A Machine Learning Approach to Predict the Consumer Purchasing Behavior on E-commerce Website During Covid-19 Pandemic

Muhtasim[1], Redoanul Haque[2], Md. Sumon Ali[3] , Md. Showrov Hossen[4]

[1] Lecturer, Dept. of CSE, Bangladesh Army University of Engineering & Technology (BAUET), Natore, Bangladesh
[2] Lecturer, Dept. of CSE, Bangladesh Army University of Engineering & Technology (BAUET), Natore, Bangladesh
[3] Lecturer, Dept. of CSE, Bangladesh Army University of Engineering & Technology (BAUET), Natore, Bangladesh
[4] Lecturer, Dept. of CSE, City University, Dhaka, Bangladesh

---

Abstract - *COVID-19 is the name of a recent terror around the whole world. This pandemic situation rapidly affecting every stages of a human life. To decrease the outbreak of COVID-19 and keeping everyone safe from the bad impact of this situation some precautionary steps like lockdown and curfew are taken places all over. In this current pandemic situation e-commerce sites are playing an important role to fulfil daily life needs and now a days the consumers are getting dependent on e-commerce services. Variety of consumers purchasing behavior helps the owners of e-commerce sites to improve business and serving their consumers properly. In this research work we will try to determine the consumers purchasing behavior using several machine learning algorithms. We have used here Linear Regression, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest and Naïve Bayes algorithms to predict the consumers purchasing behavior on e-commerce website.*

Key Words: (COVID-19, Pandemic, E-Commerce, Purchasing Behaviour.

1. INTRODUCTION

Now a days in the modern era of technology people rapidly getting dependent on purchasing product from various e-commerce website. But in this pandemic situation of COVID-19 this dependency has increased. Due to safety consumers are preferring to purchase products from online rather than purchase from market in this current situation. On the other hand, e-commerce sites are also developing and growing at a good number day by day. During pandemic humanity faced severe problems due to staying at home. So, they had to purchase things online to survive. Clothes, foods and all other necessary goods had to order online. As dependency is increasing, the owners of these e-commerce sites should be more alert about their services and products. This improvement can be made more precisely if the holders of this sites have an average idea about the consumers purchasing behaviour. This research work will help them to have an overview knowledge of consumer needs in this current pandemic situation so that they can improve their service according to this. Our proposed system can easily predict the demand of a product before adding the product on e-commerce website.

1.1 Motivation

Analyzing the importance of e- commerce websites in the recent situation of COVID-19 pandemic and some concerned research paper, we have motivated to work on this topic and that is prediction of the consumer purchasing behaviour on e-commerce website during Covid-19 pandemic. Some significant issues noticed by us not analyzed yet:

-Determination of the consumers purchasing behavior according to the recent situation of COVID-19 pandemic from e-commerce websites.

-Comparison of the consumers purchasing behavior during pandemic and before pandemic situation.

2. Literature Review

Customer purchasing behavior is an important and interesting issue in the field of digital marketing or e-commerce service. Generating this research work we have reviewed here some related papers.

In paper [1] consumer purchasing behavior of an e-commerce website is estimated by tracking the usage and sentiments attached to their products. Some data-driven marketing tools are used here such as data visualization, natural language processing and machine learning models that help in understanding the demographics of an organization.

According to the paper [2] a predictive framework, Customer Purchase Prediction model (COREL) has created for determining customer purchase behavior in the e-commerce context. This model works in two stages, when a product purchased by a particular consumer is submitted to COREL, the program can return the top n products most likely to be purchased by that customer in the future.

Similarly in paper [3] the authors employed collaborative filtering to predict a user's selection of a new advertisement based on the user's viewing history. The recommendations are made using the N-CRBM model (neighborhood conditional RBM) with joint distribution conditional on similarity and popularity of scores of the neighboring users.

Paper [4] uses customer's purchase and click history to build customer's profile and then incorporate the profile into a one-class CF model. The factorizing results of the model are used to make product recommendation.

In paper [5] Sentimental Analysis of comments on social media for predicting the person's mood which can further affect the stock prices. The authors categorized the person's mood into happy, up, down, and rejected, and the polarity index is calculated which is further supplied to an artificial neural network to predict the results.

In paper [6] Sentimental Analysis can be used to extract meaningful insights from customer reviews and feedback. Previous studies have focused on the sentimental analysis of micro-blog services such as Twitter. The sentiments attached to tweets are analyzed and classified into positive and negative sentiment.

In the paper [7] a prediction model is proposed to anticipate the consumer 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 site. The results indicate the model constructed using decision tree ensembles with Bagging achieved the best prediction of consumer behavior with the accuracy of 95.3%.

In online based shopping platform, the purchasing behavior of the consumers has a great impact. The knowledge-based economy has gotten a lot of hype recently, especially in online shopping apps that track all of the transactions and customer feedback.

3. Methodology

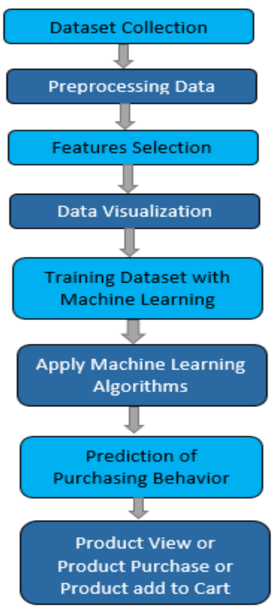

Fig. 3.1. Methodology Diagram

Total research work is accomplished in some stages. Scraping the dataset of consumers purchasing behavior at Covid-19 pandemic situation. Then removing the junk values all the dataset is preprocessed and valid features from the dataset is selected to perform the research work. After exploring all the dataset, selected features are visualized with different prospect. Now the research work is proceeded after training the selected data with different machine learning algorithms and purchasing behavior at this pandemic situation is predicted.

3.1 Overview of Proposed Methodology

To implement this task, first we collected a huge amount of data from different online sources. Then, we analyzed the dataset to determine which features were useful for our specific requirements. We then selected those features to make our proposed system effective. We also took help from Random Forest algorithm to select the best suited features for our proposed method.

Then we visualized our preprocessed data using Python and different python libraries to understand the demand of different products those were sold online. By visualizing the dataset in such a way, we got a clear conception about the relation between consumer behavior and different online e-commerce services. We took the help of pie chart and bar chart to understand the consumer behavior graphically that can be used to make people understand easily.

After that, we trained our preprocessed dataset with some classification machine learning algorithm. 70% data were used to train our model and the remaining 30% data were used to test the validity of our machine learning algorithm. Thus, we created a trained model those were used to predict the consumers behavior during pandemic situation.

3.2 Data Preprocessing

Collected datasets generally appears with unnecessary values. These values should be removed to get the better performance without affecting the contribution of trained models that predict the outcome.

3.2.1 Datasets

We have considered here the dataset of consumers purchasing behavior during Covid-19. The source of this dataset is Kaggle. There were different datasets in relevance of our requirements. But we intentionally selected this one because of the data quantity. Other datasets were available but the data quantity were so little that those were not enough to properly train our machine learning model.

event_time	event_type	product_i	category_	category_	brand	price	user_id	user_session
2019-10-01 00:0(	view	44600062	2.1E+18		shiseido	35.79	5.4E+08	72d76fde-8bb3
2019-10-01 00:0(	view	3900821	2.05E+18	appliance:	aqua	33.2	5.5E+08	9333dfbd-b87a
2019-10-01 00:0(	purchase	17200506	2.05E+18	furniture.living_roc		543.1	5.2E+08	566511c2-e2e3
2019-10-01 00:0:	view	1004237	2.05E+18	electronic	apple	1082	5.1E+08	d7c4761f-de75-
2019-10-01 00:0:	purchase	1004856	2.05E+18	electronic	samsung	130.8	5.4E+08	8187d148-3c41-
2019-10-01 00:0(	view	1801995	2.05E+18	electronic	haier	193	5.4E+08	e3151795-c355-
2019-10-01 00:0(	cart	10900029	2.05E+18	appliance:	bosch	58.95	5.2E+08	901b9e3c-3f8f-
2019-10-01 00:0(	view	1306631	2.05E+18	computer:	hp	580.9	5.5E+08	7c90fc70-0e80-
2019-10-01 00:0(	cart	1005135	2.05E+18	electronic	apple	1748	5.4E+08	c6bd7419-2748-
2019-10-01 00:0(	view	1003306	2.05E+18		apple	588.8	5.6E+08	6ec635da-ea15
2019-10-01 00:0(	cart	4803399	2.05E+18	electronic	jbl	33.21	5.6E+08	8a6afed4-77f8-
2019-10-01 00:0(	view	1480714	2.05E+18	computer:	pulser	921.5	5.1E+08	0d0d91c2-c9c2-
2019-10-01 00:0:	purchase	1005115	2.05E+18	electronic	apple	975.6	5.1E+08	d7c4761f-de75-
2019-10-01 00:0:	view	4300376	2.05E+18		polaris	40.93	5.4E+08	bb8e28c8-d11f-
2019-10-01 00:0:	cart	1307370	2.05E+18	computer:	acer	257.2	5.5E+08	7c90fc70-0e80-
2019-10-01 00:0:	view	3200321	2.05E+18	appliance:	redmonc	91.38	5.2E+08	d35a2b1e-a781
2019-10-01 00:0:	purchase	1002532	2.05E+18	electronic	apple	642.7	5.5E+08	3c80f0d6-e9ec-
2019-10-01 00:0:	view	1004856	2.05E+18	electronic	samsung	130.8	5.4E+08	8187d148-3c41-

Fig. 3.2. Sample Dataset

3.2.1 Feature Selection

Performing this research some concerned features are selected. Selected features are Event Time, Event Type, Product Id, Category Id, Category Code, Brand, Price, User Id and User Session.

3.2.2 Data Conversion into Appropriate Forms

Conducting the research work more precisely some data should be converted into an eligible form. The value of Event Type, Category Code and Brand needs to be converted into numeric value.

3.2.2 Missing Value Handle

After collecting data, there remains some missing values present in the dataset. Some anomalies is observed when we train the dataset with missing values. Because, these missing values creates some anomalies or noise in our trained model. So, we must remove them from the dataset. Missing values from the dataset should be filled using the mean or median function.

3.2.2 EXPLORATORY DATA ANALYSIS

In this section we have analyzed our dataset in different aspect to perform the research in a sound way.

3.2.2.1 Category vs Total Number of Ordered Product

Here this Pie Chart represents the total number of ordered products against that product category. This chart also express Top Categories of the dataset which product order more than 10,000. Among the top categories we can see that, total number of orders were the most for the electronic devices like smartphones, computers, notebooks, laptops and other electronic accessories. The least ordered categories were house appliances like refrigerators, vacuum, air_heater, kitchen washer and other home appliances.

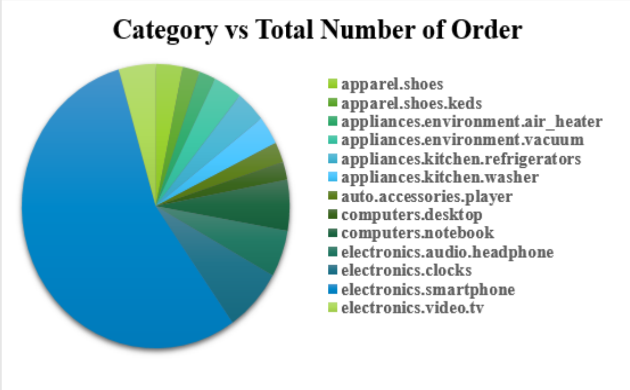

Fig. 3.3 Pie Chart of the total number of orders against all categories.

Table 3.1. Category vs Orders

Category vs Orders	
Category	**Total Order**
apparel. shoes	16867
apparel. shoes. keds	10781
appliances. environment. air_heater	10388
appliances. environment. vacuum	18145
appliances. kitchen. refrigerators	18781
appliances. kitchen. washer	16831
auto. accessories. player	11991
computers. desktop	10502
computers. notebook	30108
electronics. audio. headphone	27567
electronics. clocks	37264
electronics. smartphone	286986
electronics.video.tv	23216

3.2.2.2 Brand vs Total Number of Ordered Product

This Bar Chart represents the number of total ordered products against the product's Brand. It also express Top Brands of the dataset which product order more than 10,000. Among the top brands we can see that, total number of orders were the most for the brand apple and samsung. People believed their product much other than the brands mentioned later during pandemic situation. The least ordered brands were Lenovo, Lg, Sony and other brands like these.

Fig. 3.4. Bar Chart of the total number of orders against all brands

Table 3.2. Brands vs Orders

Brands vs Orders	
Brand Name	Total Order of This Brand
acer	10214
apple	107634
huawei	29360
bosch	12959
lenevo	12567
lg	11495
lucente	17799
samsung	126027
sony	10015

3.2.2.3 Purchase from E- Commerce Sites Before Covid-19 Pandemic vs During Pandemic

Here we have analyzed the purchasing rate from the e-commerce sites before the pandemic situation and during the Covid-19 pandemic situation. Analyzing this we have represented it with a graphical curve. According to the curve we have explored that at the year of 2019 when the pandemic situation not occurred then the purchasing rate of the products from e- commerce websites is less than the current pandemic situation. Gradually the situation became worst and spread all over the world and lockdown took place to decrease the outbreak.

Fig.5. Visual representation of purchasing rate from e-commerce site during Covid-19 pandemic

People started getting dependent on e-commerce sites to purchase their daily life needs. From graph we can visualize that product purchasing rate has increased in the year 2020. Conducting the research work more precisely some data should be converted into an eligible form. The value of Event Type, Category Code and Brand needs to be converted into numeric value.

4. Result Analysis

Analyzing the consumer behavior for purchasing products we have used here some Machine Learning Algorithms and shows 3 cases for purchasing observed here- only view, add to cart or purchase. Resulted output 1 indicate that the product is only view by the consumer, 2 indicates that the product is added to cart and 3 indicates that the product is purchased by the consumer. Here the input attributes are category, brand and price and the output attribute is event type (consumer purchasing behavior). Used Machine Learning Algorithms are – Linear Regression, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest and Naïve Bayes.

4.1 Linear Regression

When modeling the connection between a scalar answer and one or more explanatory factors, linear regression is a linear method. Future predictions may now be made scientifically and with high reliability using linear-regression models. The features of linear-regression models are well understood and can be trained extremely rapidly since linear regression is a statistical technique that has been around for a very long time.

Table 4.1. Predicted result of consumer purchasing behavior by Linear Regression.

Consumers Purchasing Behavior			
Category	**Brand**	**Price($)**	**Predicted Output**
electronics.smartphone	samsung	450	3(purchase)
computer.notebook	hp	1240	1(view)
appliances.kitchen.refrigerators	lg	360	3(purchase)
electronics.audio.headphone	xiaomi	75	2(cart)
apparel.shoes	respect	51	1(view)
electronics.video.tv	elenberg	415	1(view)
appliances.water_heater	artel	72	3(purchase)

4.2 Logistic Regression

Table 4.2. Predicted result of consumer purchasing behavior by Logistic Regression.

Consumers Purchasing Behavior			
Category	**Brand**	**Price($)**	**Predicted Output**
electronics.smartphone	samsung	450	2(cart)
computer.notebook	hp	1240	1(view)
appliances.kitchen.refrigerators	lg	360	1(view)
electronics.audio.headphone	xiaomi	75	3(purchase)
apparel.shoes	respect	51	2(cart)
electronics.video.tv	elenberg	415	3(purchase)
appliances.water_heater	artel	72	2(cart)

4.3 Support Vector Machine (SVM)

Table 4.3. Predicted result of consumer purchasing behavior by Support Vector Machine (SVM).

Consumers Purchasing Behavior			
Category	Brand	Price($)	Predicted Output
electronics.smartphone	samsung	450	2(cart)
computer.notebook	hp	1240	3(purchase)
appliances.kitchen.refrigerators	lg	360	2(cart)
electronics.audio.headphone	xiaomi	75	3(purchase)
apparel.shoes	respect	51	1(view)
electronics.video.tv	elenberg	415	2(cart)
appliances.water_heater	artel	72	2(cart)

4.4 K-Nearest Neighbor (KNN)

Although the K-NN approach is most frequently employed for classification issues, it may also be utilized for regression. Since K-NN is a non-parametric technique, it makes no assumptions about the underlying data. The KNN method simply saves the information during the training phase, and when it receives new data, it categorizes it into a category that is quite similar to the new data.

Table 4.4. Predicted result of consumer purchasing behavior by K-Nearest Neighbor (KNN).

Consumers Purchasing Behavior			
Category	**Brand**	**Price($)**	**Predicted Output**
electronics.smartphone	samsung	450	3(purchase)
computer.notebook	hp	1240	3(purchase)
appliances.kitchen.refrigerators	lg	360	1(view)
electronics.audio.headphone	xiaomi	75	2(cart)
apparel.shoes	respect	51	2(cart)
electronics.video.tv	elenberg	415	1(view)
appliances.water_heater	artel	72	1(view)

4.5 Random Forest

Table 4.5. Predicted result of consumer purchasing behavior by Random Forest.

Consumers Purchasing Behavior			
Category	**Brand**	**Price($)**	**Predicted Output**
electronics.smartphone	samsung	450	3(purchase)
computer.notebook	hp	1240	1(view)
appliances.kitchen.refrigerators	lg	360	3(purchase)
electronics.audio.headphone	xiaomi	75	1(view)
apparel.shoes	respect	51	2(cart)
electronics.video.tv	elenberg	415	1(view)
appliances.water_heater	artel	72	3(purchase)

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 01 | Jan 2023 www.irjet.net p-ISSN: 2395-0072

4.6 Naive Bayes

Table 4.6. Predicted result of consumer purchasing behavior by Naive Bayes.

Consumers Purchasing Behavior			
Category	**Brand**	**Price($)**	**Predicted Output**
electronics.smartphone	samsung	450	3(purchase)
computer.notebook	hp	1240	2(cart)
appliances.kitchen.refrigerators	lg	360	3(purchase)
electronics.audio.headphone	xiaomi	75	1(view)
apparel.shoes	respect	51	2(cart)
electronics.video.tv	elenberg	415	3(purchase)
appliances.water_heater	artel	72	1(view)

4. Conclusion

The main objective of this research work is to assist the owners of the E-commerce websites. As the dependency on e-commerce is increasing at the Covid-19 pandemic according to this research they will be able to know about the consumers purchasing choice and analyze about their products. Consumers purchasing intentions help the owners to have a precise knowledge and make proper scheme about their products using consumer's sentiment. Here consumer's behavior is visualized in various aspects and different machine learning algorithms are used to identify their shopping demands. We found that Random Forest algorithm provides the better prediction than any other algorithms. Such factors identified by the owners of e-commerce organizations will help them take necessary steps to provide a better service to the consumers and add more lifetime value to their business.

REFERENCES

[1] J. S. H. S. D. M. V. Vishwa Shrirame, "Consumer Behavior Analytics using Machine Learning Algorithms," in IEEE, India, 2020.

[2] Z. L. Y. L. Jiangtao Qiu, "Predicting customer purchase behavior in the e-commerce context," Springer, 2015.

[3] U. Z. A. A. L. T. Ben Yedder, "Modeling prediction in recommender," in IEEE, 2017.

[4] J. H. C. Z. Y. C. Yanen Li, "Improving one-class collaborative filtering by incorporating rich user information," in ACM International Conference on Information and Knowledge Management, 2010.

[5] S. K. K. a. A. Srivastava, "Using sentimental analysis in," in 5th International Conference on Reliability, Infocom Technologies, and Optimization (Trends and Future Directions) (ICRITO), Nodia, 2016.

[6] B. R. Chamlertwat, "Discovering consumer insight from twitter via sentiment analysis," Journal of universal computer science, pp. 973-992, 2012.

[7] F. Safara, "A Computational Model to Predict Consumer Behaviour During COVID-19 Pandemic," Springer, 2020.

BIOGRAPHIES

B.Sc. in CSE from PSTU
M.Sc. in CSIT from PSTU
Research Interest:
Data Science, Machine Learning and Data Mining.
Email: muhtasim222@gmail.com

B.Sc. in CSE from CUET
Research Interest:
Internet Security, Machine Learning and NLP
Email: redoanulroni@gmail.com

B.Sc. in CSE from DUET
Research Interest:
Deep Learning, Machine Learning and Image Processing
Email: sumanalicse@gmail.com

B.Sc. in CSE from PSTU
M.Sc. in CSIT from PSTU
Research Interest:
Data Science, Machine Learning and Data Mining.
Email: sourovhossen96@gmail.com