



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

**Volume 10, Issue 7, July 2022**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.165**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com



# Customer Purchase Intention Prediction and Recommender System

**Sahil Kumar Mittal, Satvik Katoch, Shikharashree G, Shruti Kashyap, K N Prashanth Kumar**

Student, Department of Computer Science and Engineering, Bangalore Institute of Technology, Bangalore, India

Assistant Professor, Department of Computer Science and Engineering, Bangalore Institute of Technology,  
Bangalore, India

**ABSTRACT:** This paper proposes a methodology in order to estimate the customer purchase intention prediction using the SVM algorithm. We have also proposed a recommender system which takes in the customer's requirements and displays the appropriate products. We provide an option of giving reviews of a particular product which in turn is fed into a review classifier which uses Logistic Regression. This would classify the review as positive or negative so that the user can make better decision about his buys. We use flask to let the manufacturer or any other person with a product to enter the specifications of their product and our ML algorithm will give the customer purchase intention prediction in terms of a percentage score. We use Web Scraped data from "Flipkart" to train our SVM model. For the purpose of our experiment, we used data only from a single website.

**KEYWORDS:** Customer; Machine Learning

## I. INTRODUCTION

In different nations, e-commerce has been a huge success and has given tremendous economic and social benefits. Due to the Covid-19 pandemic in 2020, E-Commerce has been thrust to the forefront of retail, resulting in a structural change in demand toward digital commerce that is expected to last for years. Customers explore online the items they need and buy them through online transactions. All these operations can happen at one place which saves money and time. But at the same time sellers also need to know about the trends or seasonality factors that are going on in the market, the requirements of the customers so that they can provide the best products to its customers and grow their businesses. As a result, understanding the ethics and motivations of online consumers has become incredibly influential in marketing, in order to enhance consumer service and, as a result, sales. Lots of products with the same qualities and slight variations result in confusion and the customer ends bare handed. Hence we have used machine learning algorithms to predict the intentions of online customers. Machine learning predicts a greater amount of information based on previous experience. ML is a unique AI that allows you to programme your laptop specifically. The fundamentals of machine learning, using PHP, run simple machine learning algorithms. Machine learning focuses on the occurrence of changing laptop programmes after dealing with new information. In the employment and prediction process, special algorithms are used. It provides coaching rules with an algorithm, and the algorithm uses this coaching knowledge to make predictions about different test knowledge. Machine learning is usually classified into three categories. This includes supervised, unsupervised and reinforcement learning.

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is

just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

Logistic regression is basically a supervised classification algorithm. In a classification problem, the target variable(or output),y, can take only discrete values for a given set of features(or inputs), X.

Contrary to popular belief, logistic regression IS a regression model. The model builds a regression model to predict the probability that a given data entry belongs to the category numbered as “1”.

Just like Linear regression assumes that the data follows a linear function, Logistic regression models the data using the sigmoid function.

$$g(z) = \frac{1}{1+e^{-z}}$$

To create Python models that produce useful statistics, scientists use a variety of machine learning algorithms. We use SVM and LR to provide the solution to existing problem of purchase intention prediction.

## II. RELATED WORK

This section presents the existing methods and relevant approaches which are surveyed as follows.

Danah Algawiaz proposed a model which identifies eight factors that can influence a user’s purchase intention, and categorizes them as user and population behaviors. These features will be fed to the AdaBoost algorithm to predict the user’s purchase intention.

Ahmed M. Elmogy proposed a TF-IDF model to extract the features of the contents in two languages models; mainly bi-gram and tri-gram. The TF\*IDF algorithm is used to weigh a keyword in any content and assign importance to that keyword based on the number of times it appears in the document. 70% of the dataset is used for training while 30% is used for testing. They are tested for different classifiers.

Shashank kumar Chauhan proposed a model in which Sentiment analysis is taken into account for spam detection instead of factoring only the rating score. Used a self-developed dictionary to categorize various words into producing a weightage score. The above is called creating a sentiment score (how good or bad a review is) Understanding negating statements is important to correctly measure a review. e.g.: "The built-in speaker also has its uses but so far nothing revolutionary." The reviews wherein the difference of the sentiment score and the rating is greater than 0.5, is considered to be spam.

## III. SYSTEM ANALYSIS AND DESIGN

### A. Existing system and their drawbacks:

The drawbacks of existing systems are:

- Current systems that predict the intention of purchase of products by online customers already exist but it depends on customers’ activity while they use their software, our project uses the data of customer’s history, item features and ratings to predict the intention of the customers.
- Recommendation websites that exist only tells us about the products and its features whereas our project allows the customer to input the features of the products they prefer and according to those features and its ratings our project will recommend the items to them so that they can get the best product they want.

### B. Description of the Proposed Algorithm:

Aim of the proposed algorithm is to predict the customer purchase intention based on how the previous products in the market have performed which is shown in the ratings. (Extracted through web-scraping flipkart)



The methodology involves the following steps:

- 1) Implementation of this customer purchase intention prediction needs a correct website/app to be chosen through which the required data needs to be scraped. Here, we have considered Flipkart.
- 2) After scraping any data, if Machine Learning algorithms has to be run, pre-processing steps is a must. There are a lot of pre-processing ways that could be followed and later normalize it.
- 3) Once the data is ready to be trained upon, the normalized data was trained upon by the SVM algorithm because the algorithm is as good as the data it is trained on. This algorithm is well observed to be a good classification and regression algorithm.
- 4) Using logistic regression to classify the reviews provided by the users on the website. These reviews are classified into categories which influences the prediction.
- 5) We later take in the specifications provided by the manufacturer to predict the customer purchase intention for which the SVM algorithm was trained. So this would tell us if that product which the manufacturer is intending to buy would be a successful product in the market or not.
- 6) We have also developed a website of our own with mobile phones, where users could enter their preferred features needed in their phone and the similar products will be recommended to the user. The user could also give reviews over that product which again would be categorized as a positive or negative review using LR and also displayed to the user.

#### IV. PSEUDO CODE

##### For Prediction of Customer Intent:

Input from manufacturer: {I ∈ string array}

Use CSV file extracted from flipkart {file ∈ .csv file}

Training and pre-processing:

```
data['Display(incm)'].fillna(data['Display(incm)'].mean(),inplace=True)
```

```
data['Battery(inmAH)'].fillna(data['Battery(inmH)'].mean(),inplace=True)
```

For i,do:

- Run svm algorithm on i
- Display output to UI

##### For Recommender System:

Step1: Start

Step2: Get customer requirements i.e., Product features and provide it to recommender system.

Step3: Match with products already in database and recommend it to user.

Step4: Classifier takes input of reviews and uses Logistic Regression to classify review as positive or negative

#### V. SIMULATION RESULTS

The images contained in the simulation simulation describes how the data was visualized according to the web scraped data from flipkart. The first image contains no. of mobiles vs ram which give us clear indication of the no. of mobiles spread across the RAM data. Similarly, the second image contains the no, of mobiles vs rom.

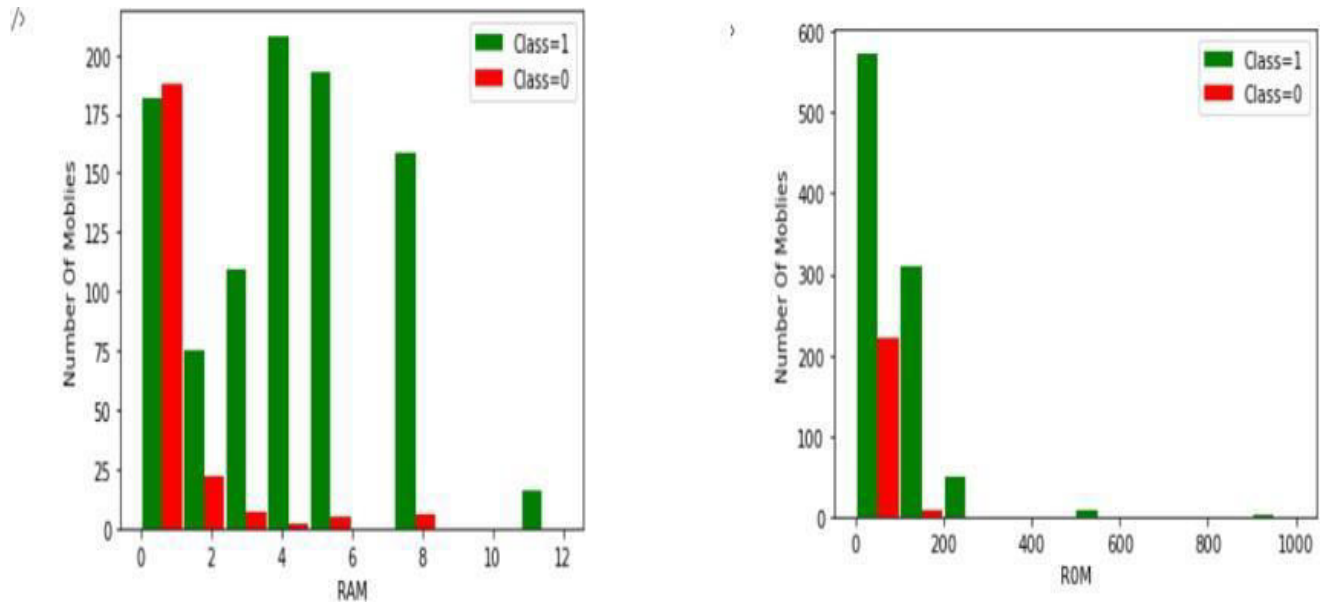


Fig 1: Data Visualization

We've also added the cross-validation test-results of accuracy of the SVM algorithm that was used to train and test the data on the manufacturer's input.

Classification report of Support Vector Machines Results:

	precision	recall	f1-score	support
0	0.82	0.93	0.87	188
1	0.92	0.80	0.85	189
accuracy			0.86	377
macro avg	0.87	0.86	0.86	377
weighted avg	0.87	0.86	0.86	377

Confusion Matrix result of Support Vector Machines is:

```
[[174 14]
 [ 38 151]]
```

Cross validation test results of accuracy:

```
[0.75132275 0.75132275 0.7287234 0.76595745 0.70744681 0.93617021
 0.85638298 0.92553191 0.92553191 0.92553191]
```

Accuracy result of Support Vector Machines is: 82.73922098390184



The figure below is a result of the LR-Classifier which classifies the review given by the user into positive or negative.

CLASSIFICATION OF REVIEWS	
S.No.	Positive Reviews
0	GOOD PRODUCT
S.No.	Negative Reviews
1	BAD PRODUCT

Fig 2: Classification of Reviews

## VI. CONCLUSION AND FUTURE WORK

In this project, we propose an architecture which will allow the consumers to be able to correctly judge a product's popularity in the market. We've used a recommender system for that. For the manufacturers we've web-scraped data from flipkart and implemented the SVM algorithm in order to predict the phone's reachability in the market. The SVM algorithm was trained on the web-scraped data on the input provided by the manufacturer.

Future work also includes considering reviews instead of only ratings and also users personal history.

## REFERENCES

- [1]. D. Algawiaz, G. Dobbie and S. Alam, "Predicting a User 's Purchase Intention Using AdaBoost," 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2019, pp. 324-328, doi: 10.1109/ISKE47853.2019.9170316.
- [2]. S. Mootha, S. Sridhar and M. S. K. Devi, "A Stacking Ensemble of Multi Layer Perceptrons to Predict Online Shoppers' Purchasing Intention," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2020, pp. 721-726, doi: 10.1109/ISRITI51436.2020.9315447.
- [3]. Prediction of Customer Purchase Intention Using Linear Support Vector Machine in Digital Marketing to cite this article: G. Saranya et al 2020 J. Phys.: Conf. Ser. 1712 012024.
- [4]. K Radha, V Karthik , K Manish "Predicting the Consumer's Product Purchase Intention Using Regression Analysis at Attribute Level" International Journal of Computer Trends and Technology 67.9 (2019):11-20.



- [5]. Ahmed M. Elmogy, Usman Tariq, Ammar Mohammed and Atef Ibrahim, "Fake Reviews Detection using Supervised Machine Learning" International Journal of Advanced Computer Science and Applications(IJACSA), 12(1), 2021. <http://dx.doi.org/10.14569/IJACSA.2021.0120169>
- [6]. S. K. Chauhan, A. Goel, P. Goel, A. Chauhan and M. K. Gurve, "Research on product review analysis and spam review detection," 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), 2017, pp. 390-393, doi: 10.1109/SPIN.2017.8049980.
- [7]. M. Yasen and S. Tedmori, "Movies Reviews Sentiment Analysis and Classification," 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019, pp. 860-865, doi: 10.1109/JEEIT.2019.8717422.
- [8]. N. Kurmashov, K. Latuta and A. Nussipbekov, "Online book recommendation system," 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO), 2015, pp. 14, doi: 10.1109/ICECCO.2015.7416895.
- [9]. M. O. Pratama, R. Meiyanti, H. Noprisson, A. Ramadhan and A. N. Hidayanto, "Influencing factors of consumer purchase intention based on social commerce paradigm," 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2017, pp. 73- 80, doi: 10.1109/ICACSIS.2017.8355015.



INNO  SPACE  
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**<sup>®</sup>  
**CROSS** **ref**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  [ijircce@gmail.com](mailto:ijircce@gmail.com)



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details