**ONLINE SHOPPERS REVENUE PREDICTIONS**

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**

in

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

by

**RISHITHA.G (2103A52016)**

**SIRICHANDANA.M (2103A52057)**

Under the guidance of

**Mr. RAMESH**

Assistant Professor, Department of CSE.



**Department of Computer Science and Artificial Intelligence**



**Department of Computer Science and Artificial Intelligence**

**CERTIFICATE**

This is to certify that project entitled **“ONLINE SHOPPERS REVENUE PREDICTION** " is the bonafied work carried out by **RISHITHA.G, SIRICHANDANA.M** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

**Mr. S Naresh Kumar Dr. M.Sheshikala**

Asst. Professor, Assoc. Prof. & HOD (CSE),

S R University, S R University,

Ananthasagar, Warangal. Ananthasagar, Warangal.

**ACKNOWLEDGEMENT**

We express our thanks to Course co-coordinator **Mr.** **S.Naresh Kumar, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude toHead of the department CS&AI, **Dr. M.Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

**ABSTRACT**

**ANALYSIS OF ONLINE SHOPPERS WISH LIST AND PREDICTING THE REVENUE**

n this study we investigate what motivates online shoppers and whether they have any

expectations which if fulfilled will further enhance their behavioral intention to shop online. We

use the grounded theory approach to generate the initial set of response from online shoppers

through 10 focus groups sessions. The responses were content analyzed by experts and after

codification reduced to a list of 39 unique items. A questionnaire containing these items were then

administered to 222 participants and their responses analyzed. Factor analysis revealed 6 distinct

factors with Psychological Safety emerging as a new 9-item construct demonstrating the highest

correlation with shoppers’ online shopping intentions followed by utilitarian value, quality value,

hedonic value, epistemic value and social value derived by shoppers from shopping online. The

implications of these findings for online shopping business as well as for marketing of products

and services in general are discussed.

n this study we investigate what motivates online shoppers and whether they have any

expectations which if fulfilled will further enhance their behavioral intention to shop online. We

use the grounded theory approach to generate the initial set of response from online shoppers

through 10 focus groups sessions. The responses were content analyzed by experts and after

codification reduced to a list of 39 unique items. A questionnaire containing these items were then

administered to 222 participants and their responses analyzed. Factor analysis revealed 6 distinct

factors with Psychological Safety emerging as a new 9-item construct demonstrating the highest

correlation with shoppers’ online shopping intentions followed by utilitarian value, quality value,

hedonic value, epistemic value and social value derived by shoppers from shopping online. The

implications of these findings for online shopping business as well as for marketing of products

and services in general are discussed.

In this study we investigate what motivates online shoppers and whether they have any

expectations which if fulfilled will further enhance their behavioral intention to shop online. We

use the grounded theory approach to generate the initial set of response from online shoppers

through 10 focus groups sessions. The responses were content analyzed by experts and after

codification reduced to a list of 39 unique items. A questionnaire containing these items were then

administered to 222 participants and their responses analyzed. Factor analysis revealed 6 distinct

factors with Psychological Safety emerging as a new 9-item construct demonstrating the highest

correlation with shoppers’ online shopping intentions followed by utilitarian value, quality value,

hedonic value, epistemic value and social value derived by shoppers from shopping online. The

implications of these findings for online shopping business as well as for marketing of products

and services in general are discussed.

In this study we investigate what motivates online shoppers and whether they have any

expectations which if fulfilled will further enhance their behavioral intention to shop online. We

use the grounded theory approach to generate the initial set of response from online shoppers

through 10 focus groups sessions. The responses were content analyzed by experts and after

codification reduced to a list of 39 unique items. A questionnaire containing these items were then

administered to 222 participants and their responses analyzed. Factor analysis revealed 6 distinct

factors with Psychological Safety emerging as a new 9-item construct demonstrating the highest

correlation with shoppers’ online shopping intentions followed by utilitarian value, quality value,

hedonic value, epistemic value and social value derived by shoppers from shopping online. The

implications of these findings for online shopping business as well as for marketing of products

and services in general are discussed.

In this study we investigate what motivates online shoppers and whether they have any expectations which if fulfilled will further enhance their behavioral intention to shop online. We use the grounded theory approach to generate the initial set of response from online shoppers through 10 focus groups sessions. The responses were content analyzed by experts and after codification reduced to a list unique items. A questionnaire containing these items were then administered to participants and their responses analyzed. Analysis revealed distinct factors with the highest correlation with shoppers' online shopping intentions followed by utilitarian value, quality value, hedonic value, epistemic value and social value derived by shoppers from shopping online. The implications of these findings for online shopping business as well as for marketing of products and services in general are discussed. Statement of Contribution This study makes unique and important contributions to our body of knowledge in online shopping. Today, and which leads to a revenue improvement when we know what soppers are liking and what they are not liking the business persons can improve the product and generate good revenue .Instore and Online shopping outlets not only compete among themselves for shopper attention but also with each other. It is therefore important to understand evolving shopper expectations. Fulfilling the expectations, will not only result in well-being of online shoppers but will also make online shopping businesses more competitive.

Our project is focused on applying Machine Learning classification models to e-commerce website data. Specifically, we analyze if machine learning is an effective method in predicting whether or not a visitor to an e-commerce website will make a purchase or not. We performed exploratory data analysis in order to visually analyze our data and identify any patterns or trends. We then performed model selection, testing a selection of 4 classification models against a base model. Our main metric in assessing our models was recall. In model selection we determined that the model that appeared to be the best suited to answer our research question was a Random Forest classifier. We tuned the hyperparameters of this model further, and obtained a final recall score of 0.81 on our test set. Ultimately, we concluded that it is feasible to create a machine learning model to predict purchase conversion for an e-commerce website, however, before an organization commits to implementing a machine learning model, we would recommend more simple methods be considered first.

**Table of Contents**

**Chapter No. Title Page No.**

1. Introduction
   1. Overview 1
   2. Problem Statement 1
   3. Existing system 1
   4. Proposed system 2
   5. Objectives 2
   6. Architecture 2-3
2. Literature survey
   * 1. Document the survey done by you 4
3. Data pre-processing
   1. Dataset description 5
   2. Data cleaning 6
   3. Data augmentation 6
   4. Data Visualization 7
4. Methodology
   1. Procedure to solve the given problem 8-9
   2. Model architecture 9
   3. Software description 9
5. Results and discussion 10
6. Conclusion and future scope 11
7. References 12

**1.INTRODUCTION**

1.1,Overview:

In this project , we propose a real-time online shoppers behavior analysis system consisting of two modules which simultaneously predicts the visitor’s shopping intent and Web site abandonment likelihood .In the first module ,we predict the purchasing intention of the visitor using aggregated data kept track during track during the visit along with some sessions and user information . purchasing interest improves the success rate of the system.  The modules are used together to determine the visitors which have purchasing intention but are likely to leave the site in the prediction horizon and take actions accordingly to improve the Web site abandonment and purchase conversion rates and thus tells us about the revenue profit and loss .

1.2, Problem Statement:

The problem is that how to identify the intensions of the bulk number of costumers on a particular e-commerce purchase, predicting review ,most purchased days / high revenue days . Given clickstream and session data of a user who visits an e-commerce website, can we predict whether or not that visitor will make a purchase?

Nowadays, it is common for companies to sell their products online, with little to no physical presence such as a traditional brick and mortar store. Answering this question is critical for these types of companies in order to ensure that they are able to remain profitable. This information can be used to nudge a potential customer in real-time to complete an online purchase, increasing overall purchase conversion rates. Examples of nudges include highlighting popular products through social proof, and exit intent overlay on webpages.

1.3.Existing system:

The present scenario for shopping is to visit the shops and market manually and then from the available product list one needs to choose the item he or she wants and then pay for the same item mainly in cash mode is done, as not every society is well educated and aware to use net banking or card modes or wallets etc.

This system is not much user-friendly as one needs to go to the market physically and then select items only from the available list. So mostly it is difficult to get the product as per our desire. Description About the products is less available and are mostly verbal only. For this type of shopping, one needs to have an ample amount of free time.

Also, not really good markets exist everywhere, so many times good markets become out of reach for certain people. In the proposed system customers need not go to the shops for purchasing the products. He/she can order the product he/she wishes to buy through the use of this system. The shop owner can be the admin of the system.

The shop owner can appoint officials particularly to handle this, who will help the owner in managing the customers and product orders. The system also endorses a home delivery system for delivering the purchased products.

1.4.Proposed system :

The proposed system involves a costumer buying or purchasing a product through an e-commerce website there is no traditional way of purchasing available it can be accessed by everyone its less time taking and the executives can identify high selling product and display It on front page so that many people buys it and the revenue can be increased this way the revenue will be generated .

1.5.Define Objectives:

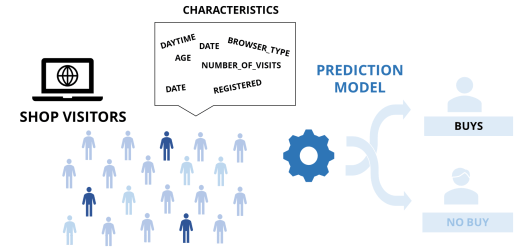
This project aims to use the information customers may leave in the form of the trace of browsing history data or user information when they visit an online shopping site. With the help of this information, the project aims to predict online shoppers' purchasing intention by using clickstream and session information data. The project aims to create a machine learning model based on this information to predict costumer’s purchasing intensions. The objective of the project is to build a machine leaning that can predict costumer purchase revenue as accurately as possible

1.6.Overall architecture :

The system being designed is a software application that takes in specific data collected from google analytics and predicts the customer purchase intention. The application produces a result in as either 1 or 0 for the feature revenue which is represented by Boolean value True or False. Generally, the machine learning algorithm model will learn the probability of each event based on the variables or features of the dataset and make a prediction. The internal working of each machine learning algorithm will differ. I shall be measuring the fitness and accuracy of each of the machine learning algorithms on the processed dataset before and after optimization of the models and comparing the performance for most accurate solution. In this project, I will be aiming to achieve as high an accuracy as possible in predicting the revenue.

The system architecture is shown below:





The data will be processed ,the computer takes the data use suitable algorithm and classifies into buys and no buys which is revenue in terms of Boolean value yes/no.



**2.LITERATURE SURVEY**

* + 1. Document the survey done by you related to your problem statement

This survey combines factors that other surveies have done that will act upon the consumer ‘s buying determination in online and offline shops for dresss. It includes the monetary value attraction, clip economy, perceived hazard, enjoyment and exhilaration, tangibleness and high interactivity. All of these factors will lend to the survey of client ‘s buying purpose for dresss on both shops which includes online and offline shopping. In on-line shopping, it is expected that shoppers are more likely to tie in monetary value attraction and clip salvaging with their purpose to shop while in offline shopping, consumers are more likely to tie in tangibleness, high interactivity and enjoyment with their purpose to shop. As a consequence, on-line sellers or retail merchants should be cognizant of the jobs faced by the consumers and their perceived hazard to increase their purpose to shop in online. Interior designers must take note of consumers ‘ demands because the serviceability is the get downing point to acquire the assurance and support of the consumers .

Nowadays, people use internet as a daily device to access their emails, do online tasks, to read online headline, look for certain information and many other functions (Bellman, 2001). This daily usage of the Internet by these people takes them naturally to operate it as a shoppers stop too. Additional constituents found which makes a difference to the buying behaviour of the buyers and their attitudes include their culture surroundings, particular needs, commitment of the product, mood to trust, to what extent buyers can easily share their personal information and their willingness to invest on internet buying (Bellman et al 1999).

**3.DATA PRE-PROCESSING**

1.1 Describe dataset :

A good data set is one that suits your purpose.

On the whole, look for data sets that meet the following conditions:

1.Contain the elements you need

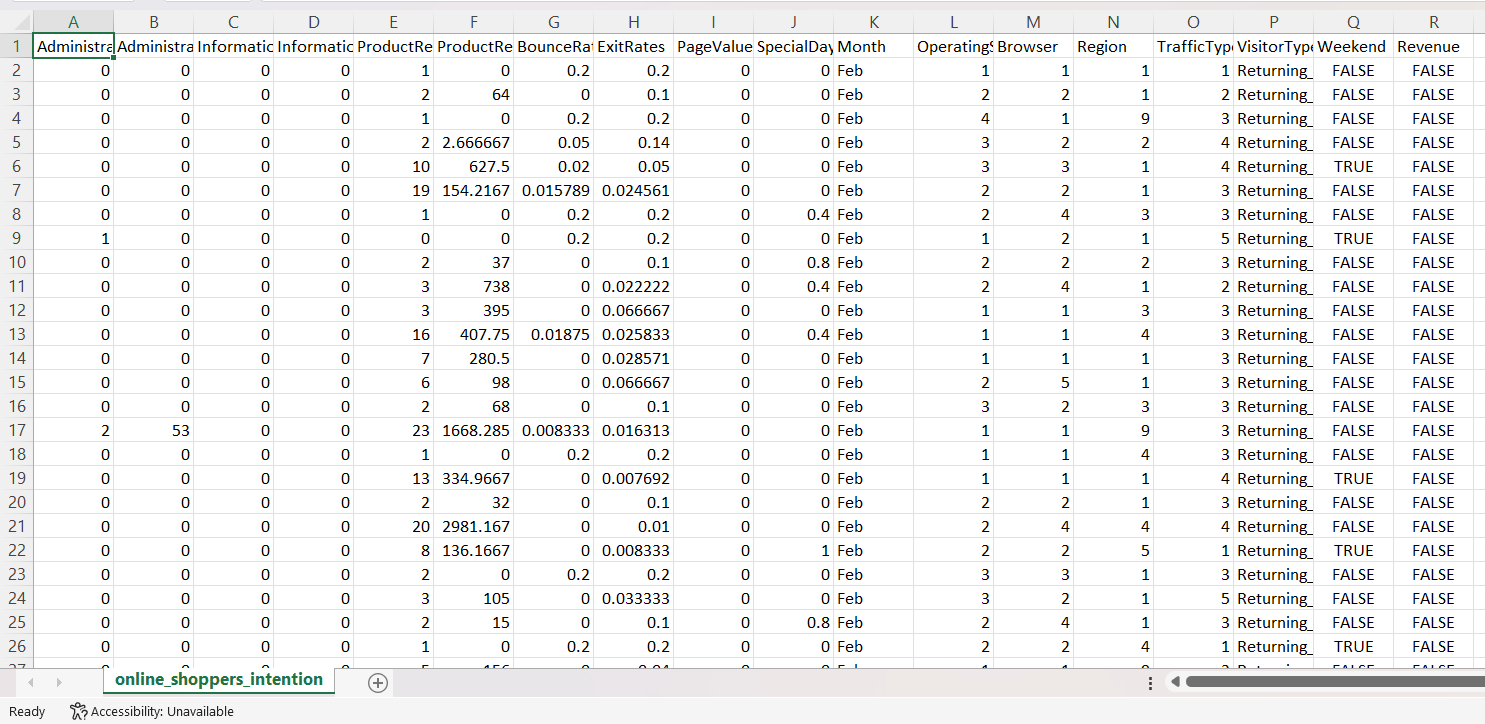
2.Are disaggregated data

3.Have at least a couple dimensions and a couple measures

4.Have good metadata or a data dictionary

5.Are useable

In our data set several features are involved to make a perfect data set like :



1.**Administrative**: Number of pages of this type that the user visited.

2**.Administrative duration**: Amount of time spent in this category of pages.

3.**Informational:** Number of pages of this type that the user visited.

4.**Informational duration**: Amount of time spent in this category.

5.**Product related**: Number of pages of this type that the user visited.

6.**Product related duration**: Amount of time spent in this category of pages.

7.**Bounce rates:** Percentage of visitors who enter the website through that page and exit without triggering any tasks

8.**Exit rates:** Percentage of pageviews on the website that end at that specific page

9.**page values**: Average value of the page averaged over the value of the target page and/or the competition of eCommerce.

10.**Special days** : This value represents the closeness of the browsing date to special days or holidays(eg : mother’s day).

11.**Month:** contains the month the pageview occurred  
 12.**Operatingsystems:** an integer value representing the operating system that the user was on when viewing the page  
 13.**Browswer**: an integer value representing the browser that the user was using to view the page  
 14.**Region:** integer value representing which region the user located in   
 15.**Traffic type** : integer value representing what type of traffic the user is categorised into  
 16.**Visitor type** : string representing either the user is new visitor ,returning visitor , or other  
 17.**Weekend:**A Boolean representing whether the session is on a weekend  
 18**.Revenue**:A Boolean representing whether the user completed the purchase or not.

1.2.Data cleaning :

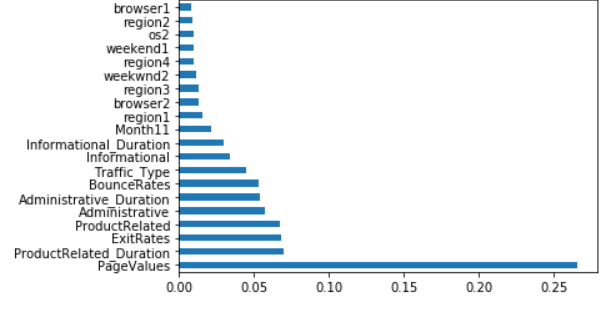
Data cleaning is the process of fixing or removing incorrect ,corrupted ,incorrectly formatted, duplicate, or incomplete data within a dataset. , I will be cleaning the dataset of missing values and other such human errors. These could include wrong entry of data in a field for example, String value in an Integer Datatype or even typing errors. However, the dataset being used for this project has data formatting at the point of collection and so typing errors and incorrect datatypes are unlikely. The dataset used for this project is data collected through google analytics which has certain formatting in place at point of data collection. Although areas such as missing values, imbalanced dataset, outliers, and other abnormalities in the dataset will have to be addressed before the dataset can be used for building the machine learning model. By removing unwanted data from the dataset, the final dataset used will be more effective as there will be less incorrect information or noise in the data.

1.3. Data augmentation :

Data augmentation is a process of artificially increasing the amount of data by generating new data points from existing data. This includes adding minor alterations to data or using to generate new data points in the latent space of original data to amplify the dataset. In our model we have generated or added certain data points from existing data .This will make our prediction or result more accurate .

1.4. Data Visualization:

The representation of information in the form of a chart, diagram , picture , etc.,



**4.METHODOLOGY**

1.4.Procedure to solve the given problem :

We use logistic regression and knn to solve this problem . I have trained the model on algorithms such as Logistic Regression, KNN,. Logistic regression model is used to caluculate probabilities of a class label or even existing such as pass/fail, win/lose. Knn is a non-parametric ,supervised learning classifier,which uses proximity to make classifications or predictions about the grouping of individual data point. But I had the most accurate validation score in KNN (K=1 with uniform weights). It gives an Accuracy of 84.03% of Training Accuracy and about 80.6% of Validation Accuracy, and about 57% of Validation with some extra features added to it.

**Code for finding predictions:**

import pandas as pd

df = pd.read\_csv('osi.csv')

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

df['VisitorType'] = le.fit\_transform(df['VisitorType'])

df['Month'] = le.fit\_transform(df['Month'])

df['Weekend'] = le.fit\_transform(df['Weekend'])

df['Revenue'] = le.fit\_transform(df['Revenue'])

df['ProductRel\_per\_dur'] = df['ProductRelated']/(df['ProductRelated\_Duration']+0.00001)

df['Admin\_per\_dur'] = df['Administrative']/(df['Administrative\_Duration']+0.00001)

df['Inform\_per\_dur'] = df['Informational']/(df['Informational\_Duration']+0.00001)

X1 = df[['ProductRel\_per\_dur','Admin\_per\_dur','Inform\_per\_dur','VisitorType','SpecialDay','PageValues','Region','Month']]

y1 = df['Revenue']

from sklearn.model\_selection import train\_test\_split

X1,x2,y1,Y2 = train\_test\_split(X1,y1,test\_size=0.25)

from imblearn.combine import SMOTETomek

smk = SMOTETomek(random\_state=42)

X\_res,y\_res = smk.fit\_sample(X1,y1)

from sklearn import preprocessing

scaler = preprocessing.StandardScaler().fit(X\_res)

X1 = scaler.transform(X\_res)

from sklearn.neighbors import KNeighborsClassifier

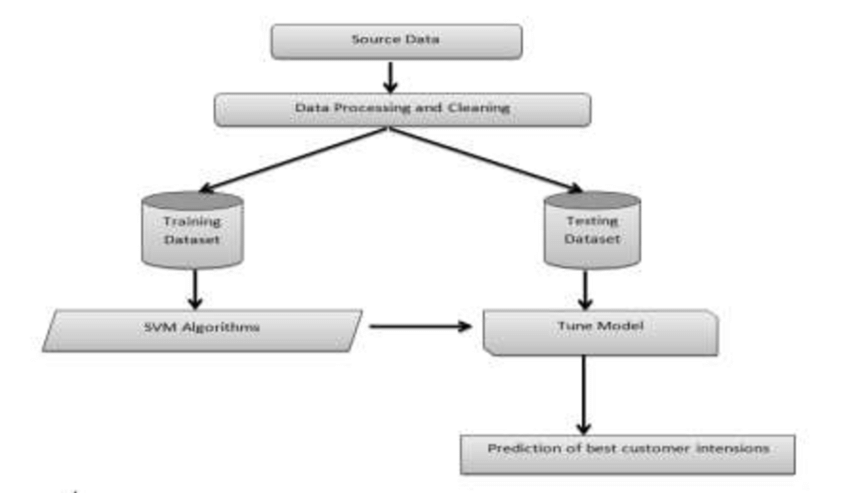
neigh = KNeighborsClassifier(n\_neighbors=1,algorithm='ball\_tree',weights='uniform')

neigh.fit(X1,y\_res)

neigh.score(X1,y\_res)

**The Training Acc:** 0.840375899760064  
**The Validation Acc:** 0.8053843658773921  
  
**Training Precision:** 0.8588896697118763  
**Validation Precision:** 0.43122270742358076

# 1.2 .Model architecture :



# 

* 1. Software description :

**Python:** is a programming language that support large number of open-source libraries that are used in implementing machine learning algorithm and for data visualization and analysis.

**Pandas**: is open-source data analysis and manipulation tool, built on top of the Python programming language.

**Google CoLab:** It is an online programming platform that is especially well suited to machine learning, data analysis and education. I have used google colab as it provides better hardware for Machine Learning computing for example for training the Machine Learning.

**5. RESULTS AND DISCUSSION**

With e-commerce becoming more and more prevalent in today’s economy, businesses within this sector need to understand what factors influence a visitor to transform into a purchaser. In their study, our learners aimed to build a machine learning set-up for the online shopping environment to predict the purchase intentions of prospective buyers through various analytical models. The sample data had the information of 12330 users, each containing metrics of web visits of a user within a one-year timeframe out of which 15% users have made a purchase. The dataset had Administrative, Product, Demographic and Navigation Information for each user such as different types of pages visited by the visitor in a session and total time spent in each of these page categories. Information like bounce rate, exit rate and page value features are also available. The bounce rate feature indicates the percentage of visitors who entered the site from a particular page at the site and left the website without any activity. The exit rate of a page is the percentage of users who have their last session on that page. The page value indicates the average value of a page that visitors have visited before purchasing any product.

**6.CONCLUSION AND FUTURE SCOPE**

Before we launch into the future and scope of eCommerce in India, let us first understand what is e-commerce. To put it simply, electronic commerce refers to the purchase and sale of goods online or via the [**internet**](https://webandcrafts.com/services/internet-marketing). Sellers make websites where they display images of their products with price and description. Shoppers who buy the products have multiple payment options like COD, e-wallet, net banking, credit card, and so on. Online sellers have the responsibility of shipping the product to the buyer and ensuring safe and timely delivery. Today with the high penetration of the internet and mobile devices – currently there are about 776.45 million internet connections in India – it has also become very easy to buy and sell online. Online selling and purchasing offer innumerable benefits to both sellers and buyers, and these advantages are also the reasons for the rising scope of eCommerce. As of 2017, the total value of the e-commerce business in India was USD 38.5 billion, according to [IBEF](https://www.ibef.org/industry/ecommerce.aspx), and it is expected to hit USD 200 billion by 2026. That gives us a good idea about the future scope of e-commerce in India.

-Well, to put it bluntly, the scope of e-business in the near future looks to be ever-increasing and growing, because the trend has really caught on here. E-commerce giant Amazon is keen to conquer the Indian market and has already invested a great deal, especially with its 49% stake in the Future Group. Indian online retail giant Flipkart has already opened a few offline stores and plans more stores in smaller cities. They plan to combine online and offline stores to maximize their selling potential. Google and Tata Trust have launched a joint program to increase internet and mobile penetration among rural women.

-The Government of India is also making a huge push for Ecommerce by providing numerous sops to start ups, cyber parks, and so on through its Digital India program. As of now, there are close to 20,000 E-commerce companies in India, with many more expected to join the bandwagon every month. Experts are of the opinion that by 2034, we will outperform the US where online shopping is concerned, becoming the second-largest e-commerce economy in the world.

**7.REFERENCES**

[1] Stuart J. Russell, Peter Norvig (2010) Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall ISBN 9780136042594.

[2] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN 9780262018258.

[3] https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590 [Accessed 11/01/2021]

[4] https://www.ibm.com/cloud/learn/unsupervised-learning [Accessed 16/02/2021]

[5] Roman, Victor (2019-04-21). "Unsupervised Machine Learning: Clustering Analysis". Medium. Retrieved 2019-10-01.

[6] <https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a>

[7] https://courses.cs.duke.edu//fall03/cps260/notes/lecture18.pdf [Accessed 16/02/2021]

[8] Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes"

[9] Greene, William H. (2012). Econometric Analysis (Seventh ed.). Boston: Pearson Education. pp. 803–806. ISBN 978-0-273-75356-8.

[10] https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm