support for them. Be that as it may, utilizing AI strategies isn't confined to projects with immense examination financial plans. Little and Medium Enterprises (SMEs) could profit from these strategies to work on their productivity also. Thus, it is vital to comprehend the reason why, when, how, and what different elements that impact purchasing choice of the shoppers. Foreseeing shoppers conduct keen strategies required other than figuring out what they had bought before. Shoppers can characterize as indicated by their real way of behaving (H. T. Cheng, L. Koc, J. Harmsen, et al, 2016).

The utilization of E-commerce has boundlessly expanded in this modern mechanical world. Consumers like to shop online as opposed to shopping markets. In online business, the information created by customers as appraisals and surveys of a specific item can be utilized for legitimacy and exposure of the item. Purchasers settle on whether to purchase a given item by checking out at these appraisals and audits. Such satisfaction can be positive or negative surveys made by purchasers who have recently utilized the item. A precise examination of this client-produced content can be useful to online business

associations to acquire experiences and get their buyers' goals and necessities. AI Algorithms can assist us with plotting exact visual portrayals of such shopper conduct. Such visual portrayals would help study the dataset in a more itemized way to comprehend and make specific finishes of the general customer conduct on the online business stage. Natural Language Processing, a sub-field of AI, is utilized to examine the text and recognize positive or negative audits given by the buyers. This is likewise called sentiment analysis (P. Covington, J. Adams, and E. Sargin, 2016).

This paper recommends that there are a few elements that influence the purchasing choice of customers. To improve and deal with various wellsprings of upgrading client relationships it is vital to know once conduct. There are a few elements like outer factors, for example, social variable and social element which relies upon ethnicities, geographic districts, racial gatherings, religions, pay, calling, and training while referencing gatherings, family, job, and status. Inward factors, for example, age, calling, instruction, pay, character, and way of life are relied upon what to purchase for sure not to purchase. Some of the time purchasing conduct relies upon mental factors, for example, insight, inspiration, gaining from previous experience, convictions, and disposition.

2. Literature Review

The literature review has been committed to showing the exact visual portrayal of buyer conduct to such an extent that it covers every one of the areas of premium expected for an internet business organization to make upgrades in their items and advertising methodologies. Past studies have focused on techniques on how friendly research can be a significant element in deciding the purchasing conduct of any client. This would incorporate the consumer's emotional, or objective conclusions on items. Snowball inspecting was utilized to separate profoundly persuasive clients (Kwanruan Rusmee, Narumol Chumuang, 2021).

Sentiment Analysis was applied to tweets separated from Twitter and news features to create new indicators for investment (Bin Wen at. Al). From the gathered information, they pick an irregular example and characterized each tweet as bullish or negative assuming that it contains those terms. They showed that Twitter feeling marker and the event of money-related terms on Twitter are measurably critical indicators of normal market returns. Sentiment examination was likewise performed on miniature writing for a blog administration committed to the stock market.

(M.P. Rajakumar, Dr. V. Santhi) They gathered 62,100 blog entries from stocktwits.com, for some time of 90 days. The feeling of the posts was arranged to utilize an AI calculation known as the J48 classifier to create a learning model (Kwanruan Rusmee, Narumol Chumuang, 2021).

They demonstrated that the mined opinion has a solid insightful incentive for coming business sector bearings. Sentiment Analysis was utilized to estimate the end record of Tata Services and exactness of 85.99% was found simultaneously. Sentiment investigation is frequently used to construct a social conduct diagram on human's web-based conduct to observe the relationship among exchanging and volume costs of stocks (Jiajia Li). Sentimental Analysis was additionally performed on the information removed from Senti-WordNet utilizing a crossbreed choice model to show what market patterns mean for an item prominence and rate (MICHAEL M. TADESSE, 2020; A.Ashraf, 2021; W.H.Bangyal, 2021;)[24-25].

(I. M. W, Dr Rizwana Bashir) The way of behaving of consumers shopping over the web has been broken down into five variables. These variables are time, protection, trust, comfort, and item assortment. The exploration led for this design was a survey and the outcomes were examined measurably. Trust was asserted as such a human quality that influences their purchasing propensities (Martin, 2020).

The shopping conduct of individuals in Pakistan gets impacted by their mental and passionate mentalities. Protection is additionally considered as the most noticeable variable in this online shopping pattern in the way that individuals may now and then not have a real sense of reassurance sharing their data over the web. Then again, entrancing costs of different stuff may likewise draw in individual consideration and help to ask online shopping. Thus, trust over the source is the issue that influences individuals purchasing ways of behaving (H. u. R. I. J. Sajjad Nazi) (Kapoor, 2020; W.H.Bangyal, 2021; A.Ashraf, 2021; W.H.Bangyal, 2022)[22,23,26].

Sr no.	Authors	Title	Year	Accuracy rate	Research gap
1	Harsh Valecha, Aparna Varma	Prediction of Consumer Behavior using Random Forest Algorithm	2018	89%	This paper proposes a time-evolving random forest classifier that leverages unique feature engineering to predict the behavior of consumer that affect the choice of purchasing the product significantly.
2	Vishwa Shrirame	Consumer Behavior Analytics	2020	0.93%	Employ data-driven marketing tools, such as data visualization, natural language processing, and machine

		using Machine Learning Algorithms			learning models that help in understanding the demographics of an organization.
3	Sunil Kumar Khatri	Using Sentimental Analysis in Prediction of Stock Market Investment	2016	90%	Sentimental analysis is performed on the data extracted from Twitter and Stock Twits.
4	Aly Ahmed, Ljiljana Trajković	Modeling Prediction in Recommender Systems Using Restricted Boltzmann Machine	2017	78%	Employ the Restricted Boltzmann Machine (RBM) for collaborative filtering and propose the neighborhood-conditional RBM (N-CRBM) model based on joint distributions of similarity and popularity scores.
5	Kwanruan Rusmee, Narumol Chumuang 3	Predicting System for the Behavior of Consumer Buying Personal Car Decision by Using SMO	2020	95.13%	Data analysis using SMO techniques for predict system for the behavior of consumers buying personal car decisions.
6	Michael M. Hongfei	Personality Predictions Based on User Behavior on the Facebook Social Media Platform	2018	78.6%	Analyze and compare four machine learning models and perform the correlation between each of the feature sets and personality traits.
7	Fatemeh Safara	A Computational Model to Predict Consumer Behavior During	2020	95.3%	The prediction model is proposed to anticipate the consumer's behavior using machine learning methods. Five individual classifiers and their ensembles with Bagging and Boosting are examined on the dataset collected from an online shopping

		COVID-19 Pandemic			site
8	Shruti Goyal, K. P. Miyapuram , Uttama Lahiri	Predicting Consumer's Behavior Using Eye- Tracking Data	2015	89%	Investigate the components of eye-tracking data that can predict consumers' final choices with better precision.
9	Valentina Chkoniya	Challenges in Decoding Consumer Behavior with Data Science	2020	89%	contribution is explaining the principles of different machine learning techniques and their applicability in various real-world application domains, such as cybersecurity systems, smart cities, healthcare, e-commerce, agriculture, and many more.
10	Seren Başaran and Obinna H. Ejimogu	A Neural Network Approach for Predicting Personality from Facebook Data	2019	85%	Combines the inference features highlighted in three different relevant studies which are; the number of likes, events, groups, tags, updates, network size, relationship status, age, and gender.
11	J. Scott Armstrong	Prediction of Consumer Behavior by Experts and Novices	2012	69.4%	To obtain 20 articles that met the four criteria, this research examined 94. Thirty-two percent exceeded 10 pages, 40 percent did not test any hypotheses, and 6 percent lacked clarity
12	Chun- Cheng Peng, Yuan-Zhi Wang	Artificial- Neural- Network- Based Consumer Behavior Prediction	2020	94.73%	Comparison of the differences between AI and statistical systems

3. Datasets

To analyze character characteristics from informal communities, this research utilized the myPersonality dataset as a contextual investigation. This research developed the review with 250 clients and 9917 announcements from the my Personality test. The dataset of Facebook clients was named by the Big 5 model. As indicated by the dissemination of character types in Table, every client in the dataset had numerous posts accumulated in one file (Singh, 2021).

With some of the datasets, this research chose the consumer data, for example, the consumer's interpersonal organization structure, client's status, and text posts. The final dataset contains the Facebook situations with the crude message, writer data, character names (scores and classes), and five informal community proportions of the clients in surmising with character qualities, that is network size, betweenness, thickness, financier, and transitivity (Zhai, 2020).

4. Outcomes

The dataset comprises Amazon item audits sorted by different classifications like apparel items, home and kitchen items, mobiles, and substantially more. For our review, this research has zeroed in on a particular sort i.e Home and Kitchen items to deal with. The Count Vectorizer is used to tokenize a collection of text documents and build a vocabulary of all the words in that document. This vocabulary helps in encoding new documents. The focus is on the occurrence of words in that document to find out a pattern. This research has used the following classification models for analyzing the sentiment score of each associated product review. In all of these models, 80% of data is used for training and 20% is used for testing purposes. This classification model is used to find K nearest neighbors of a particular data point, according to a certain distance measure, and decide which category it belongs to. Random forests or random decision forests are an ensemble learning method that takes multiple learning algorithms or multiple instances of an algorithm and puts them together to give the best possible result. The SVM classification model is generally used to classify between different categories using some decision boundary. A decision boundary separates the points into two classes.

5. Conclusion

The goal of this research is to aid e-commerce organizations in analyzing the sale of products and also gain insights into customer intentions while purchasing any particular product. These studies help the organizations in knowing their customers and incorporating targeted marketing techniques to increase

their customer base and profits (Martin, Vincent). To understand consumer interactions with e-commerce sites, this research employed various data visualization techniques and did a thorough sentiment analysis on product reviews. Sentiment analysis helped us evaluate consumers' sentiments related to various products which in turn helped us analyze the product's performance in the market (Surendro, 2019). This research used five classification models for sentiment analysis of reviews with the help of natural language processing and obtained the best results in the Naive Bayes classifier. This research also trained an LSTM network which went one step further and found out the amount of positive and negative sentiment in the review, which helps in understanding neutral reviews as well.

6. Future work

This research employed three recommender systems using nearest neighbors, a priori, and Boltzmann machines. The objective of a recommender system is not to get the best accuracy, but to recommend similar items to a consumer based on their recent purchase activity. This personalized approach to a consumer's needs increases their trustworthiness towards the website and influences them to buy more products. Nearest neighbors helped us filter out similar products and a priori algorithms helped us gather item sets that are frequently bought together. The Boltzmann machine helped in predicting the ratings of all the products that the consumer hasn't bought yet, and gives the most promising result among all. Using all these results in this research concluded that any customer decision is influenced by multiple aspects when he/she is using any e-commerce platform for shopping. Such factors can be identified by e-commerce organizations and they can take necessary steps to facilitate better service to the customer and add more lifetime value to their business.

References

- [1] T.Yoshida,M.Hasegawa,T.Gotoh,H.Iguchi,K.Sugiokaand K. Ikeda,, "Consumer behavior modeling based on social psychology and complex networks," The 9th IEEE International, " *Conference on E-Commerce Technology* , no. 6th edition, 2020.
- [2] H. Valecha, "Prediction of Consumer Behaviour using Random Forest Algorithm," 2018. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8597070>
- [3] A.H.Kumar, F.S.John, & S. Senith, "A Study on factors influencing consumer buying behavior in cosmetic Products," *International journal of scientific and research*, vol. 4, no. 9, pp. pp. 1-6, 2015.
- [4] H. T. Cheng, L. Koc, J. Harmsen, et al, "Wide and deep learning for recommender systems," 2016. [Online]. Available: ACM Workshop on Deep Learning for Recommender Syst., Boston, MA, USA,
- [5] S. K. Singh, "A Machine Learning Model for Effective Consumer Behaviour Prediction," 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9702495>

- [6] Kwanruan Rasmee, Narumol Chumuang, "Predicting System for the Behavior of Consumer Buying Personal Car Decision by Using SMO," vol. 2, no. 3, pp. 10-15, 2021
- [7] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for YouTube recommendations," *ACM Conf. Recommender Syst., Boston, MA, USA,*, pp. pp. 191-198, 2016
- [8] X. Zhai, "Prediction Model of User Purchase Behavior Based on Machine Learning," 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9233677>
- [9] K. Surendro, "Predictive Analytics for Predicting Customer Behavior," 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8834571>
- [10] V. Martin, "Predicting the french stock market using social media analysis," *International workshop on sentimentic*, pp. pp/3-7, 2020
- [11] N. Kapoor, "A Review on Consumer Behavior Prediction using Data Mining Techniques," 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8991415>.
- [12] K. Yada, "Prediction of Consumer Purchasing in a Grocery Store Using Machine Learning Techniques," 2016. [Online]. Available: <https://ieeexplore.ieee.org/document/7941935/similar#similar>
- [13] S.R. Sujithra; R.K. Kavitha, "Consumer Online Buying Behaviour - A Perspective Study Analysis using R," 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9675566>
- [14] Muqaddas Gull; Arshi Pervaiz, "Customer Behavior Analysis Towards Online Shopping using Data Mining," 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8467262>.
- [15] M. Kosinski, S. C. Matz, S. D. Gosling, "Facebook as a research tool for the social sciences: Opportunities, challenges," Vols. 70., no. 6., pp. 543, 2015, 2017
- [16] Harsh Valecha, Aparna valecha, "Prediction of Consumer Behaviour using Random Forest Algorithm," *Prediction of Consumer Behaviour using Random Forest Algorithm* , vol. 3, no. 5, pp. 5-9, 2018
- [17] Hanene Ben Yedder and Umme Zakia, "Modeling Prediction in Recommender Systems Using Restricted Boltzmann Machine," vol. 2, no. 4, pp. 7-12, 2017
- [18] Muqaddas Gull and Arshi Pervaiz , "Customer Behavior Analysis Towards Online Shopping using Data Mining," vol. 4, no. 2, pp. 6-9, 2018
- [19] Sunil Kumar Khatri, Ayush Srivastava, "Using Sentimental Analysis in Prediction of Stock Market Investment," vol. 2, no. 1, pp. 13-19, 2016
- [20] Vishwa Shrirame, Juyee Sabade , "Consumer Behavior Analytics using Machine Learning Algorithms," vol. 2, no. 3, pp. 3-6, 2020
- [21] MICHAEL M. TADESSE, , "Personality Predictions Based on User Behaviour on the Facebook Social Media Platform," vol. 2, no. 3, pp. 1-4, 2020.
- [22] Bangyal, W.H.; Nisar, K.; Ag. Ibrahim, A.A.B.; Haque, M.R.; Rodrigues, J.J.P.C.; Rawat, D.B. Comparative Analysis of Low Discrepancy Sequence-Based Initialization Approaches Using Population-Based Algorithms for Solving the Global Optimization Problems. *Appl. Sci.* 2021, *11*, 7591. doi: 10.3390/app11167591
- [23] Ashraf, A.; Pervaiz, S.; Haider Bangyal, W.; Nisar, K.; Ag. Ibrahim, A.A.; Rodrigues, J.P.C.; Rawat, D.B. Studying the Impact of Initialization for Population-Based Algorithms with Low-Discrepancy Sequences. *Appl. Sci.* 2021, *11*, 8190. doi: 10.3390/app11178190
- [24] W.H.Bangyal, A. Hameed et al, A New Initialization Approach in Particle Swarm Optimization for Global Optimization Problems " , Computational Intelligence and Neuroscience, Volume 21, PP:1-1
- [25] A.Ashraf, A. A. Almazroi ,W.H.Bangyal, et. al, "Particle Swarm Optimization with New Initializing Technique to Solve Global Optimization Problems", Intelligent Automation & Soft Computing , 2021
- [26] W.H.Bangyal, K. Nisar, et . al , " New Modified Controlled Bat Algorithm for Numerical Optimization Problem " , Computers, Materials & Continua Vol.70,Issue No.2, pp.2241-2259

- [27] W.H.Bangyal, J.Ahmad, " Evolving Artificial Neural Network Using Improved Bat Algorithm for Data Classification" in Journal of Medical Imaging and Health Informatics.
- [28] W.H.Bangyal, J.Ahmad, " An Improved Particle Swarm Optimization Algorithm with Chi-Square Mutation Strategy ", International Journal of Advanced Computer Science and applications(IJACSA).
- [29] W.H.Bangyal, J.Ahmad, ," A Modified Bat Algorithm With Torus Walk for Solving Global Optimization Problems" , in International Journal of BioInspired Computation (IJBIC), 2019
- [30] Waseem, Q.; Alshamrani, S.S.; Nisar, K.; Wan Din, W.I.S.; Alghamdi, A.S. Future Technology: Software-Defined Network (SDN) Forensic. Symmetry, MDPI, 13, 767, 2021.
- [31] Nisar, Kashif; Sabir, Zulqurnain; Zahoor Raja, Muhammad A.; Ag. Ibrahim, Ag. A.; Rodrigues, Joel J.P.C.; Shahid Khan, Adnan; Gupta, Manoj; Kamal, Aldawoud; Rawat, Danda B. 2021. Evolutionary Integrated Heuristic with Gudermannian Neural Networks for Second Kind of Lane–Emden Nonlinear Singular Models" Appl. Sci. 11, no. 11: 4725.