

# Prediction of Consumer Purchase Intension

**Prof. Vrushali Paithankar**

Assistant Professor

Department of Computer Engg.

Smt. K. N. C. O. E. Pune, India

vrushali.paithankar@gmail.com

**Shruti Pardeshi**

Student

Department of Computer Engg.

Smt. K. N. C. O. E. Pune, India

shrutipardeshi.skncoe.comp@gmail.com

**Esha Tate**

Student

Department of Computer Engg.

Smt. K. N. C. O. E. Pune, India

eshatate.skncoe.comp@gmail.com

**Sakshi Darwatkar**

Student

Department of Computer Engg.

Smt. K. N. C. O. E. Pune, India

sakshidarwatkar.skncoe.comp@gmail.com

**Purva Wagh**

Student

Department of Computer Engg.

Smt. K. N. C. O. E. Pune, India

purvawagh.skncoe.comp@gmail.com

**Abstract:** The change in recent years from physical store visits to online purchasing has made it more crucial than ever to predict client behavior in the context of e-commerce. It can increase consumer satisfaction and sales by offering a more customized purchasing experience, improving conversion rates and giving businesses a competitive edge. Customer data can be added to and used to build models for predicting consumer behavior. In this study, a big German clothing shop uses machine learning models to forecast a purchase, which is an important use case. This study goes beyond simply evaluating the performance of the models on sequential and static customer data by conducting a descriptive data analysis and individually training the models on the various datasets. Total of three different algorithms.

**Keywords:** SVM, Customer Purchase Intention, Dataset, Machine Learning.

## I. INTRODUCTION

Like Twitter and Facebook, social networking is a sort of real-time microblogging where users constantly receive, send, and share information. Consumer product-related interactions are more prevalent on the microblogging application Twitter. Tweets about shopping websites, as well as tweets about other Twitter pages, businesses, mobile phone companies, clothing brands, and live events like sporting events and elections, can be retrieved.

The service provider will be able to learn more about how clients feel about their offerings thanks to these outcomes. Twitter has grown to be a well-liked online social networking tool for exchanging up-to-the-minute news about current affairs.

Several studies are currently being done to determine the best way to use the vast amount of data that users on Twitter are currently posting. The research covers topics including finding communities in social networks, summarizing information, examining tweets, and more. A social media site that allows users to write tweets with 280 characters. Tweets only have a certain number of characters, making sentiment analysis simple. Every day, 550 million tweets are published on Twitter. Twitter also represents all age group people and a fair representation of gender. Therefore, the sentiment analysis of twitter data becomes somewhat general sentiments of society [4]. In this paper, we use different machine learning algorithms to analyze the

purchase intention of the user. We have used algorithms like Support Vector Machine and LSTM. We will compare the accuracy and precision of all these methods to find out which one works the best.

Unstructured data can be extracted in bulk from webpages using the automated technique known as web scraping.

Web scraping assists in gathering this unstructured data and storing it in a structured format to help avert this issue. As a result, it is now vital to conduct research in the field of customer price and rating of product evaluation and forecast. The purpose is to examine a dataset using machine learning-based methods for product rating predicting to achieve the highest level of accuracy.

## II. RELATED WORK

Proposed [1] Review uses the updates life structure response (SOR) speculation, which exhibits that particular normal enhancements influence the customers' inward state or animal and fundamentally merges their approach to acting, in this manner. There study includes charity as the improvement, supporting neighborhood creators, straightforwardness, satisfaction with checking, and hankering for naming as the customers' internal state (living being), and purchase assumptions and brand love as the response. Cross-sectional data were assembled from 2045 neighborhood food customers related with Facebook-based REKO (fair use) bundles in Finland.

In [2] the point of the paper was to look at chain steadiness effects of clients' evident worth of endurance programs in staple retailing. The essential finding was that clients' obvious worth of a steadiness program influences chain dependability, and the control influences are chain subordinate. Bosses should therefore consider satisfaction creation, picture building, and commitment program regard creation as equivalent cycles.

This paper [3] study intended to conclude the connection between individual's web based participation and e-cigarette purchase objective from Facebook vape social affairs. A cross-sectional survey was coordinated among 214 respondents using the Facebook stage from September to December 2019. Data were accumulated through an independent survey.

In [4] The results of 1726 datasets from two web-based business stages suggest that client responsibility is generally associated with followership and purchase assumption in live-streaming automated promoting. Whiles cost is a basic middle person, its effects become immaterial on their purchase assumptions once clients become allies. The results highlight the constructive outcomes of social parts, including likes, talks, visits, and transparency time in well disposed exchange towards contingent (purchase) and non-esteem based (followership) benefits.

In [5] framework is making an informational collection from two unique sources, for instance, using beforehand existing data and the data eliminated from Twitter and the data considered was film assessments. For removing data from Twitter we utilized the normal language dealing with thoughts using python. At the point when the complete data made then the data provided for the BERT model to perceive the various components which expect a urgent part in equal gathering as per the assessments as lucky or sad

Proposes [6] client point assumption model was been comprehensively used for the coupon piece, notice and idea on Taobao stage, which staggeringly further foster the client experience and shopping capability, and benefit the gross item volume (GMV) progression also

Proposed [7] Model Considering assessment, they manufacture two sorts of pointers: (1) a marker for strange gatherings that can unequivocally expect purchase reason in obscure gatherings, beating a creation arranged marker by over 17.54 percent F1 and (2) a pointer for perceived clients that includes meeting data along with client history and achieves a F1 of 96.20 percent on held-out data assembled from a genuine retail stage.

In the proposed framework [8], focus on gave an observational preliminary of this fresher model by separating every one of the four of its parts with shopper steadfastness and purchase assumption. Using fundamental condition exhibiting to take apart 508 impending area first-time homebuyers, this study surveys the occupation of the pieces of Displaying 4.0 in supporting customer devotion and influencing purchase objectives

They [9] The assessment gives an extensive manual for responsiveness examination of model limits as to execution in assumption for thing assessments with cost nuances by tracking down precision computation.

In this paper [10] study looks at simulated intelligence models to predict a purchase, which is a huge use case as applied by a colossal German dress retailer. Then, to differentiating models this concentrate further gives information into the presentation qualifications of the models on continuous clickstream and the static client data, by coordinating an obvious data assessment and freely setting up the models on the different datasets. The results exhibit that a Sporadic Forest computation is the best for the assumption task, showing the best show results, reasonable latency, offering fathomability and a high strength.

## III. SYSTEM ARCHITECTURE

- **Data pre-processing:** It is a data mining approach that involves converting raw data into an understandable format. Filling in missing numbers, smearing noisy data, and resolving data discrepancies are all examples of

data cleansing techniques. Because there are some missing values in the dataset, decimal values are converted to appropriate float values.

- **Data splitting:** The new dataset is divided into two parts: the training and testing sets. For distribution, an 80-20 split is used. The training set for the model is made up of 80% of the dataset. The remaining 20% is converted into the Test Set, which is used to evaluate and test the model's correctness. Never train with the testing set because this may cause the mode to become overfit.

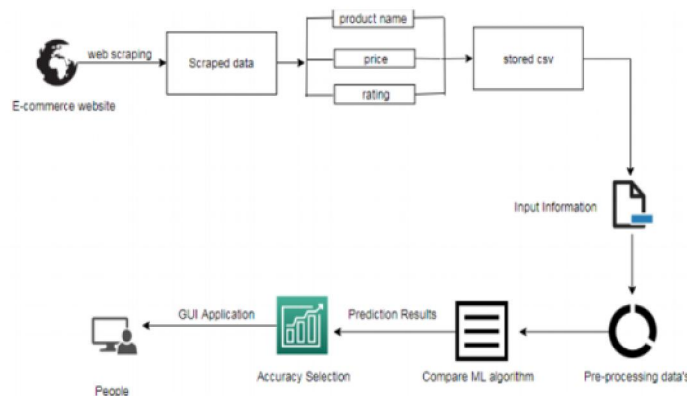


Fig: System Architecture

- **Feature Selection:** The properties of the data used to train machine learning models influence the model's performance significantly. Model performance may be harmed by features that are irrelevant or only partially relevant.
- **Classification:** The model is trained by adjusting the training set for the classifier model. Following testing, the classifier model assigns a good or bad rating to the air quality. The categorizations closely correspond to the testing set.

## IV. ALGORITHM

### 4.1 Support Vector Machine

A supervised machine learning approach called a support vector machine (SVM) can be applied to both classification and regression problems. SVM are frequently employed in classification issues.

Finding the optimum hyperplane to split a dataset into two classes is the foundation of SVM. Support vectors are the data points in a data set that, if removed, would change the location of the dividing hyperplane and are closest to the hyperplane.

These can therefore be regarded as the crucial components of a data set. The margin is the separation between the hyperplane and the closest data point in either collection. The goal is to select a hyperplane with the largest margin between the hyperplane and any point in the training set, increasing the likelihood that new data will be properly classified.

- Import the dataset
- Explore the data to figure out what they look like
- Pre-process the data
- Split the data into attributes and labels
- Divide the data into training and testing sets
- Train the SVM algorithm
- Make some predictions
- Evaluate the results of the algorithm

### 4.2 Random Forest

Step-1: Select random K data points from the training set

Copyright to IJARSCT

DOI: 10.48175/IJARSCT-9136

[www.ijarsct.co.in](http://www.ijarsct.co.in)

- Step-2: Build the decision trees associated with the selected data points (Subsets)  
 Step-3: Choose the number N for decision trees that you want to build  
 Step-4: Repeat Step 1 & 2

## V. MATHEMATICAL MODEL

Let S be as system which can identify the customer purchase intention

$S = \{In, P, Op\}$

Identify Input In as

$In = \{Q\}$

Where, Q = Input Data from User

Identify Process P as

$P = \{CB, C, PR\}$

Where, CB = Pre-processing

C = Feature Extraction

PR = Classification

Identify Output Op as

$Op = \{UB\}$

Where, UB = Output

## VI. CONCLUSION

In the e-commerce sector, the consumer's role in the purchase of such a product is critical. There is a wealth of information available on the internet that can be used for research. In order to perform the fundamental operations of machine learning algorithms, information is gathered from the internet. After data preprocessing, data cleaning and analysis are performed so that the algorithms can operate on the supplied data to generate results. The proposed system predicts client intention and expresses it in the output after prediction. Because merchants are the intended audience, the results must be presented simply so that everyone can understand them. To make the results more understandable, we use a user interface.

## REFERENCES

- [1]. Sushant Kumar , Mikko Murphy “What drives brand love and purchase intentions toward the local food distribution system? A study of social media-based REKO (fair consumption) groups”, 2021
- [2]. Erik Nettet , Ola Bergem, “Building chain loyalty in grocery retailing by means of loyalty programs – A study of ‘the Norwegian case’”, Journal of Retailing and Consumer Services 2021
- [3]. Nusrat Parvin ; Sayaka Zaman ; Samia Amin , ” E-cigarette Purchase Intention through Facebook Social Media: A Cross-Sectional Study” , Science Publication 2021
- [4]. Price Clement Adoo , ” Customer engagement and purchase intention in live-streaming digital marketing platforms”, Research Gate 2021
- [5]. Aanchal Badgaiyya and Prachi Shankarpale , “An Application of Sentiment Analysis Based on Hybrid Database of Movie Ratings ” , International Research Journal of Engineering and Technology 2021
- [6]. Jingxing Jiang, Zhubin Wang, Fei Fang, BinqiangZhao , “TPG-DNN: A Method for User Intent Prediction Based on Total Probability Formula and GRU Loss with Multi-task Learning ” , IEEE Conference 2020
- [7]. Mariya Hendriksen ,Pim Nauts , “Analyzing and Predicting Purchase Intent in E-commerce: Anonymous vs. Identified Customers ” , IEEE Conference 2020
- [8]. Ganesh Dash , Justin Paul , “Marketing-to-Millennials: Marketing 4.0, customer satisfaction and purchase intention ” , Journal of Business Research , Elsevier 2020
- [9]. V kumar and S kesharwani ,“Online Customer Satisfaction and Loyalty in Apparel Purchase: A Review of Extant Literature”,2020
- [10]. HS Seippel ,“ Customer purchase prediction through machine learning”, Electrical Engineering, Mathematics & Computer Science, 2018