

# Consumer purchase behaviour analysis

Using Randomforest Regression Algorithm

Developed By: Bhavani Poshala , Jyothsna Mandala , Imranullah Khan , Abhinand Pallapu

## Smart Bridge-Remote Summer Internship Program

### 1. INTRODUCTION

Machine Learning is probably one of the most commonly discussed topics in the world of technology. In fact, most technology companies from start-ups to MNCs have today invested in Artificial Intelligence (AI) and Machine Learning (ML) projects.

Consumer Behavior:

The ever-changing market trends make it difficult for businesses to understand consumer behavior accurately. A detailed insight is needed to understand buying and consumption pattern of consumers, needs/wants/motives to buy/sell/use a particular product, and capture which type/category of customers are based on demographic variables. Techniques like Machine Learning can bring the desired results oriented towards user engagement and revenue generation. Marketing departments are more than ever engaged in discovering newer ways of making advertising and branding campaigns effective.

Consumer behavior is comparatively a new field of study which evolved just after the Second World War. The sellers market has disappeared and buyers market has come up. This led to paradigm shift of the manufacturer's attention from product to consumer and specially focused on the consumer behavior. The evaluation of marketing concept from mere selling concept to consumer-oriented marketing has resulted in buyer behavior becoming an independent discipline. The growth of consumerism and consumer legislation emphasizes the importance that is given to the consumer. Consumer behavior is a study of how individuals make decision to spend their available resources (time, money and effort) or consumption related aspects (What they buy? When they buy?, How they buy? etc.

Consumer Decisions & Machine Learning

The Consumer Decision Processes (also known as Buyer Decision Process) refers to the decision-making stages that a consumer undergoes before, during, and after purchasing a product or service.

- Need Recognition: This is recognized as the first and most crucial step in the

process. Need in fact, is the catalyst which triggers buying decision of consumers.

- Research: After the consumer has developed a want or a need, they initiate information search about the different alternatives available in the market
- Evaluation of Alternatives and Purchase: An individual, after gathering relevant information, tries to choose the best option available, as per their needs, tastes and budget. Finally they purchase the product

## **1.1 Overview:**

Machine learning and data analytics play a vital role in the consumer buying behavior scenario.

Source: Internet

1) Consumer Online Search Google and its competitors are constantly improving on what the search engine understands. Whenever we search on Google, the search engine watches how you respond to results. If you click the top result and stay on that page, it assumes that we received the right information and that the search was a success. If you click on the second page (of results), or type in a new search string without clicking any of the results, the search engine will learn from that mistake and use it to deliver a better results in future.

2) Marketing Personalization Most of us had the experience of visiting a physical store, looking at a product and not buying it. We then visit the web for the same product. This is something that eCommerce companies can use increasingly for personalization. This can be done through promotional mails. After analysing consumer search patterns, such companies can also target people through coupons, offers and product recommendations which ultimately culminate in a sale.

3) Learning from past behaviour of consumers: A major advantage of Machine Learning is that we can learn from past predictions and outcomes, and continually improve predictions based on new and different data.

4) Recommendation Systems: Recommendation systems are used by eCommerce sites to suggest relevant products to their customers. Types of recommendation include suggesting products to the consumer, which products others bought, summarisation of community opinion, and providing customer reviews.

5) Product Matching to Preferences: Machine learning coupled with data analytics can present a crystal-clear picture of consumer behavior. Algorithms play a crucial role when it comes to determining price points that can help leading brands position products to the right target market.

6) "Up-sell" and "Cross-sell" Cross-selling is when you recommend that your customer buy a product that compliments an existing purchase from a different category or

vendor. Such recommendations are a great way to help visitors discover other products on your online store and deliver a personalised shopping experience to them.

Final Word : Machine Learning is a great way to gain insights into customer behavior and give them what they want well ahead. It ensures their loyalty when combined with the right price, quality and innovation. Shopping behavior is now being tracked relentlessly to make the most out of customer dollars.

## **1.2 Purpose**

Our aim from the project is to make use of pandas, matplotlib, & seaborn libraries from python to extract the libraries for machine learning for the purchase prediction

Secondly, to learn how to analysis the consumer purchase behavior using Regression algorithms in machine learning .

And in the end, to predict the consumer purchase behavior analysis .

## **2. LITERATURE SURVEY**

Research indicates that the cost of retaining a customer is less than attracting new ones. This is due to marketing costs required to appeal to new customers. For this reason, together with the increase of competition it has become pivotal that the current customers base is retained. Normally, customers churn gradually and not abruptly. This means that by analyzing customers historic buying patterns one can adopt a proactive approach in predicting churn. Since all transactions are inserted through POS and stored in databases, understanding customers' needs and patters is possible as data is accessible.

According to, executives are dedicating marketing budgets to focus on customer retention campaigns. Various models designed to predict churn focus on statistical and renowned machine learning algorithms including Random Forest and Logistic Regression. This paper focuses on two aspects when predicting churn within the grocery retail industry. The first is based on the features which will be passed on to the model. Instead of using customers buying trends to cluster the individuals, these values will be created as features and are passed to the model. Therefore, for each customer various features are created to allow the model to learn and identify patterns per individual. For this reason, two datasets are created to test and evaluate how data should be represented to predict churn. The second aspect is the implementation of the algorithms. The novelty of this study is the use of deep learning to predict churn within the grocery industry. To our knowledge, this is the first study which implements deep

learning within this industry. The strength of using deep learning is that it can reveal hidden patterns within the available dataset. An important aspect within the business is to have a good understanding of customers' needs, whereby holistic views of their patterns may be analyzed. When customers are satisfied with the service or products, customer loyalty increases.

## **2.1 Existing Problem**

The previous models have less accuracy and the predictions are not accurate whereas this model is constrained with the lot of advantages and with a higher accuracy than any other model already proposed. In this model we used Machine learning algorithm named Randomforest Regression which give an accuracy 67% of the predicted problem and there is an user friendly user interface to check the consumer purchase behavior analysis.

## **2.2 Proposed Solution**

Machine Learning ( Randomforest Regression):

Randomforest Regression is one of the simplest and most common supervised machine learning algorithms that data scientists use for predictive modeling. We'll use Randomforest regression to build a model that predicts the consumer purchase behavior analysis .And also we have created an UI using the Flask for the purchase prediction and this UI will allow the users to predict their purchase behavior analysis very easily and the User interface is user friendly not at least one complication in using the interface, and it can be used just by entering some necessary details into the UI in real time it'll give the predicted value like the purchase prediction.

Randomforest Regression is one of the simplest and most common supervised machine learning algorithms that data scientists use for predictive modeling. We'll use Randomforest regression to build a model that predicts the consumer purchase behavior analysis .And also we have created an UI using the Flask for the purchase prediction and this UI will allow the users to predict their purchase behavior analysis very easily and the User interface is user friendly not at least one complication in using the interface, and it can be used just by entering some necessary details into the UI in real time it'll give the predicted value like the purchase prediction.

## **3. THEORETICAL ANALYSIS**

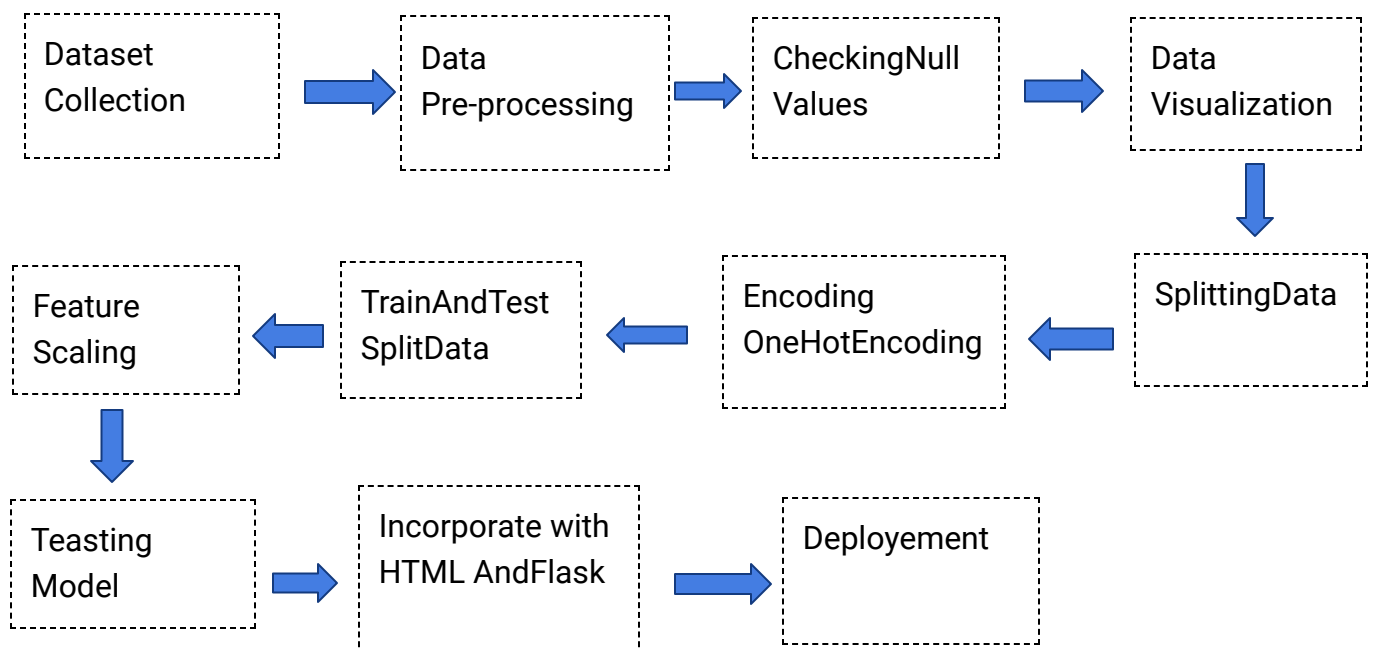
While selecting the algorithm that gives an accurate prediction we gone through lot of algorithms which gives the results abruptly accurate and from them we selected only

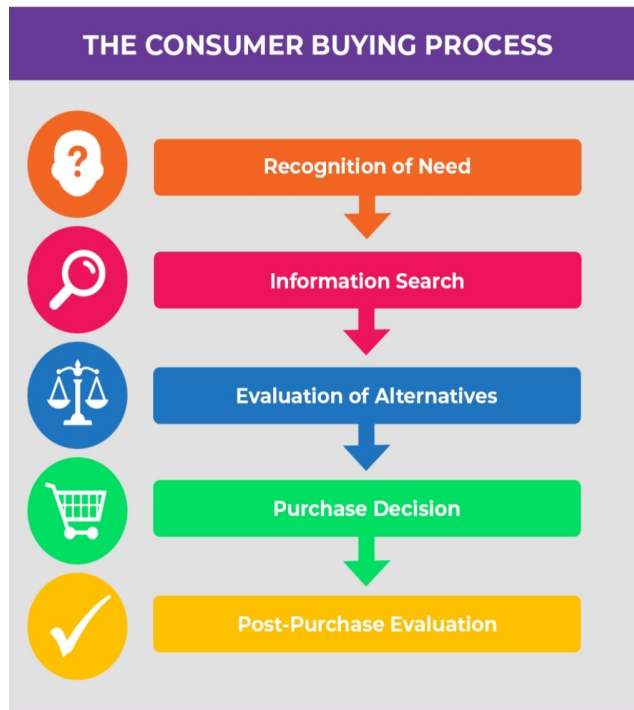
one algorithm for the prediction problem that is Randomforest Regression , it gives the output based on the independent variables accurately.

The peculiarity of this problem is collecting the consumer dataset real time and working with the prediction at the same time, so we developed an user interface for the consumer purchase behavior analysis for purchase prediction. There are several ways to check the Randomforest Regression model accuracy. Usually we use Root Mean Squared Error. We train Randomforest regression model by adding or removing the features to dataset, and see which one has the lowest RME- is the best one.

At first we got like lot of worst accuracies because we tried lot of algorithms for the best accurate algorithm , finally after all of that we tried the best suitable algorithm which gives the prediction accurately is Randomforest Regression Algorithm. And developed it to use as a real time prediction probelm for the admission prediction.

### 3.1 Block Diagram





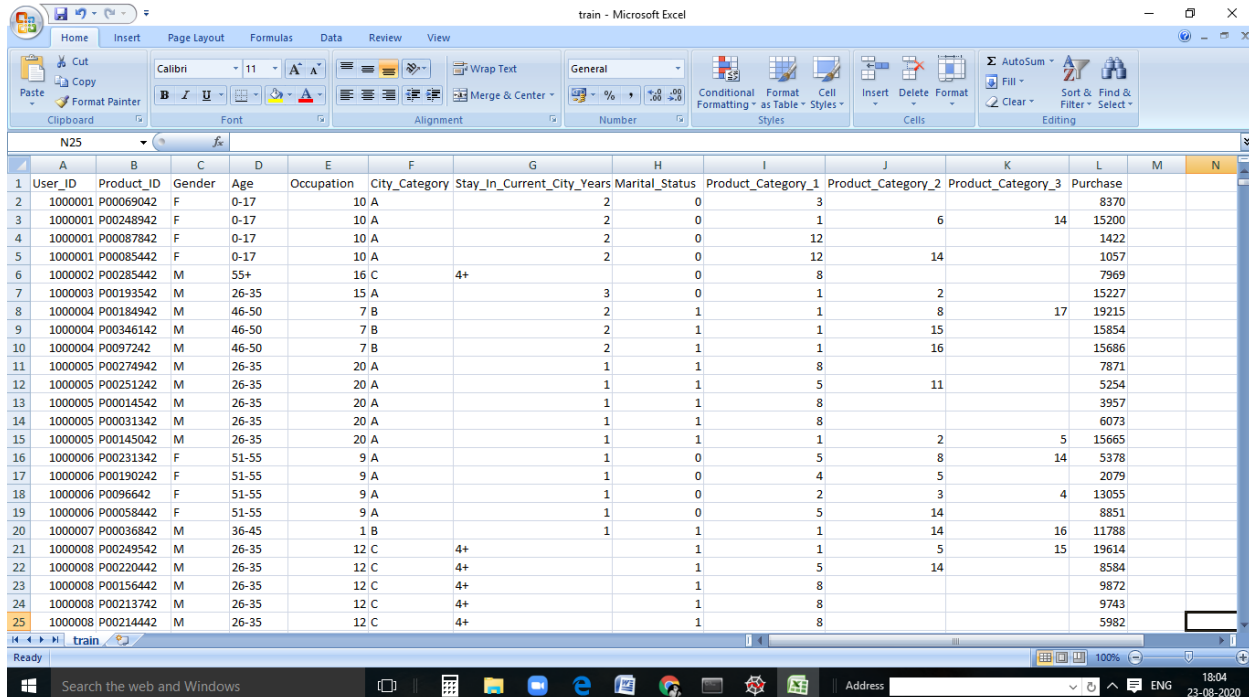
## 3.2 Software Designing

- Jupyter Notebook Environment
- Spyder Ide
- Machine Learning Algorithms
- Python (pandas, numpy, matplotlib, seaborn, sklearn)
- HTML
- Flask

We developed this purchase prediction by using the Python language which is a interpreted and high level programming language and using the Machine Learning algorithms. for coding we used the Google Colab and Jupyter Notebook environment of the Anaconda distributions and the Spyder, it is an integrated scientific programming in the python language. For creating an user interface for the prediction we used the Flask. It is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions, and a scripting language to create a webpage is HTML by creating the templates to use in the functions of the Flask and HTML.

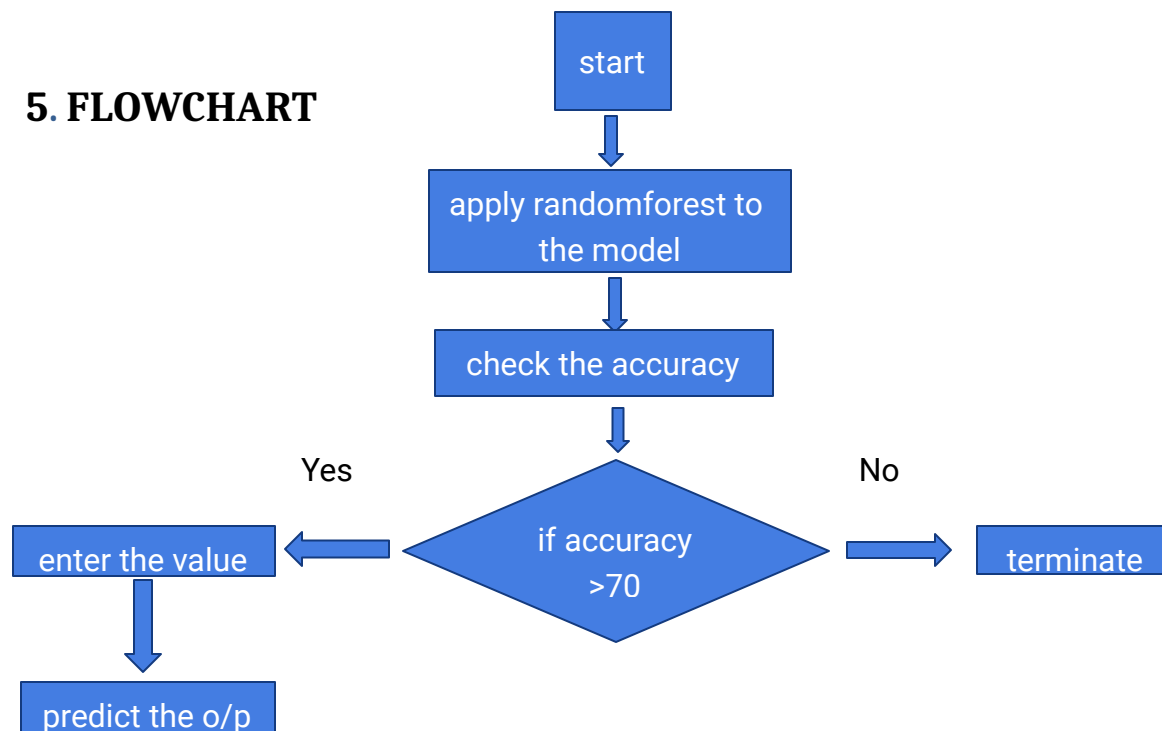
## 4. EXPERIMENTAL INVESTIGATION

The dataset we used is derived from [www.kaggle.com](http://www.kaggle.com). It contains plenty of datasets which are real time. We choosed consumer purchase behavior analysis prediction dataset which contains 11 attributes and 5lakhs rows. After that, the missing values are checked and the unwanted columns are deleted, finally we have retained to 10 attributes. Those are shown below.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	User_ID	Product_ID	Gender	Age	Occupation	City_Category	Stay_In_Current_City_Years	Marital_Status	Product_Category_1	Product_Category_2	Product_Category_3	Purchase		
1	1000001	P00069042	F	0-17	10 A		2	0	3			8370		
2	1000001	P00248942	F	0-17	10 A		2	0	1	6	14	15200		
3	1000001	P00087842	F	0-17	10 A		2	0	12			1422		
4	1000001	P00085442	F	0-17	10 A		2	0	12	14		1057		
5	1000002	P00285442	M	55+	16 C	4+		0	8			7969		
6	1000003	P00193542	M	26-35	15 A		3	0	1	2		15227		
7	1000004	P00184942	M	46-50	7 B		2	1	1	8	17	19215		
8	1000004	P00346142	M	46-50	7 B		2	1	1	15		15854		
9	1000004	P0097242	M	46-50	7 B		2	1	1	16		15686		
10	1000005	P00274942	M	26-35	20 A		1	1	8			7871		
11	1000005	P00251242	M	26-35	20 A		1	1	5	11		5254		
12	1000005	P00014542	M	26-35	20 A		1	1	8			3957		
13	1000005	P00031342	M	26-35	20 A		1	1	8			6073		
14	1000005	P00145042	M	26-35	20 A		1	1	1	2	5	15665		
15	1000006	P00231342	F	51-55	9 A		1	0	5	8	14	5378		
16	1000006	P00190242	F	51-55	9 A		1	0	4	5		2079		
17	1000006	P0096642	F	51-55	9 A		1	0	2	3	4	13055		
18	1000006	P00058442	F	51-55	9 A		1	0	5	14		8851		
19	1000007	P00036842	M	36-45	1 B		1	1	1	14	16	11788		
20	1000008	P00249542	M	26-35	12 C	4+		1	1	5	15	19614		
21	1000008	P00220442	M	26-35	12 C	4+		1	5	14		8584		
22	1000008	P00156442	M	26-35	12 C	4+		1	8			9872		
23	1000008	P00213742	M	26-35	12 C	4+		1	8			9743		
24	1000008	P00214442	M	26-35	12 C	4+		1	8			5982		

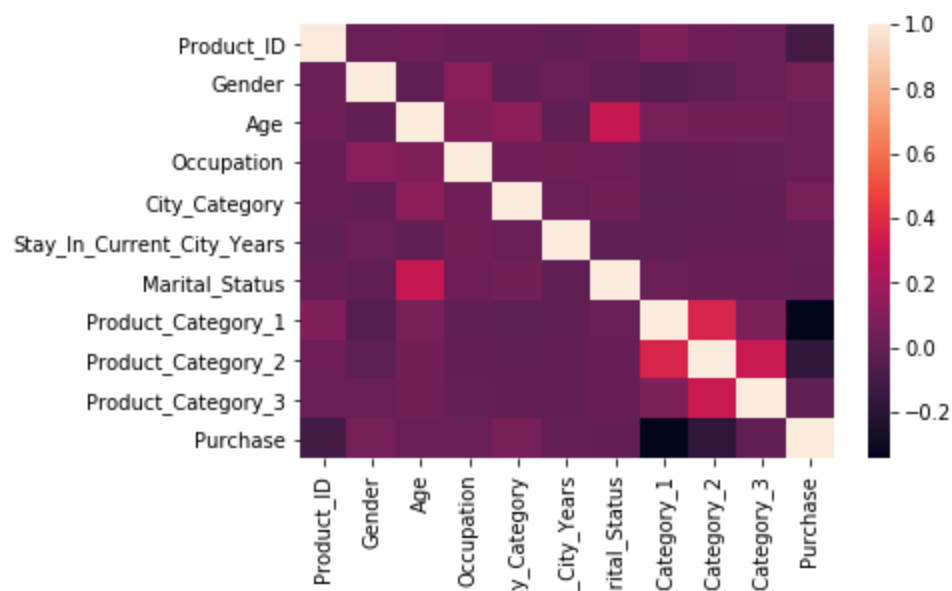
## 5. FLOWCHART



## 6. RESULT

In this paper, the Randomforest Regression algorithm is used to predict its performance, and compared with another two machine learning methods namely the decision tree, the Linear regression. The obtained results are displayed in Table below. The results show that, the performance of Regression have better performance than linear and decision tree. The Randomforest Regression is best with an accuracy of 70% higher than decision tree with an accuracy of 64%.

The given are the heatmap of the dataset represents the correlation between attributes and the boxplot of each attribute





TEAM 10

HOME PREDICTION ABOUT US

# CONSUMER PURCHASE BEHAVIOR ANALYSIS

Consumer purchase behavior analysis is the study of how people make purchasing decisions

Product\_ID

Your Product\_ID ..

Gender

-select\_gender-

Age

Enter Age..

Occupation

Enter Occupation..

City\_Category

Stay\_in\_current\_city\_year

Enter Year..

Marital\_Status

-select Marital\_Status-

Product\_category1

Product\_category2

Product\_category3

Submit

customer purchase is : 10925.126050253135

## ADVANTAGES:

- Helps in Reorientation of Packaging:
- Segmentation of Market is Helped:
- Helps in Product Orientation:

- Helps Consumers to Study their Behaviour:
- It is a feasible web blog, which can be accessed easily.
- It is composed using HTML and python for the web usage in real time.
- Helps in Formulating Right Marketing Strategy.
- It can work in real time and predicted as soon as the necessary details for prediction are given to the model.

## **Disadvantages:**

- The marketers study the behaviour of consumer to mold it in favour of their product.
- They take full advantage of weaknesses of consumers to mold it in their favour.
- It predicts the output within the range but not accurate value.
- Needs more than a single value for the prediction.

## **8. APPLICATIONS**

- In this model we can also predict in different sectors like business, bank, salaries etc...
- The most obvious is for marketing strategy.
- A second application is public policy.
- Social marketing involves getting ideas across to consumers rather than selling something.
- As a final benefit, studying consumer behavior should make us better consumers.

## **9. CONCLUSION**

In this paper, the RandomForest Regression algorithm is adopted to build a UI model for predicting consumer purchase behavior analysis and the results are compared with other two algorithms of Linear and Decision Tree. The model shows that RandomForest Regression performs best than the other two algorithms in the prediction of purchase. There is no definitive guide of which algorithms to be used. What may work on some datasets may not work on others. Therefore, always check the accuracy and predict with the dataset values.

## 10. FUTURE SCOPE

This RandomForest Regression model can also be used in the future predictions like weather forecast, job prediction, salary prediction etc. In future we can create a web application on this type of problems so that it can be accessed from everywhere with more users.

## 11. BIBLIOGRAPHY

- 📖 The science behind customer churn. [Online].  
Available:<http://financeinbusinesslife.info/thesciencebehind-customer-churn/>
- 📖 Siddique Ibrahim at. Al. A Survey on Infrequent Weighted Itemset Mining ApproachesI , IJAR CET, Vol.4, pp. 199-203 in 2015.
- 📖 Surendar Natarajan and Sountharajan Sehar proposed a idea of Distributed FP-ARMH Algorithm in Hadoop Map Reduce Framework for IEEE, in 2013.
- 📖 E. Siegel, Predictive Analytics, The power to Predict Who Will Click, Buy, Lie or Die, Wiley, 2013.
- 📖 M. Nielson, Neural Networks and Deep Learning, Online Book, 2016.
- 📖 Adcock, R. J. (1878). A problem in least squares. Analyst, 5, 53-54.
- 📖 Marquardt, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics, 11,431-441.
- 📖 Schemper. M. (1990). The explained variation in proportional hazards regression. Biometrika, 77,216-218.
- 📖 McCullagh, P., and J. A. Nelder (1989). Generalized Linear Models, 2nd ed. London: Chapman and Hall.
- 📖 Bagherpour, A. (2017),Predicting Mortgage Loan Default with Machine Learning Methods; Univ ersity of California, Riverside.
- 📖 K. Aleum and S.B. Cho, "An ensemble semi-supervised learning method for predicting defaults in social lending," Eng. Appl. Artif. Intell., vol. 81, pp. 193–199, May 2019
- 📖 Dobson. A. J. (1990). Introduction to Generalized Linear Models. London: Chapman and Hall.
- 📖 Aia, M. A., Goldsmith, R. L., and Mooney, R. W. (1961). Predicting stoichiometric CaHP04 • 2H2O.Indllstrial and Engineering Chemistry, 53, January, 55-57.
- 📖 Adcock, R. J. (1878). A problem in least squares. Analyst, 5, 53-54.

## APPENDIX

### HTML:

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <link rel="stylesheet" href="{{ url_for('static', filename='css/style.css') }}">
  <link rel="stylesheet" href="{{ url_for('static', filename='css/Form.css') }}">
```

```
<script language="JavaScript">
function showInput() {
    document.getElementById('display').innerHTML =
        document.getElementById("user_input").value;
}
</script>
```

```
<title>My Website</title>
</head>
<body>
  <!-- Header -->
  <section id="header">
    <div class="header container">
      <div class="nav-bar">
        <div class="brand">
          <a href="#hero"><h1>TEAM <span>10</span></h1></a>
```

```

</div>
<div class="nav-list">
  <div class="hamburger"><div class="bar"></div></div>
  <ul>
    <li><a href="#hero" data-after="Home">Home</a></li>
    <li><a href="#services" data-after="services">Prediction</a></li>
    <li><a href="#section-title" data-after="section-title">About
US</a></li>

  </ul>
</div>
</div>
</div>
</section>
<!-- End Header -->

```

```

<!-- Hero Section -->
<section id="hero">
  <div class="hero container">
    <div>
      <h1>Hello, <span></span></h1>
      <h1>Welcome <span></span></h1>
      <h1> TEAM 10<span></span></h1>
      <a href="#services" type="button" class="cta">Prediction</a>
    </div>
    <b> <label class="cta" style="color: antiquewhite;font-size:
30px;position: relative;top:-10rem;left:34rem;"><span>OUTPUT :</span>
</label></b>

```

```

    <p id='display' style="color:azure;font-size:20px;position:
relative;top:5rem;left:0rem;background-color: crimson ; "><b>{{y}}</b></p>

```

**</section>**

**<section id="projects">**

**<div class="container1">**

**<form action = "/login" method = "post">**

**<div class="row">**

**<div class="col-25">**

**<label>Product\_ID</label>**

**</div>**

**<div class="col-75">**

**<input type="text" name="Product\_ID" placeholder="Your  
Product\_ID .." required>**

**</div>**

**</div>**

**<div class="row">**

**<div class="col-25">**

**<label for="Gender">Gender</label>**

**</div>**

**<div class="col-75">**

**<select name="Gender">**

**<option value="">-select\_gender-</option>**

**<option value="MALE" name="Gender">MALE</option>**

**<option value="FEMALE" name="Gender">FEMALE</option>**

**</select>**

**</div>**

**</div>**

**<div class="row">**

**<div class="col-25">**

```
<label>Age</label>
</div>
<div class="col-75">
  <input type="text" name="Age" placeholder="Enter Age.." required>
</div>
</div>
<div class="row">
  <div class="col-25">
    <label>Occupation</label>
  </div>
  <div class="col-75">
    <input type="text" name="Occupation" placeholder="Enter
Occupation.. " required>
  </div>
</div>
<div class="row">
  <div class="col-25">
    <label>City_Category</label>
  </div>
  <div class="col-75">
    <input type="text" name="City_Category" placeholder="Enter
City_Category.. " required>
  </div>
</div>
<div class="row">
  <div class="col-25">
    <label>Stay_in_current_city_year</label>
  </div>
  <div class="col-75">
    <input type="text" name="Stay_in_current_city_year"
placeholder="Enter Year.. " required>
  </div>
</div>
```

```

</div>
<div class="row">
  <div class="col-25">
    <label for="Marital">Marital_Status</label>
  </div>
  <div class="col-75">
    <select name="Marital">
      <option value="">-select Marital_Status-</option>
      <option value="Single" name="Marital">Single</option>
      <option value="Married" name="Marital">Married</option>
    </select>
  </div>
</div>
<div class="row">
  <div class="col-25">
    <label>Product_category1</label>
  </div>
  <div class="col-75">
    <input type="text" name="Product_category1" placeholder=""
required >
  </div>
</div>
<div class="row">
  <div class="col-25">
    <label>Product_category2</label>
  </div>
  <div class="col-75">
    <input type="text" name="Product_category2" placeholder=""
required>
  </div>
</div>
<div class="row">

```



```

    <div class="col-25">
      <label>Product_category3</label>
    </div>
    <div class="col-75">
      <input type="text" name="Product_category3" placeholder=""
required>
    </div>
  </div>
  <br>
  <br>
  <div class="row">
    <input type="submit" value="Submit" onclick="showInput();">
    <div> <h1 id="user_input" style="color:aliceblue;font-size:
20px;background-color:
black;position:relative;top:6rem;left:0rem;border:2px solid crimson;
padding:30px ;text-align:center;"><b >{{y}}</b></h1></div>
    </div>

</form>
</section>

```


```

<section id="services">
  <div class="services container">
    <div class="service-top">
      <h1 class="section-title">Consumer <span>purchase</span>
<span>behavior</span> analysis</h1>
      <p><b>Consumer purchase behavior analysis is the study of how

```


people make purchase decisions with regard to a product, service, or organization. Studying consumer behavior would allow you to answer several questions

ABOUT US

 style="border-radius: 30px;"


IMRANULLAH KHAN

4<sup>th</sup> YEAR   
Computer science Engineering

 style="border-radius: 30px;"


JYOTHSNA MANDALA

4<sup>th</sup> YEAR   
Computer science Engineering

 style="border-radius: 30px;"

P ABHINAND

4<sup>th</sup> YEAR   
Computer science Engineering

 style="border-radius: 30px;"

P BHAVANI

4<sup>th</sup> YEAR   
Computer science Engineering

```
</div>
</div>
</section>
</section>
```

```
<!-- Footer -->
<section id="footer">
  <div class="footer container">
    <div class="brand"><h1><span>Team</span>10</h1></div>
    <h2></h2>
    <div class="social-icon">
      <div class="social-item">
        <a href="#"></
a>
      </div>
      <div class="social-item">
        <a href="#"><
/a>
      </div>
      <div class="social-item">
        <a href="#"></a>
      </div>
      <div class="social-item">
        <a href="#"></a>
      </div>
    </div>
```

```
<p>Copyright © 2020 Team 10. All rights reserved</p>
</div>
</section>
<!-- End Footer -->
<script src="./app.js"></script>
</body>
```

### APP.PY:

```
from flask import Flask , render_template,request
import pickle
model = pickle.load(open('purchase.pkl','rb'))
app = Flask(__name__)
global m2
@app.route('/') #when the browser is routed towards url execute below function
def hello_world():
    return render_template("webapp.html")
@app.route('/login',methods=["POST"])

def func2():
    Product_ID = request.form['Product_ID']
    Gender = request.form['Gender']
    if(Gender=='MALE'):
        g2=1
    if(Gender=='FEMALE'):
        g2=0
    Age = request.form['Age']
    Occupation = request.form['Occupation']
    City_Category = request.form['City_Category']
    Stay_in_current_city_year = request.form['Stay_in_current_city_year']
    Marital = request.form['Marital']
    if(Marital=='Single'):
        m2=1
    if(Marital=='Married'):
        m2=0
```

```
Product_category1 = request.form['Product_category1']
Product_category2 = request.form['Product_category2']
Product_category3 = request.form['Product_category3']
data =
[[int(Product_ID),g2,int(Age),int(Occupation),int(City_Category),float(Stay_in_current_city
_year),m2,int(Product_category1),float(Product_category2),float(Product_category3)]]
pred = model.predict(data)
print(pred[0])
return render_template("webapp.html", y = "purchase is ." + str(pred[0]))
if __name__ == '__main__':
    app.run(debug = True)
```

