

Customer Behavior Analysis in E-Commerce using Machine Learning Approach : A Survey

Siddhant Sharma¹, Akhilesh A. Wao^{*2}

AKS University, SATNA, MP, India

* Corresponding Author- Dr. Akhilesh A. Wao, akhileshwao@gmail.com

ARTICLE INFO

Article History:

Accepted: 05 March 2023

Published: 27 March 2023

Publication Issue

Volume 10, Issue 2

March-April-2023

Page Number

163-170

ABSTRACT

These days, consumer behaviour models are often founded on machine learning and the data mining of customer data. Forecasting client behaviour is an unclear and challenging endeavour. Thus, one must use the appropriate strategy while constructing consumer behaviour models. Once a prediction model has been built, it is difficult to manipulate it for the marketer to determine precisely what marketing actions to take for each customer or group of customers. This is because once the model has been built, it is impossible to change the variables that make up the model. While this formulation is very complicated, most customer models are, in practice, relatively straightforward. Because of this requirement, most customer behaviour models neglect so many essential elements that their forecasts are often untrustworthy. The purpose of this study is to present different research work that has been done on the analysis of consumer behavior using various machine learning and data mining approaches.

Keywords: Customer, Machine Learning, Prediction, Accuracy, Error, Data Mining.

I. INTRODUCTION

The advent of the internet has had a significant impact on many aspects of our day-to-day lives. In today's internet-driven world, one of the most dynamic development sectors is electronic commerce. Consumers are eager to purchase goods from online retailers such as Amazon, eBay, and Flipkart, amongst others. Online sites additionally give the ability for consumers to post a review on things they purchase. Customers and sellers alike may benefit from the

decisions made based on these assessments about marketing techniques, as well as the enhancement of goods and services [1]. Today consumers are very much eager to read reviews before buying any goods and obtaining services. This opens the door for opinion spammers, who will submit bogus evaluations in an attempt to boost or lower the reputation of businesses, goods, and services.

This sort of action is sometimes described as Review spam.

The study of consumer behavior raises several basic concerns, including: Why do customers purchase the products that they do?

- How does the consumer acquire the product?
- In what manner does the product be consumed or used?
- How does the client create a product after purchasing it?
- In what way does the consumer become exempt from the product (or his packaging) after using it?

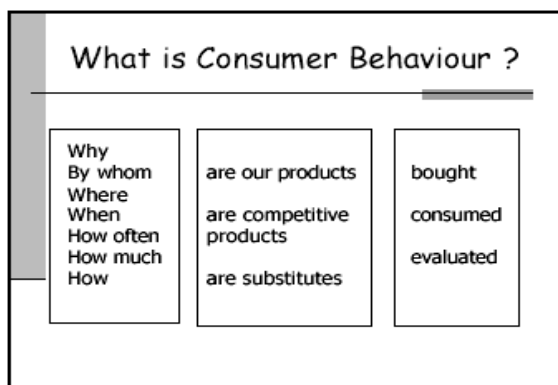


Figure 1: Customer behavior [Prof. Dr. Maggie Geuens, Customer Behaviour, 1999]

The investigation of the elements that affect the actions taken by customers yields the answers to these questions. These influences may be broken down into four distinct categories: the social, the cultural, the demographical, and the psychological [16].

The study of customer behavior begins with the assumption that consumers always base their choices on a certain quantity of information before making a choice. This knowledge may be broken down into two categories: internal (gained through one's own prior experiences), and external (type of product, word of mouth, etc.) According to this presumption, a business would be unable to successfully market a product if it did not have a solid understanding of the kind of information customers rely on when making purchasing decisions, as well as how customers perceive and make use of that information—in other words, the decision-making processes [5].

The processes that go into making a decision are heavily influenced by three major categories of

variables: those that are directly related to the customers themselves, those that are related to the context or situation in which the purchase is being made, and those that are related to the products or services that are being considered. The "fundamental trio" consists of these three different factors. This chapter devotes a significant amount of its content to discussing the decision-making processes that are used by consumers as well as the several methods in which the information that customers are likely to employ is processed.

II. BACKGROUND

E. Manohar et al.,[1] The client-generated information such as surveys, evaluations, and notes may be broken out for more prominent experiences that can be used by large businesses. It is helpful to understand the requirements of the client and to anticipate their future expectations towards the administration by investigating the buyer's behavior such as this. With the help of this psychological research, Online Business Associations can monitor the applications and outcomes associated with their products, as well as adopt appropriate marketing strategies to provide a personalized shopping experience for their customers, which ultimately increase their authoritative benefit. This study intends to leverage information-driven marketing tools such as information representation, common language training, and AI models that aid in understanding the socioeconomics of an organization.

V. Shrirame et al.,[2] Recent events have shown that the unmistakable proof of web-based media networks is a big cause for concern. This is because customers who take an interest in such networks might contribute to viral advertising efforts. We focus on clients' communication in our study, considering their nature as an important identifier for open organizations, which we define as networks with a large volume of data streams. We illustrate the

Twitter Character based Open People group Extraction (T-PCCE) framework by depicting it in a Twitter network chart. This framework identifies the most informative networks by taking into consideration the clients' personalities. After this, we expand upon previously established approaches as a part of the client character extraction process by combining data from several different aspects of client behavior and using AI techniques. We make use of an existing particularity-based network recognition computation, which we then expand by including a post-handling phase that removes chart edges depending on the nature of the customers.

B. Lebichot et al.,[3] The objective is to devise a system for doing forecasts inside a Cloud application that is based on AI indicators. Any improvement made to the accuracy of the projection will have a direct impact on the key execution markers for cloud providers as well as cloud occupiers and consumers. The results of the tests demonstrate that our method is capable of dealing with and improving cloud asset planning for a cloud server farm.

S. Shahriar et al.,[4] One of the brilliant applications for urban communities that is making gains these days is smart transportation, which is one of the sharp city applications that is transitioning from the stage of calculated models to the stage of progress. Electric vehicles, sometimes known as EVs, are increasingly being recognized as a crucial component of innovative mobility solutions. Electric vehicles are gaining popularity rapidly because of the predicted contribution they will make toward reducing dependency on ozone-depleting substances and petroleum derivatives. Despite this, the widespread distribution of EV charging stations creates several challenges to the power matrix and the public foundation. The basic plan of sending extra charging stations to increase charging capacity does not work as a solution to the problem of delayed charging time.

J. Edmond Meku Fotso et al.,[5] This is one of the primary factors that contributes to the high dropout rate, as well as the poor rate of contentment and accomplishment saw in MOOCs. The majority of these models focus on predicting dropout, finishing, and progress; however, the vast majority of these models don't generally give enough consideration to one of the key advances (student conduct), that precedes, and can explain exiting and disappointment. Many research works have recommended distinct, prescient, and prescriptive models to address this issue. Our study aims to construct a comprehensive learning model that can predict student behavior (student connections) in the learning cycle. This model will be used to provide students and course instructors with information and an understanding of student behavior in the learning cycle. The models were generated using L2 Regularization, and based on the results of the predictions; we were surprised to find that a model with simple RNNs had the greatest performance and accuracy on the dataset that was used as compared to the other RNN designs. We came to a few conclusions, one of which was that there is a connection between the video survey and test administration, as well as the level of participation of the student in the learning cycle.

P. A. Savenkov et al.,[6] This article analyses the development of numerical help and programming for recognizing irregular behavior of customers depending on biometric characteristics of their conduct evaluation. While developing sensible UBA (Client Conduct Examination) frameworks, one of the challenges that might arise is the challenge of safeguarding important data from vast amounts of unstructured, unmatched information. Methods and computations of intelligent data handling and artificial intelligence (AI) that are used in UBA/DSS frameworks aid to take on an endeavor of taking care of concerns of information analysis of varied directivities. The use of AI methods in the operation of a flexible UBA framework is an idea that has been

offered. Throughout the evaluation, the examination led to the formation of the rundown of the primary components that were subjected to the participation of the breaking down approaches. There have been two different methods described for spotting unusual behavior shown by customers. When used in intelligent UBA systems, the application of AI approaches will make it possible to anticipate data risks, detect, and prevent the theft of sensitive information by insiders at these organizations in advance.

J. R. Goodall et al.,[7] Despite the finest efforts of digital protection investigators, organized figuring resources are routinely undermined, bringing about the lack of permitted innovation, the exposure of state-protected insights, and big monetary damages. Irregularity identification algorithms are helpful for spotting new types of attacks and odd organization movements. Nevertheless, such computations might be difficult to grasp and it may be difficult to know whether to believe the results. Administrators of organizations and digital auditors require rapid and flexible tools to assist in recognizing suspicious behavior that circumvents automated security systems. Nevertheless, administrators do not need another automated gadget with calculations that they do not trust. The specialists require tools that will help them improve their area aptitude as well as provide them with a relevant understanding of questionable behavior that will assist them in making decisions. Situ is a visual assessment tool that we describe in this research for spotting questionable behavior in streaming company information. Situ provides a versatile setup that combines the discovery of inconsistencies with the interpretation of data.

D. Damkevala et al.,[8] This article provides a tutorial on how to use the Watson AI Programming interface on IBM Cloud to do serverless data analysis with the assistance of AI. Using advanced analysis procedures, such as applying an altered Mahalanobis Distance

computation for amalgamation of connection information within the realm of artificial intelligence, it ought to be able to transform the massive amount of data produced by an organization into insight. Using a Multivariate Dependability Classifier model, more refining of connection information is completed. The use of this high-level investigation service should be possible in a serverless manner, where the engineer is only responsible for worrying about how the data is broken down, such as scoring, cluster, or stream models with a persistent learning framework, without the expense of equipment whereupon those models can be prepared. The purpose of this study is to investigate the use of serverless computer-based intelligence frameworks in the context of client behavior analysis across varying socioeconomic conditions.

Asniar et al.,[9] The proliferation of the internet has led to the digitization of information, which has resulted in the development of several new information channels. Computerized information in large numbers leaves hints of what customers see, what they read, their inclusion and behavior, judgment, and about their preferences and inclinations to give a lot of information that can be dug up for learning experiences. This information can be mined for learning experiences. The tremendous value of the information may be found in the findings of the research as well as any predictions or actions that are derived from the findings of the research. Using data, doing calculations based on facts, and employing AI techniques, prescient analysis is the process of identifying prospective patterns, events, and practices in the future based on information that has been documented. This paper makes an effort to propose that prescient examination should be able to anticipate client behavior by utilizing conduct informatics and an investigative approach. The goal of this is to enable prescient examination to obtain a more profound understanding of client behavior,

which would help prescient examination improve business dynamics.

F. D. Pereira et al., [10] A large number of researchers have started eliminating student behavior by cleaning information acquired from online conditions and using it as highlights in artificial intelligence (ML) models. This process was previously described. We have compiled several useful features connected with the student grade by using log information that was acquired from an online adjudicator. These features were then applied to a data collection that included information about 486 students who were enrolled in CS1 courses. We made use of this arrangement of features in enhanced machine learning pipelines, which included a combination of a computerized technique with a developmental computation and hyperparameter-tuning with irregular search. After that, we were able to predict the students' final grades with an accuracy of 75.55 percent by using data from the first fourteen days of the semester. We demonstrate how our pipeline is superior to cutting-edge chip away by comparing the two in various scenarios.

M. A. Salitin et al., [11] Organizations are increasingly turning to more advanced security solutions to protect their data assets. Despite such high stakes, traditional security measures still failed to protect the organizational structure from the most sophisticated attacks. There is a rise in the use of novel preventative approaches to dealing with security issues, such as Client Element Behavior Investigation (UEBA). UEBA is a kind of internet security mechanism that uses artificial intelligence (AI), computations, and factual inspections to determine whether a business is being attacked continuously. The purpose of this article is to assess the usefulness and success of applying conduct examination in protecting an organization against attacks that have not been observed previously, such as zero-day attacks.

Specifically, this evaluation will focus on zero-day attacks.

A. Bouhoute et al., [12] the continuing computerizations of automobiles, along with the advancement of sensor innovations and vehicle specialist gadgets have converted the vehicles into rich wellsprings of data. The analysis of data that is continuously produced by cars has the potential to make significant contributions to the improvement of driving safety and the comfort of drivers. There are still a few significant problems in driving safety that we acknowledge we are not effectively attended to, just as there are a variety of numerical approaches whose application in driving conduct examination is to be researched. This is even though distinct scientific arrangements have emerged in recent times. In this article, we built a philosophy to measure and examine information produced by vehicles, with the center of our focus being on two investigation objectives: 1) programmed confirmation of drivers' behavior adjustment to traffic rules, and 2) perception and correlation of drivers' practices. Both of these objectives are centered around the vehicle. The technique that has been suggested may be broken down into three distinct parts. From the very beginning, the reflection that makes use of mathematical regions is used to reduce the total amount of information that is created [13].

Jain, Jay Kumar et al., [14] combine ML and cyber security to talk about two distinct notions. They also talk about the benefits, problems, and difficulties of combining ML and cyber security. In addition, they explore several attacks and present a thorough analysis of various tactics in two different categories.

Jain, Jay Kumar et al., [15] has discussed about the intrusion detection system, often known as IDS, is a piece of equipment or a piece of software that monitors a network or collection of devices in order to search for indications of possible intrusion. The frequency of cyber assaults has

grown in recent years, and with it, the damage they do to society. The study of cyber security and the avoidance of cyber assaults, such as the use of intrusion detection as a defensive mechanism, is therefore needed. They provide a neural network approach to intrusion detection system threat prediction.

III. PROPOSED STRATEGY

- Load the Amazon Review Dataset from the Kaggle

In this step, the customer review dataset will be downloaded from the Kaggle source. It is a large dataset-providing company. Then load this dataset into the python environment.

- Visualizing the Dataset

Now open the dataset files and view the various data in terms of features like product name, quantity, review, purchasing time, number of visits, add to cart, etc.

- Pre-process the Dataset

Now the data preprocessing step is applied, here data is finalized for processing. Missing data is either removed or replaced from constant one or zero in this step.

- Splitting the Dataset into training and testing

In this step, the final preprocessed dataset is divided into the training and the testing dataset. In machine learning, firstly the machine is trained through a given dataset then it comes in a tested period for the remaining dataset.

- Classification Using Machine Learning Algorithm

Now apply the machine learning technique to find the performance parameters. The existing work applied several techniques and find Naïve Bayes is a better method than others. In the proposed method, we apply the logistic regression method and optimize the better results than other approaches; According to

the researchers, the logistic regression method is good for optimization to enhance the accuracy.

- Performance Metrics (Accuracy, Precision, Recall, F1 - Score)

Now the performance parameters are calculated in terms of precision, recall, f-1 measure, accuracy, etc by using the following formulas-

True Positive (TP): predicted true and event are positive.

True Negative (TN): Predicted true and event are negative.

False Positive (FP): predicted false and event are positive.

False Negative (FN): Predicted false and event are negative.

$$Precision = \frac{|TP|}{|TP| + |FP|}$$

$$Recall = \frac{|TP|}{|TP| + |FN|}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}$$

IV. CONCLUSION

There are various types of customer reviews available on the internet that increasingly affects businesses and customers. Hence, detecting and eliminating such fake reviews from online websites is essential. Machine learning techniques are suitable for predicting and analysis of various problems. This paper reveals several approaches used for customer review performance measures are identified. This topic needs further research in the Big Data approach to reduce the number of features and computational complexity, which helps to improve the detection methods, and also consider other kinds of media such as forums, blogs, etc. Still, it needs to be exposed yet in this regard. The prediction model can identify and review the online data of customer reviews. Therefore need to implement an analysis of the customer review model based on machine learning.

Further, implement the machine learning-based methods and optimize the improved results.

V. REFERENCES

- [1] E. Manohar, P. Jenifer, M. S. Nisha and B. Benita, "A Collective Data Mining Approach to Predict Customer Behaviour," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1310-1316, doi: 10.1109/ICICV50876.2021.9388558.
- [2] V. Shrirame, J. Sabade, H. Soneta and M. Vijayalakshmi, "Consumer Behavior Analytics using Machine Learning Algorithms," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-6, doi: 10.1109/CONECCT50063.2020.9198562.
- [3] B. Lebichot, T. Verhelst, Y. -A. Le Borgne, L. He-Guelton, F. Oblé and G. Bontempi, "Transfer Learning Strategies for Credit Card Fraud Detection," in IEEE Access, vol. 9, pp. 114754-114766, 2021, doi: 10.1109/ACCESS.2021.3104472.
- [4] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou and M. Nijim, "Machine Learning Approaches for EV Charging Behavior: A Review," in IEEE Access, vol. 8, pp. 168980-168993, 2020, doi: 10.1109/ACCESS.2020.3023388.
- [5] J. Edmond Meku Fotso, B. Batchakui, R. Nkambou and G. Okereke, "Algorithms for the Development of Deep Learning Models for Classification and Prediction of Behaviour in MOOCS," 2020 IEEE Learning With MOOCS (LWMOOCS), Antigua Guatemala, Guatemala, 2020, pp. 180-184, doi: 10.1109/LWMOOCS50143.2020.9234363.
- [6] P. A. Savenkov and A. N. Ivutin, "Organizations Data Integrity Providing through Employee Behavioral Analysis Algorithms," 2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2020, pp. 1-3, doi: 10.1109/MECO49872.2020.9134271.
- [7] J. R. Goodall et al., "Situ: Identifying and Explaining Suspicious Behavior in Networks," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 204-214, Jan. 2019, doi: 10.1109/TVCG.2018.2865029.
- [8] D. Damkevala, R. Lunavara, M. Kosamkar and S. Jayachandran, "Behavior Analysis using Serverless Machine Learning," 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2019, pp. 1068-1072.
- [9] Asniar and K. Surendro, "Predictive Analytics for Predicting Customer Behavior," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT), Yogyakarta, Indonesia, 2019, pp. 230-233, doi: 10.1109/ICAIIIT.2019.8834571.
- [10] F. D. Pereira, E. H. T. Oliveira, D. Fernandes and A. Cristea, "Early Performance Prediction for CS1 Course Students using a Combination of Machine Learning and an Evolutionary Algorithm," 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), Macei??, Brazil, 2019, pp. 183-184, doi: 10.1109/ICALT.2019.000066.
- [11] M. A. Salitin and A. H. Zolait, "The role of User Entity Behavior Analytics to detect network attacks in real time," 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakhier, Bahrain, 2018, pp. 1-5, doi: 10.1109/3ICT.2018.8855782.
- [12] A. Bouhoute, R. Oucheikh, K. Boubouh and I. Berrada, "Advanced Driving Behavior Analytics for an Improved Safety Assessment and Driver Fingerprinting," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 6, pp. 2171-2184, June 2019, doi: 10.1109/TITS.2018.2864637.

- [13] Chauhan, Dipti, Jay Kumar Jain, and Sanjay Sharma. "An end-to-end header compression for multihop IPv6 tunnels with varying bandwidth." 2016 Fifth international conference on eco-friendly computing and communication systems (ICECCS). IEEE, 2016.
- [14] Jain, Jay Kumar, Akhilesh A. Wao, and Dipti Chauhan. "A Literature Review on Machine Learning for Cyber Security Issues." (2022).
- [15] Jain, Jay Kumar, and Akhilesh A. Wao. "An Artificial Neural Network Technique for Prediction of Cyber-Attack using Intrusion Detection System." Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN) ISSN: 2799-1172 3.02 (2023): 33-42.
- [16] Jain, Jay Kumar, and Dipti Chauhan. "AN ENERGY-EFFICIENT MODEL FOR INTERNET OF THINGS USING COMPRESSIVE SENSING." Journal of Management Information and Decision Sciences 24 (2021): 1-7.

Cite this article as :

Sidhhant Sharma, Akhilesh A. Wao, "Customer Behavior Analysis in E-Commerce using Machine Learning Approach: A Survey", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 2, pp.163-170, March-April-2023. Available at doi : <https://doi.org/10.32628/CSEIT239028>
Journal URL : <https://ijsrcseit.com/CSEIT239028>