CUSTOMER BEHAVIOUR PREDICTION FOR ONLINE SHOPPING: A REVIEW

Aryan Srivastava , Asad Mahmood , Apoorv Chaudhary, Anikesh Sahai, Pradeep Kumar , Kakoli Banerjee and Ajay Kumar
JSS Academy of Technical Education, Noida

Abstract: **Prediction client behavior in the context of e-commerce is becoming more and more important with the rapid growth in the number of internet transactions brought on by today's shift from visiting physical businesses to online purchasing. By enabling a more individualized shopping experience, it can boost consumer happiness and sales, resulting in improved conversion rates and a competitive edge. Therefore, being able to estimate customer purchases accurately has become a crucial tool for e-commerce businesses. Such data aids in the strategic choice-making during the manufacturing processes as well as the planning of the inventory at the warehouse and point of sale.**

Keywords: **Customer Purchasing Behaviour, Machine Learning Classification, Customer Churn Analysis.**

I. INTRODUCTION

Technology has had an unquestionable and inconceivable impact on the modern period. It has aided humanity in overcoming their dependence on conventional methods. Online purchasing is become a more frequent occurrence in our daily life. Customers are a company's most valuable asset and they provide the groundwork for success and business expansion. Most businesses and startups are spending money on fostering positive customer relationships. To maintain a positive customer relationship, client data is stored. As a result, businesses are able to analyze client purchase patterns and develop the most effective tactics for marketing their goods and services. Online sales are dramatically increasing in the modern day, and clients are connected via digital media platforms. Online transactions are categorized as distance contracts, which means that the trader (service provider, seller) and the consumer (person acting for purposes unrelated to his or her trade, business, or profession, customer), who are not present at the same time and in the same physical location, enter into a contract through an online platform without actually meeting (for example, in a store, open-air market, through a trade agent, etc.) [18]. Nowadays, a lot of e-commerce companies concentrate on luring clients via various social media channels. Typically, a customer will only purchase a thing once and won't do so again. The expense of recruiting new consumers is relatively fairly expensive, and businesses have created a variety of techniques and plans for producing/selling their goods and services. Old clients are the most significant buyers. As a result, the company's

sources of income are reduced. Finding potential repeat customers who will buy the company's products again is vital to solve this issue. Businesses have begun concentrating on devoted clients who are more inclined to purchase their goods and services in order to increase the effectiveness of the marketing.

Consumers are unaware of the factors that affect their purchasing decisions. Economic, demographic, technological, and other factors affect the buyer's decision-making process, which is influenced by the buyer's characteristics and personality. As a result, the buyer's response can be observed, with their purchasing attitudes, preferences, and behavior playing a significant role. Additionally, businesses use user data to transmit, communicate, and disseminate information in order to sell products, get client feedback, and perform customer surveys. To achieve the greatest outcomes, machine learning techniques are deployed appropriately. Data suggests that machine learning algorithms look at customer trends and deliver desired outcomes.

Fig. 1. Factors affecting purchase behaviour [12]

It is difficult to predict a self-initiated purchase since so many variables can influence a buyer's purchasing attitude, and many of them cannot be obtained in an electronic context. Ex: A customer can purchase a TV because an old one is available for purchase online but is inoperable. In an e-commerce setting, it is nearly impossible to anticipate buyer desires. From the

standpoint of the consumer, a customer's purchasing decisions are typically driven by their requirements and preferences. Additionally, a customer's decision to make a purchase is influenced by the popularity of the product [19]. For instance, it can be challenging to pique a customer's interest in a product in which they have no interest.

Fig. 2.Reasons for making purchase decisions [17]

Customers' product preferences also have a big impact on their choice to buy. We investigate relationships between items and make use of them to anticipate client wants. When a customer makes an online purchase, their behavior toward e-shopping refers to their emotional and mental state. The act of buying things online is known as online purchase behavior.

The five main steps of online purchasing behavior are very similar to those of conventional shopping [19]. For instance, when a customer feels the need or desire to purchase a certain good or service, they turn to the internet to make their purchase. They then begin to research all of their options and choose the one that best suits their needs and wants. A number of recommendations are made to buyers prior to making the final purchase, limiting or influencing their choice.

This study's primary goal was to identify the variables that affect consumers' attitudes and behaviors concerning internet buying. This study will also concentrate on how customers develop such attitudes with the use of various models and who actually shop online.

Fig. 3. More than 93% of consumers' shopping decisions are influenced by online reviews. [12]
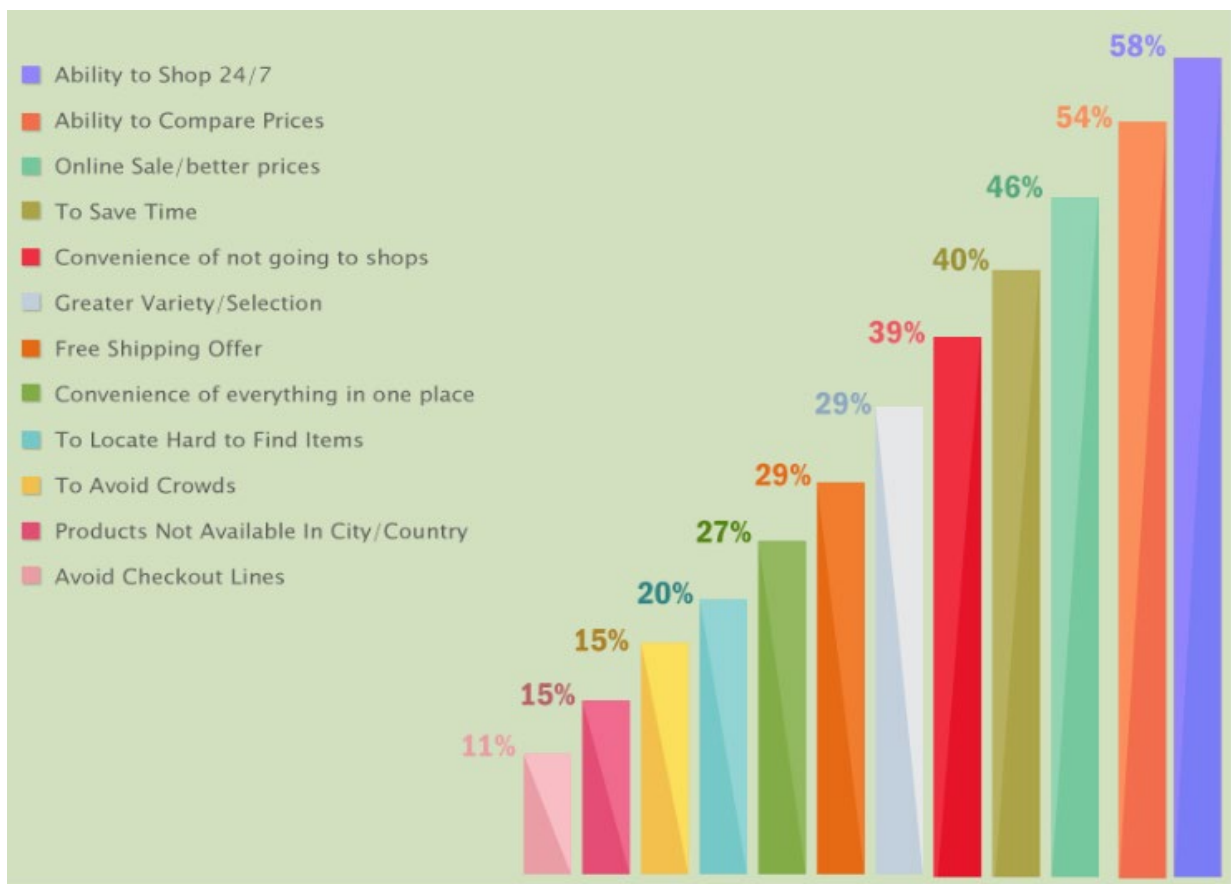

Fig. 4. Reasons for Consumers Shifting to Online Shopping [17]

II.   RELATED WORK

Since firms are now recording consumer data, predicting client purchasing behavior is a significant subject that has been studied by numerous researchers. There are numerous methods that have been used to forecast consumer purchase behavior. The majority of them made use of data mining and machine learning techniques. The following sentences present some current research in this field:

Cardoso [1] reported experiments that used logical discriminant analysis to forecast client purchase behavior. The highest level of accuracy was 90.9%. (making use of a decision tree).

Das [2] created a prediction model to gauge customer acceptance of a company's services based on this hypothesis. Other research have made use of algorithms like naive Bayes, K-nearest Neighbor, and support vector machines. The results show that the naive Bayes algorithm has the highest accuracy (90%).

Additionally, Vafeiadis et al. [3] used ANN, SVM, decision trees, naive Bayes, logistic regression, and their boosted versions to predict customer turnover. This was accomplished using knowledge about clients' prior purchase patterns. With an accuracy of 92%, the boosted SVM was the top classifier overall.

Studies that sought to precisely forecast client attrition were presented by Momin et al. [4]. Different algorithms, including K-nearest neighbor, decision tree, random forest, logistic regression, naive Bayes, and ANN, were contrasted. ANN had the highest accuracy, with 82.83%.

E-commerce is another area that is expanding quickly. In this area, several researchers have worked. In this regard, Kachamas et al[5] .'s proposal for an analytical tool for online retailers to forecast customer behavior. The Naive Bayes approach was mostly used to construct this classification model. More than 86% of accuracy was reached by the model.

The majority of clients who purchase goods on an online marketplace are once-off bargain seekers. Charanasomboon and Viyanon [6] offered a technique for recurring buyer prediction to discover devoted consumers. With the use of the leave one offer out strategy, random forest regression performed better than the other methods.

Do and Trang [7] recently reported experiments that predicted customers' buying decisions. Decision trees, multilayer perception, naive Bayes, radial basis function (RBF), and SVM were all used to assess the sample data. According to the results, the decision tree performed better than the other strategies, with an accuracy of 91.67%.

Additionally, Ullah et al. [8] provided a prediction model to identify churning customers as well as the reasons behind it. The categorization of customers who left the market was attempted using several classification approaches, and a random forest (RF) algorithm proved to be successful with an accuracy rate of 88%.

Alloghani et al. [9] described how machine learning is used to anticipate student progress and learn software engineering. While the random forest produced the best accuracy for the second dataset, the neural network produced the highest accuracy for the first dataset.

Finally, Niu, Li, and Yu [10] used RFs to predict purchase probabilities while primarily focused on search keywords and click locations rather than the actual click stream data. They used data from numerous mouse click events, static and dynamic session data, static customer data, and other information to develop the model. A fair number of records were utilized to train and test the model. According to their investigation, an RF performed 76% more accurately than an LR.

III.   SUMMARY OF RESEARCH LITERATURE

Table 1. Summary of Research Literatures

Paper	Techniques	Accuracy
"An approach based on machine learning techniques for forecasting Vietnamese consumers' purchase behaviour" (2021).[7]	Naïve Bayes Decision Tree SVM	91.67%
"A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector" (2020)[8]	Random Forest	88%

"Application of artificial intelligent in the prediction of consumer behaviour from Facebook posts analysis" (2020)[5]	Naïve Bayes	86%
"Prediction of customer churn using machine learning" (2019)[4]	Naive Bayes Random Forest Logistic Regression ANN	82.83%
"Predictive analytics of e-commerce search behaviour for conversion"(2018).[10]	Logistic Regression Decision Tree	76%
"A comparison of machine learning techniques for customer churn prediction "(2015)[3]	Naïve Bayes SVM ANN Logistic Regression	92%
"Logical discriminant models. In: Quantitative modelling in marketing and management " (2014).[1]		90.9%
"Customer purchase prediction through machine learning" (2018)[11]	RNN	82.0%
"Automated Feature Selection and Churn Prediction using Deep Learning Models" (2017)[15]	CNN LFNN	71.66%
"Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning" [14] (2018)	Multi Layered ANN	80.03%

IV. LIMITATIONS

The related work has already been done in sector of machine learning and deep learning. According to the previous works, Researchers had built a model which takes standalone algorithms for predictions having a maximum precision of 92%.

Very less work had been done that takes makes use of hybrid algorithms or boosted algorithms for predictions.

V. MACHINE LEARNING TECHNIQUES

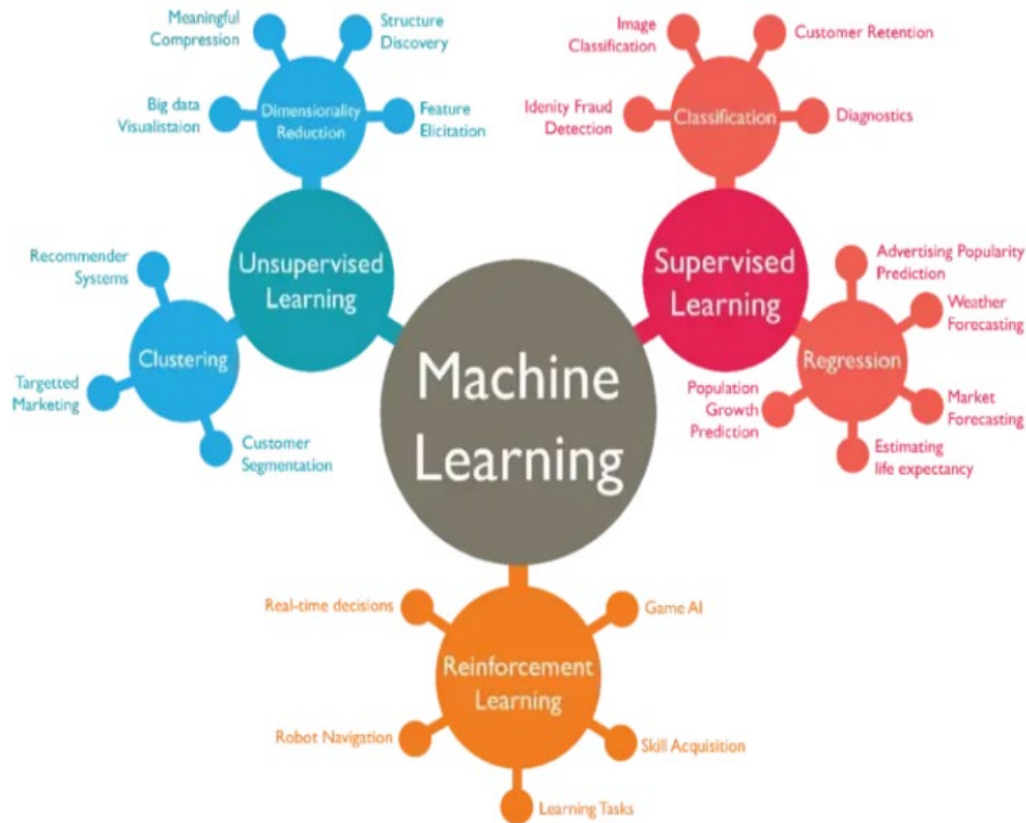Several machine learning techniques will be used on the datasets.

Fig. 5. Machine Learning Techniques [16]

1.1 Decision Tree

A decision tree is a type of supervised machine learning method used to categorize or make predictions based on the answers to a previous set of questions. It is a form of supervised Learning since the model is developed and tested on certain datasets that include the desired classification.

The decision tree might not always provide a clear solution. Instead, it provides options so that the data scientist may decide for themselves. Decision trees simulate human cognition, making it normally easier for data scientists to understand and assess the results..
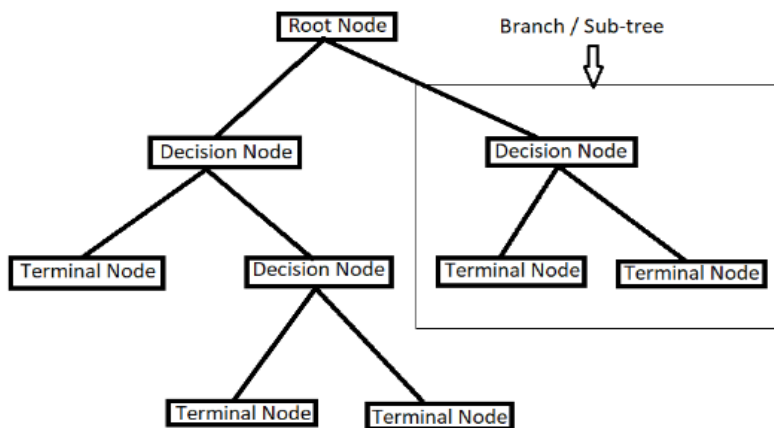

Fig. 6. Decision Tree Algorithm [21]

1.2 Logistic Regression

The category of Supervised Machine Learning algorithms includes the classification algorithm logistic regression. This just indicates that it has its origins in the study of statistics.

This algorithm's primary use in machine learning for logistic regression is to forecast the results of a categorical dependent variable from a set of independent variables. Generally speaking, categorically dependent variables are those whose data is encoded in binary form, either 1 (success/yes) or 0 (failure/no), and are binary in nature (0 or 1).
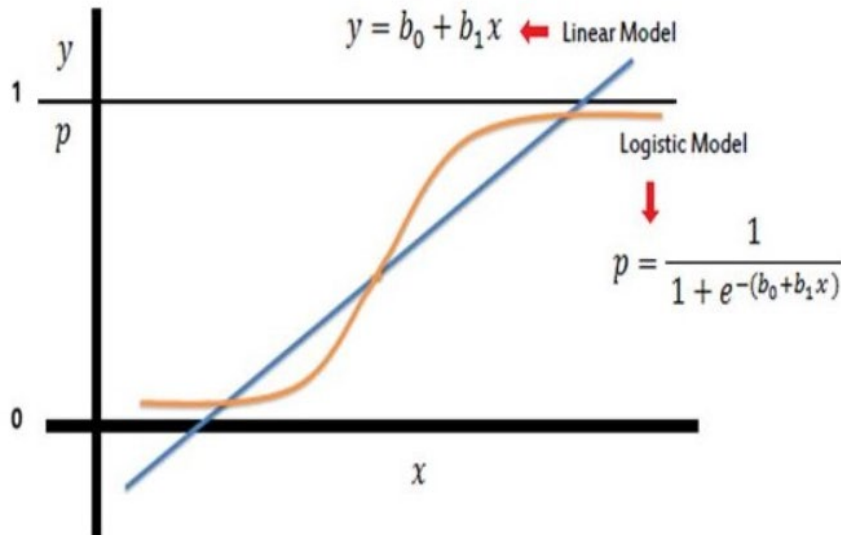

Fig. 7. Logistic Regression [20]

1.3 Naïve Bayes

Naive Bayes is a machine learning approach for classification problems on datasets that is based on the Bayes Theorem. Despite its simplicity, this classifier is fairly effective and may therefore be applied to large and difficult issues. The algorithm has long been a favorite among machine learning techniques because of its prowess in tackling extremely complicated jobs. Sentiment analysis, spam filtering, and other Naive Bayes applications are only a few.

Fig. 8. Naive Bayes Algorithm [22]

1.4 Support Vector Machine

Outlier identification, classification, and regression are all carried out using support vector machines (SVMs), a supervised type of machine learning technique. Because SVMs use some training point subsets in their decision function and are a common and memory-efficient machine learning technique, they are frequently used in the classification of tasks. SVMs are particularly effective in high-dimension fields.
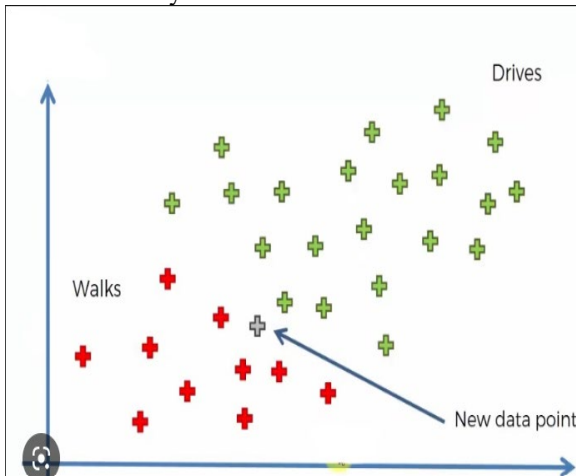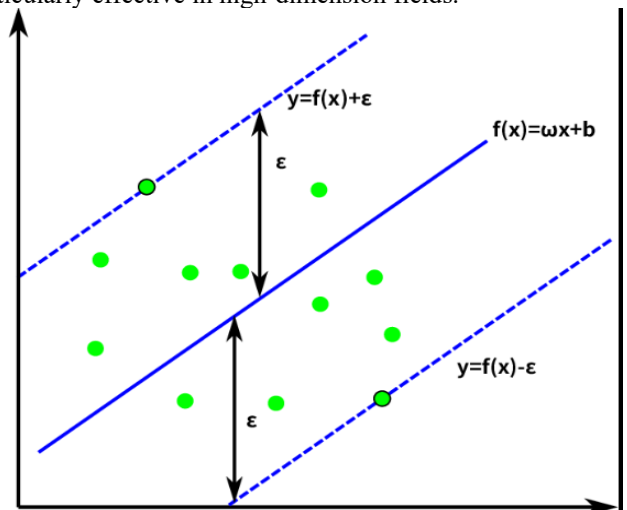

Fig. 9. Support Vector Machine [23]

1.5   Random Forest

Decision trees are used as the individual models for the Random Forest algorithm, an ensemble of models that uses "Bagging" as the ensembling approach. This algorithm is a learning technique that builds numerous decision trees, then uses a random forest to choose the final decision based on the majority of the trees.

It falls within the category of supervised machine learning and can be applied to both classification and regression issues. However, it is typically employed to solve classification issues.
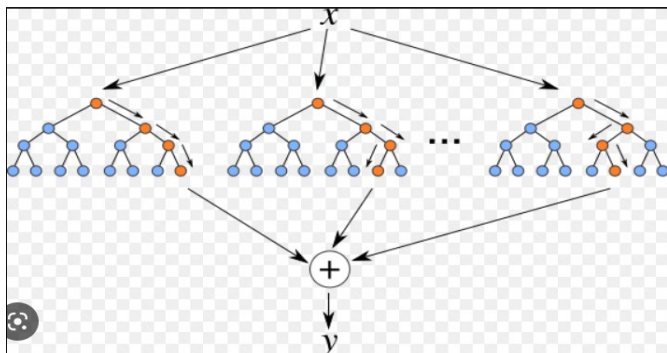

Fig. 10. Random Forest algorithm [24]

1.6   Artificial Neural Network

A weighted and directed graph can be used to describe artificial neural networks (ANNs), where all of the nodes are made up of artificial neurons, and the connections between their inputs and outputs are shown using directed edges with weights. The input signals for an artificial neural network (ANN) are received from outside sources as patterns and images in the form of any vector. The mathematical notations $x(n)$ for every n inputs are then used to identify these inputs [27].
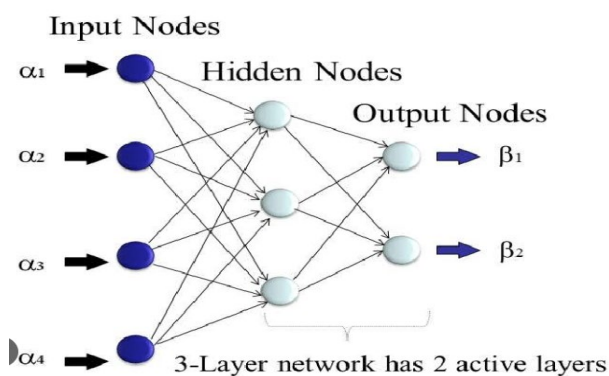

Fig. 11. Artificial neural networks [25]

1.7   K-Nearest Neighbours

To solve classification and regression issues, one can utilize the supervised machine learning method k-nearest neighbors (KNN). Although this machine learning method is simple to use, put into practice, and comprehend, it has a big downside in that it can occasionally become significantly slower as the vol-

ume of data we are examining rises. KNN computes the distances between a query and each example in the data while classifying data. It then chooses the K instances that are the closest to the query, averages the labels, or chooses the label with the highest frequency (in the case of regression).
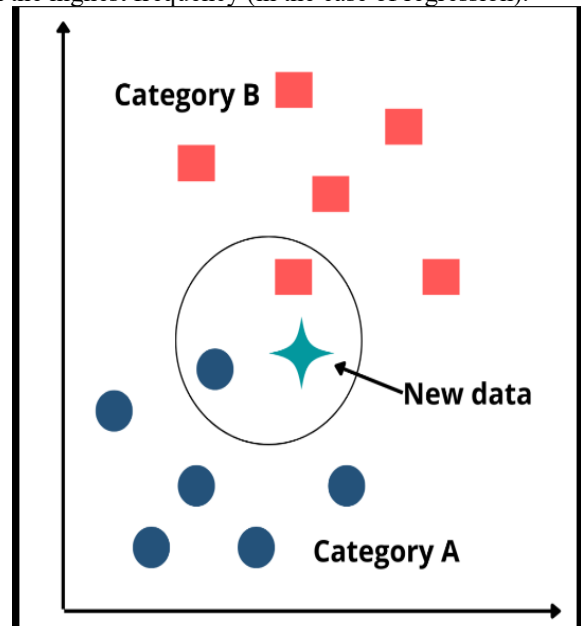

Fig. 12. K Nearest Neighbours [25]

VI.   CONCLUSION

Through a thorough understanding of machine learning algorithms, this work will aid in the development of an understanding of customer purchasing behavior analysis and provide procedures for solving machine learning problems. We will give a comparative study comparing several supervised machine learning methods, such as logistic regression, decision tree, KNN, naive Bayes, SVM, and random forest, to handle the problem of forecasting customer purchase behavior.

According to the study's findings, managers and marketers should always remember to create strategies based on e customers' attitudes toward the repurchasing related to their convenience with the concerns about ease of use and product information - trust and security - in order to attract and keep their customers in this online environment. To experience the repurchase, a further elevation and its ramifications for the managers are crucial.

As we come to a conclusion, it has become clear that a variety of factors, including customers' past experiences with particular brands and their perceptions of the significance of convenience, trust, and security when making purchases, influence their intention to make additional purchases of products in this online market. This research reveals the enormous possibilities of analyzing consumer purchasing behavior through the personalities and traits of the persons involved. Future research will analyze the outcomes of this strategy using various categorization techniques. In this study, we discovered that customer

reviews play a significant influence in online word-of-mouth and are strongly correlated with consumer purchases of any product. Online reviews, which are an essential component of the purchasing experience, may be just what it takes to convince a consumer who wants to buy to give them some thought. The consumer's perception of risk may be lessened by certain pertinent information provided about a particular product through the reviews, which will affect the customer's purchasing intention and behavior, according to observation. These testimonials and comments are helpful for prospective consumers in deciding whether to make a purchase, but they also assist the company in raising the caliber of its goods and services. Online reviews of items, brand, availability, price, and the necessity of the product are only a few of the numerous variables that have an influence on how customers behave while making purchases. The examination of the preliminary data, which comprised these individuals from various locations and age categories, served as the primary foundation for this study. In order to conduct the study in sample locations and gather pertinent data for research purposes, a questionnaire was devised.

VII.   FUTURE SCOPE

In order to achieve more accuracy, this implementation approach may be used in the future on bigger datasets that include enormous amounts of data—hundreds or millions of records. With machine learning hybrid algorithms SvmAda, AdaBoost, XgBoost, or KnnSgd, we may assess how well our model performs.

VIII.   REFERENCES

[1].   Cardoso MGMS, "Logical discriminant models. In: Quantitative modelling in marketing and management", (2012).

[2].   Das TK, "A customer classification prediction model based on machine learning techniques. 2015 (iCATccT)", (2015), IEEE

[3].   Chatzisavvas K, Sarigiannidis G, Diamantaras K, Vafeiadis T, "A comparison of machine learning techniques for customer churn prediction", (2015).

[4].   Momin S, Bohra T, Raut P, "Prediction of customer churn using machine learning", (2019), pp 203–212

[5].   Sinthupinyo S, Kachamas P, Akkaradamrongrat S, Chandrachai A, "Application of artificial intelligent in the prediction of consumer behaviour from Facebook posts analysis", (2019).

[6].   Viyanon W, Charanasomboon T, "A comparative study of repeat buyer prediction", (2019).

[7].   Trang TV, Do QH, "An approach based on machine learning techniques for forecasting Vietnamese consumers' purchase behaviour", (2020).

[8].   Imran M, Malik AK, Imran M, Islam SU, Kim SW, Ullah I, Raza B , "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector", (2019).

[9].   Baker T, Hussain A, Mustafna J, Aljaaf AJ, Alloghani M, Al-Jumeily D, "Applications of machine learning techniques for software engineering learning and early prediction of students' performance", (2018).

[10].   Yu, X., Li, &Niu, X. "Predictive analytics of e-commerce search behaviour for conversion", (2017)..

[11].   Hannah Sophia Seippel, "Customer purchase prediction through machine learning. Master's thesis", (2018), University of Twente

[12].   Rishita Goyal, Megha Grover, "A study on consumer's buying behaviour based on customers' online reviews", (2020).

[13].   Moaiad Ahmad Khder, Suresh Subramanian, and Samah Wael Fujo, "Predictive Analytics of E-Commerce Search Behaviour for Conversion", (2021).

[14].   Das, A. Gaikwad, S. Dhage, S. Agrawal, "Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning", (2018).

[15].   Iyakutti, K., Umayaparvathi, V., "Automated Feature Selection and Churn", (2017).

[16].   Audrey Lorberfeld, "Machine Learning Techniques in Layman's Terms", (2019) .

[17].   Chandra Gogineni, "Top 5 Valuable Insights on Online   Consumer Buying Behaviour", (2022).

[18].   Gurudatt Kakkar, "Online Shopping : The Growth", (2017).

[19].   JiangtaoQiu, "A predictive model for customer purchase behaviour in e-commerce context", (2014).

[20].   Hemant Kumar Gianey, "Comprehensive review on Supervised Machine Learning Algorithms", (2017)

[21].   Amrutha K, "Decision tree Machine Learning Algorithm", (2021).

[22].   Surabhi S, "A Guide to Naïve Bayes Algorithm", (2021).

[23].   Geronx, "An Example of SVM regression in a simple 2D , linear and hard-margin problem", (2021).

[24].   Chaya, "Random Forest Regression", (2020).

[25].   Ferry Wahyu Wibowo, "An Analysis of FGPA Hardware Platform Based Artificial Neural Network", (2019).

[26].   Keerthana Buvaneshwaran, "K nearest neighbours | KNN: One of the earliest ML Algorithm", (2021).

[27].   Ravindra Savaram, "What is Artificial Neural Network and how it Works, (2022).