

## ABSTRACT

Online shopping grows along with the growing population. An ensemble approach has been drawn for better shopping churn prediction. The algorithms used are KNN, Stacking, Random forest, XGBoost, and Logistic Regression. An accurate prediction of (90.65%) has been achieved for our ensemble approach as the best result. Customer churn refers to the number of customers who have ceased utilizing the company's product or service over a period of time.

The number of customers lost within a certain time frame divided by the number of active customers at the start of the period is one way to compute a churn rate. For example, if you gained 1000 clients last month but lost 50, the monthly churn rate would be 5%. Every month, the active customer base is fed into a Machine Learning Predictive Model, which calculates the likelihood of each client churning, will be sorted from highest to lowest probability value (or score. Clients with a low likelihood of turnover (or, in other words, customers for whom the model forecasts no churn) are satisfied customers.

# TABLE OF CONTENTS

CHAPTER	CHAPTER NAME	PAGE NUMBER
	<b>Abstract</b>	<b>iv</b>
	<b>List of Figures</b>	<b>vii</b>
	<b>List of Abbreviations</b>	<b>vii</b>
<b>1</b>	<b>INTRODUCTION</b>	<b>01</b>
	1.1 OUTLINE OF PROJECT	01
	1.2 PURPOSE OF THE PROJECT	01
	1.3 RELATED WORK	02
	1.4 SCOPE OF PROJECT	03
	1.5 PROJECT OVERVIEW	03
	1.6 TECHNIQUES USED?	04
	1.7 SYSTEM REQUIREMENTS	05
<b>2</b>	<b>LITERATURE SURVEY</b>	<b>06</b>
	2.1 DIFFERENT SURVEYS FR ONLINE PREDICTION	06
<b>3</b>	<b>METHODOLOGY</b>	<b>09</b>
	3.1 EXISTING METHODOLOGY	09
	3.2 PROPOSED METHODOLOGY	09
	3.3 SYSTEM ARCHITECTURE	10
	3.3.1 TAKING INPUT DATA	11
	3.3.2 DATA PREPROCESSING AND ACQUISITION	13
	3.3.3 VALIDATION OF DATASET	17
	3.3.4 USING DIFFERENT ML ALGORITHMS	19
	3.3.4.1 KNN	19
	3.3.4.2 RANDOM FOREST	21
	3.3.4.3 XGBOOST	24

	3.3.4.4 LOGISTIC REGRESSION	26
	3.3.4.5 STACKING	29
	3.3.5 TEST OF DATASET	32
	3.3.6 METRICS USED	32
	3.4 WORKFLOW -DIAGRAM	36
<b>4</b>	<b>RESULTS AND DISSCUSSIONS</b>	<b>37</b>
	4.1 RESULTS	37
	4.2 REQUIRED ANALYSIS	38
<b>5</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>39</b>
	5.1 CONCLUSION	39
	5.2 SCOPE OF FUTURE WORK	39
	<b>REFERENCES</b>	<b>40</b>
	<b>APPENDICIES</b>	
	<b>A. SAMPLE CODE</b>	<b>41</b>
	<b>B. SCREENSHOTS</b>	
	<b>C. PLAGIARISM REPORT</b>	

## LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO.
3.3.1	System Architecture	10
3.3.2.1	Data pre-processing and Acquisition	13
3.3.4.1.1	KNN Processing	19
3.3.4.2.1	Random Forest Processing	22
3.3.4.3.1	XGBoost Processing	24
3.3.4.4.1	Logistic Regression Processing	27
3.3.4.5.1	Stacking Process	30
3.3.4.5.2	Model training of stacking	31
3.3.6.1	Classification Graph	33
3.4.1	Work-Flow Diagram	36

## LIST OF TABLES

TABLE NO	TITLE	PAGE NO.
4.1	RESULT TABLE	37

## LIST OF GRAPHS

GRAPH NO	TITLE	PAGE NO.
4.1	TITLE COMPARSION FOR PERFORMANCE METRICS	38

## **LIST OF ABBRIVATIONS**

- 1 KNN - K NEAREST NEIGHBOURS
- 2 XGBOOST- XTREME GRADIENT BOOSTING
- 3 TP – TRUE POSITIVE
- 4 TN - TRUE NEGATIVE
- 5 FP - FALSE POSITIVE
- 6 FN – FALSE NEGATIVE

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 OUTLINE OF PROJECT**

Online Shopping is continuously growing in India. People depend on online shopping since it is easier time saving and effortless. Getting the menu at our fingertips makes it a lot time saving. An online shopping system that permits a customer to submit online orders for items and/or services from a store that serves both walk-in customers and online customers.

Today, web-based shopping is the quickest and internet promoting tool. The is vital when shopping. quantity of engineers who purchase their items on the web with the nature of online administrations and items on the web. In this way, the internet shopping climate assumes a significant part in the connection among shippers and purchasers.

ML Algorithms give incredible assets to unraveling huge, uproarious information. Machine preparing permits analysts likely comprehend buyer conduct and work on the presentation of promoting efforts. We have attempted to change the fulfillment of shoppers who shop on the web/online, also attempted to make a perfect framework for deciding shopper purchasing conduct and online store fulfillment levels. The main goal of this work is to use ML algorithms for better recommendations.

### **1.2 PURPOSE OF PROJECT**

Accurate online prediction for a product buying or not involves expert knowledge, because price usually depends on many distinctive features and factors. Typically, most significant ones are brand, Expiry date, Tenure, Membership and Average Salary of person. The main objective of the Online Shopping System is to manage the details of Shopping, Internet, Payment, Bills, Customer. It manages all the

information about Shopping, Products, Customer, Shopping. The project is totally built at administrative end and thus only the administrator is guaranteed the access based on the products purchased.

A typical online store enables the customer to browse the firm's range of products and services, view photos or images of the products, along with information about the product specifications, features and prices.

Online stores usually enable shoppers to use "search" features to find specific models, brands or items. Online customers must have access to the Internet and a valid method of payment in order to complete a transaction, such as a credit card, debit card, or a service such as PayPal. For physical products (e.g., paperback books or clothes), the e-tailer ships the products to the customer; for digital products, such as digital audio files of songs or software, the e-tailer usually sends the file to the customer over the Internet.

People in large number are doing online shopping today, and it is not only because it is convenient as one can shop from home, but also because there is an ample number of varieties available, with a high competition of prices, and also it is easy to navigate for searching regarding any particular item.

### ***1.3 RELATED WORK***

Online shopping is a form of electronic commerce which allows consumers to directly buy goods or services from a seller over the Internet using a web browser or a mobile app. Consumers find a product of interest by visiting the website of the retailer directly or by searching among alternative vendors using a shopping search engine, which displays the same product's availability and pricing at different e-retailers.

K Maheshwari, P Packia Amutha Priya has discussed, in their paper in 2018 written for Master thesis [2], that regression model that was built using Support Vector Machines (SVM). Customer buying behaviour is identified by people's personality and character. These personality characters vary from person to person. The character includes quality, motivation, occupation and income level, perception,

psychological, personality, reference groups and demographic reasons learning, beliefs, attitude, Culture and social forces. Nowadays, Data mining normally used to investigate the customer activities on shopping by using various algorithms and methods. Data mining has gradually raised and it gains numerous industries which applies this technology. Each and every activity of a customer is stored as a byte of data in a database to collect information such as how the customer spends their valuable time, day in buying decision.

Another approach by Cheng Lin in 2020 was we studied the size prediction problem and proposed a method to improve the predication accuracy. We capture the subtle semantics of customer reviews by using Bert model to learn the latent representations. Deep features are learned from the review data, combining with customers' body features and the product's information to predict the product's fitness for a customer. The experiments on the real online clothing retailer dataset validate the effectiveness of the proposed method.

## ***1.4 SCOPE OF PROJECT***

This Online prediction system can be implemented to any shop in the locality or to multinational branded shops having retail outlet chains. The system recommends a facility to accept the orders 24\*7 and a home delivery system which can make customers happy.

If shops are providing an online portal where their customers can enjoy easy shopping from anywhere, the shops won't be losing any more customers to the trending online shops such as flipcart or ebay. Since the application is available in the Smartphone it is easily accessible and always available.

## ***1.5 PROJECT OVERVIEW***

The Online Customer churn prediction form predicts whether the customer buys product from the respective website or not based on feedback or any other factors of product. The central concept of the application is to allow the customer to shop virtually using the Internet and allow customers to buy the items and articles of their desire from the store. The information pertaining to the products are stores on an RDBMS at the server side (store).



The end user of this product is a departmental store where the application is hosted on the web and the administrator maintains the database. The application which is deployed at the customer database, the details of the items are brought forward from the database for the customer view based on the selection through the menu and the database of all the products are updated at the end of each transaction. Data entry into the application can be done through various screens designed for various levels of users. Once the authorized personnel feed the relevant data into the system, several reports could be generated as per the security.

The Server process the customer to the demo page to enter the details i.e, age ,membership, has any offers, coupons , time spent on website etc things to fill it and predict the score whether they are willing to buy the product or not based on the churn score we get on the screen.

## ***1.6 TECHNIQUES USED?***

This project is an attempt to provide the advantages of online shopping to customers of a real shop. It helps buying the products in the shop anywhere through internet by using an android device by showing there the Front-End Development connecting to the backend server to store the information using Flask etc. Thus the customer will get the service of online shopping and home delivery from his favorite shop based on the tracker of the address and and may buy the product depending on the feedback of other customers. This project is a web-based shopping system for an existing shop. The project objective is to deliver the online shopping application into android platform.

Supervised machine learning is a technique whose task is to deduce a function from tagged training samples. The training samples for supervised learning consist of large set of examples for a particular topic. In supervised learning, every example training data comes in a pair of input (vector quantity) and output value (desired result). These algorithms analyze data and generate an output function, which is used to mapped new data sets to respective classes. Different algorithms used in prediction of this analysis are:-

- 1.KNN (K Nearest Neighbor)
- 2.Random Forest
- 3.XGBoost (Xtreme Gradient Boosting)
- 4.Logistic Regression
- 5.Stacking (Ensemble Method)

## **1.7 SYSTEM REQUIREMENTS**

### **HARDWARE REQUIREMENTS**

- System : Pentium Dual Core.
- Hard Disk : 120 GB.
- Monitor : 15"LED
- Input Devices : Keyboard, Mouse
- Ram : 1GB.

### **SOFTWARE REUIREMENTS**

- Operating system : Windows 7.
- Coding Language : python
- Toolkit : Jupiter Notebook
- DATABASE : EXCEL

## CHAPTER 2

### LITERATURE SURVEY

#### **2.1 DIFFERENT SURVEYS FOR ONLINE PREDICTION**

[1]Theresa Maria Rausch, Nicholas Daniel Derra, Lukas Wolf in the year 2020 Predicting online shopping cart abandonment with machine learning approaches was proposed by using the methodology of k-nearest neighbour algorithm, Decision Tree And Boosting Algorithm proposing that Thereby, we provide methodological insights to gather a comprehensive understanding of the practicability of classification methods in the context of online shopping cart abandonment prediction: our findings indicate that gradient boosting with regularization outperforms the remaining models yielding an F1-Score of 0.8569 and an AUC value of 0.8182. Nevertheless, as gradient boosting tends to be computationally infeasible, a decision tree or boosted logistic regression may be suitable alternatives, balancing the trade-off between model complexity and prediction accuracy.

[2] Cheng Lin; Qinpei Zhao; Jiangfeng Li; Weixiong Rao in the year 2020 has proposed that Size Prediction for Online Clothing Shopping with Heterogeneous Information by a methodology of Naïve Bayes Algorithm by saying in their paper, we studied the size prediction problem and proposed a method to improve the predication accuracy. We capture the subtle semantics of customer reviews by using Bert model to learn the latent representations. Deep features are learned from the review data, combining with customers' body features and the product's information to predict the product's fitness for a customer. The experiments on the real online clothing retailer dataset validate the effectiveness of the proposed method.

[3] Yuanbang Liang; Yunyu Jia; Jinglin Li; Meiyi Chen; Yifan Hu; Yinan Shi; Fei Ma in the year 2020 has proposed that Online Shop Daily Sale Prediction Using Adaptive Network-Based Fuzzy Inference System by saying that This paper proposes a method for online shop daily sale forecasting. In the method, Kalman Filter was firstly applied on the historic sale data to smooth the data. Adaptive Network-based Fuzzy Inference System (ANFIS) was then built to achieve time series forecasting. Sale histories of an online shop was used to evaluate the method. Overall performance of