Ecommerce Customer Churn Analysis and Prediction

A project report submitted to ICT Academy of Kerala

in partial fulfillment of the requirements

for the certification of

CERTIFIED SPECIALIST

IN

DATA SCIENCE & ANALYTICS

submitted by

Akshara Shaju

Aswathy G

Govind S

Kavya V

Priya Santhosh Kumar

ICT ACADEMY OF KERALA

THIRUVANANTHAPURAM, KERALA, INDIA

Nov 2022

List of Figures

Table of Contents

Abstract

E-commerce has become a major part of the retail industry, but one of the biggest challenges for e-commerce businesses is customer churn. Customer churn is when a customer stops doing business with a company, and it is a major problem because it can be costly to acquire new customers. To mitigate this issue, e-commerce businesses have turned to customer churn analysis and prediction.

Customer churn analysis involves analysing customer data to identify patterns and behaviours that may indicate a customer is about to churn. This analysis can help businesses understand why customers are leaving and what they can do to prevent it. Customer churn prediction uses machine learning algorithms to predict which customers are most likely to churn in the future. This prediction can help businesses take proactive measures to retain those customers.

There are several methods for conducting customer churn analysis and prediction, including analysing purchase behaviour, customer engagement, and demographics. Machine learning algorithms such as decision trees, logistic regression, and neural networks can be used for prediction.

The benefits of customer churn analysis and prediction for e-commerce businesses are numerous. By identifying and retaining high-value customers, businesses can increase revenue and reduce costs associated with customer acquisition.

Additionally, businesses can use the insights gained from analysis and prediction to improve the customer experience, which can lead to increased customer loyalty and retention.

In conclusion, customer churn analysis and prediction are important tools for e-commerce businesses looking to reduce customer churn and increase revenue. By analysing customer data and using machine learning algorithms, businesses can gain insights into customer behaviour and take proactive measures to retain valuable customers.

1. Problem Definition

1.1 Overview

Customer churn refers to the phenomenon of customers ceasing to do business with a company. In the context of e-commerce, customer churn can have a significant impact on the success of a business, as it can lead to decreased revenue and increased customer acquisition costs.

E-commerce businesses often use various marketing strategies to attract new customers and retain existing ones. However, despite their efforts, some customers still churn, which can be due to various reasons such as poor customer experience, lack of trust, pricing, or competition. Understanding the reasons behind customer churn and identifying customers who are at risk of churning can help e-commerce businesses take proactive measures to retain customers and reduce churn.

One approach to addressing this problem is through the use of machine learning and data analytics. By analysing customer behaviour data, such as purchase history, browsing history, and demographic information, businesses can identify patterns that can predict which customers are likely to churn. They can then use this information to implement targeted retention strategies, such as offering personalised discounts or improving customer service, to keep customers engaged and loyal.

1.2 Problem Statement

The issue of losing customers who stop doing business with an ECommerce company is referred to as the ECommerce customer churn problem. Customer churn is a major concern for ECommerce businesses because it can lead to decreased revenue and profits. The problem statement for ECommerce customer churn typically entails analysing customer behaviour, identifying factors that contribute to churn, and developing retention strategies. The goal is to predict which customers are likely to churn and take proactive steps to prevent it. Data analytics techniques can be used to analyse customer data such as purchase history, demographic information, and engagement metrics to address this issue.

2. Introduction

With the increasing popularity of e-commerce and the rapid development,the competition among e-commerce companies becomes fiercer .As known, the most significant matter is to retain customers by providing the best services along with a suitable price. Therefore, with the rapid increase of E-commerce transaction volume and intense competition in the market to meet the high demand from customers, it is necessary to attract customers through customised services and targeted strategies to increase customer loyalty. On the other hand, E -commerce customer churn shows nonlinear changes and asymmetrical customer categories. E-commerce customer churn data has a typical sample imbalance which means that the number of churn samples is significantly larger than the number of non-churned samples or vice versa. This project proposes the methods or models imposed on e-commerce to proactively reduce customer churn.

In response to the above problem, existing research on customer churn mainly includes predictions related to Exploratory Data Analysis(EDA), predictions based on statistical learning theory, predictions based on classifiers. This project proposes research on E-commerce customer churn prediction based on customer segmentation, using improved methods for data balance, and then using different machine learning algorithms for prediction. Finally, predictors importance is identified to help decision makers in choosing the proper decisions on behalf of the business organisation.

3. Literature Survey

Establishment of E-Commerce Customer Churn Prediction Model Customer churn refers to the fact that the original customers of an enterprise stop to purchase enterprise goods or accept enterprise services, and instead accept the services of competitors (Wu et al., 2017).

Churn rate prediction is applied extensively in the telecommunication sector. Ecommerce customer churn is a kind of churn that customers leave the enterprise, products or services for some reasons such as low quality or delay in delivery. E-commerce customer churn is a kind of customer churn in a non-contractual relationship scenario. In a non-contractual relationship, even if the termination of this kind of business-customer relationship occurs, it is difficult for the business to detect it in advance (Shao, 2016).

For e-commerce companies, it is important to be able to accurately predict the high-value customer groups that are about to churn, and at the same time to study the purchasing habits of customers who have not churned in order to retain this type of customer group. The value of e-commerce customer churn prediction is to merge Ecommerce customer data over some time and establish e-commerce customer churn prediction models by analysing customer purchase behaviours (Zhang, 2015). Then, provide e-commerce customer churn retention measures to reduce customer churn and identify high-value non-churn e-commerce customers and do a respectable job in customer retention. According to the research of Shao (2016), the remaining customers do not need high cost as the new customers want to bring high profit in ecommerce. Comparatively, the customer purchase behaviour differs for both existing and new customers; however, it is essential to identify the reasons leading to the customer's loss.

In support, Lu et al., (2018) stated that in the e-commerce sector, it is extremely important to analyse the loss of customers, predict the customers who might be lost, and then take corresponding measures to retain these customers and avoid their loss. At present, most e-commerce companies have conducted an in-depth analysis of customer basic characteristic information and transaction behaviour data, and then use various methods and technologies to establish and study customer churn prediction models, and finally use this to predict customer churn (Huang, 2018).

4. Project Description

This project will go through several steps to build a customer churn prediction model. First, the dataset will go through preprocessing where the dataset is cleaned and to get best performance during modelling. After that, we will go through data visualisation to get some insights about the data set in addition to the common aspects of churned customers. Finally, machine learning algorithms will be utilised to build customer churn prediction models.

4.1 Data Description

The dataset belongs to a leading E-commerce platform.It is a historical data containing customer details and experience and its outcome is customer churn flag (churn = 1, no churn = 0). The dataset shows more than 5000 customers and their interaction and preferences in the platform. The effectiveness of this data is that it contains some specific detailed attributes which will help in customer segmentation such as: preferred login device, Satisfaction score and other attributes. These attributes will help us in studying the causes of churn in each customer's segment to identify the triggers leading to customer churn.

Sl No.	Attribute	Description	Data Type
1	CustomerID	Unique customer ID.	Numeric
2	Churn	Churn flag.	Numeric
3	Tenure	Tenure of customer in organisation.	Numeric
4	PreferredLoginDevice	Preferred login device of customer.	Character
5	CityTier	City tier.	Numeric
6	WarehouseToHome	Distance in between warehouse to home of customer.	Numeric
7	PreferredPaymentMode	Preferred payment method of customer.	Character
8	Gender	Gender of customer.	Character
9	HourSpendOnApp	Number of hours spend on mobile application or website.	Numeric
10	NumberOfDeviceRegistered	Total number of devices registered per customer.	Numeric
11	PreferedOrderCat	Preferred order category of customer in last month.	Character
12	SatisfactionScore	Satisfactory score of customers on service.	Numeric
13	MaritalStatus	Marital status of customer.	Character
14	NumberOfAddress	Total number of added addresses per each customer.	Numeric
15	Complain	Any complaint has been raised in last month.	Numeric
16	OrderAmountHikeFromlast Year	Percentage increases in order from last year.	Numeric

17	CouponUsed	Total number of coupons has been used in last month.	Numeric
18	OrderCount	Total number of orders placed in last month.	Numeric
19	DaySinceLastOrder	Day Since last order by customer.	Numeric
20	CashbackAmount	Average cashback in last month.	Numeric

Fig 4.1. Table of Data Attributes Description.

4.2 Exploratory Data Analysis

Exploratory data analysis (EDA) is done to analyse and investigate data sets and summarise their main characteristics, often employing data visualisation methods. It helps determine how best to manipulate data sources to get the answers you need, making it easier to discover patterns, spot anomalies, test a hypothesis, or check assumptions.

DISCRETE DATAS	CONTINUOUS DATAS	CATEGORICAL DATAS	
		NUMERICAL TYPE	CHARACTER TYPE
Number of Device Registered.	Tenure.	Customer ID.	Preferred Login Device.
Satisfaction Score.	Warehouse to Home Distance.	Churn.	Preferred Payment Mode.
Number of Address.	Hour Spend on App.	City Tier.	Gender.
Coupon Used.	Order Amount Hike from Last Year.	Complain.	Prefered Order Category.
Order Count.	Cashback Amount.		Marital Status.
Day Since Last Order.			

Fig 4.2. Nature of Feature Variables.

4.2.1 Univariate Analysis

Uni means one and variate means variable. The objective of univariate analysis is to derive the data, define and summarise it, and analyse the pattern present in it. In a dataset, it explores each variable separately. It is possible for two kinds of variables - Categorical and Numerical.

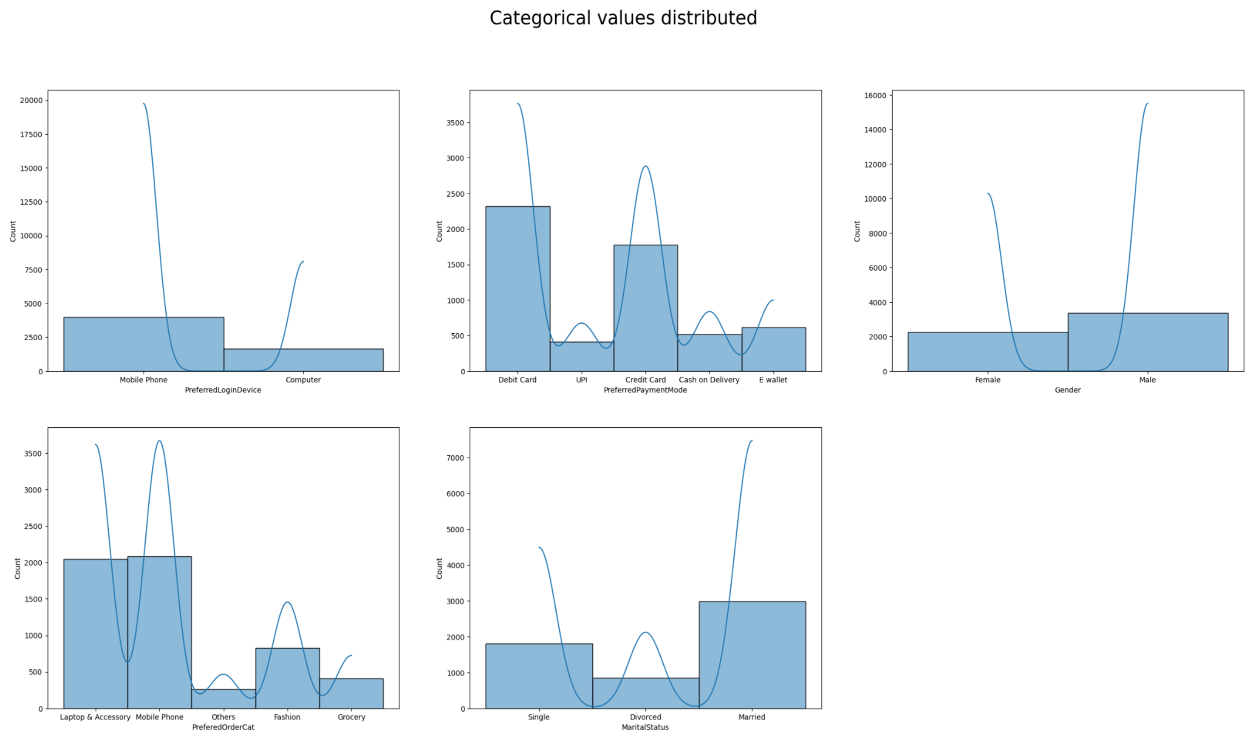


Fig 4.3. Distribution Plot for Categorical Values.

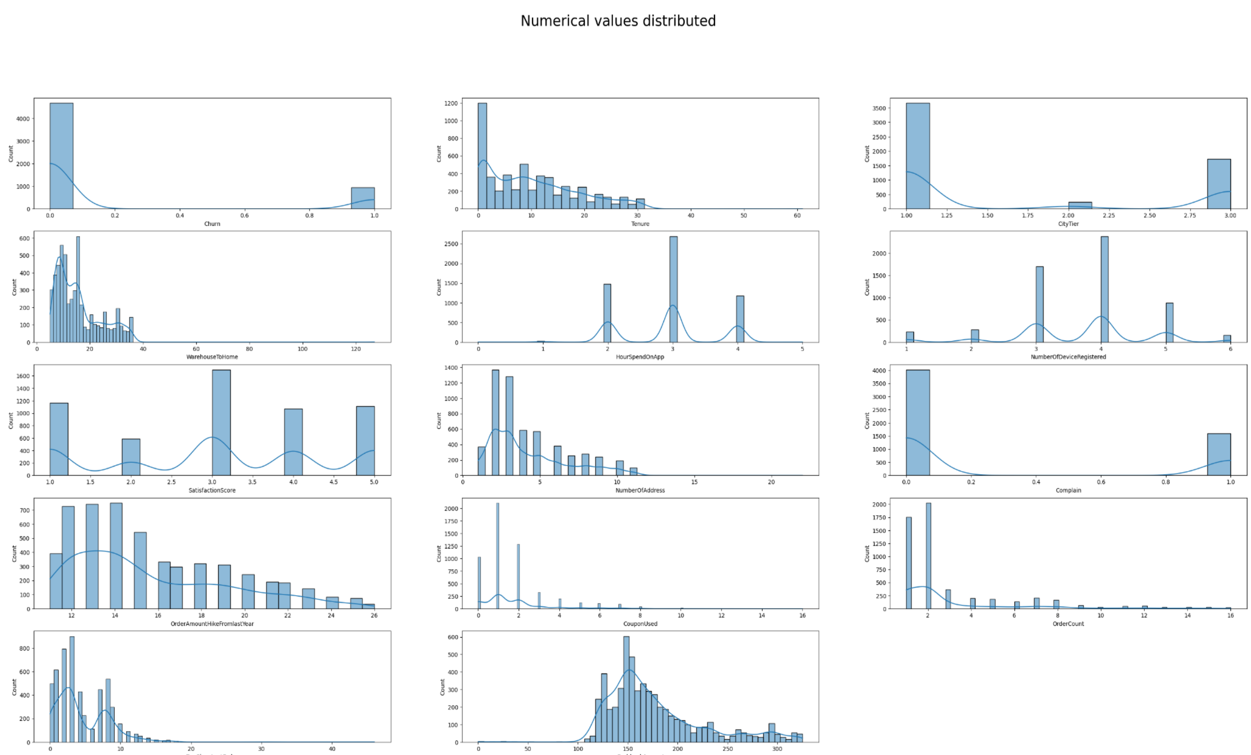


Fig 4.4. Distribution Plot for Numerical Values.

4.2.2 Bivariate Analysis

Bi means two and variate means variable, so here there are two variables. The analysis is related to cause and the relationship between the two variables.Here each feature variable is analysed against the target variable which is Churn.

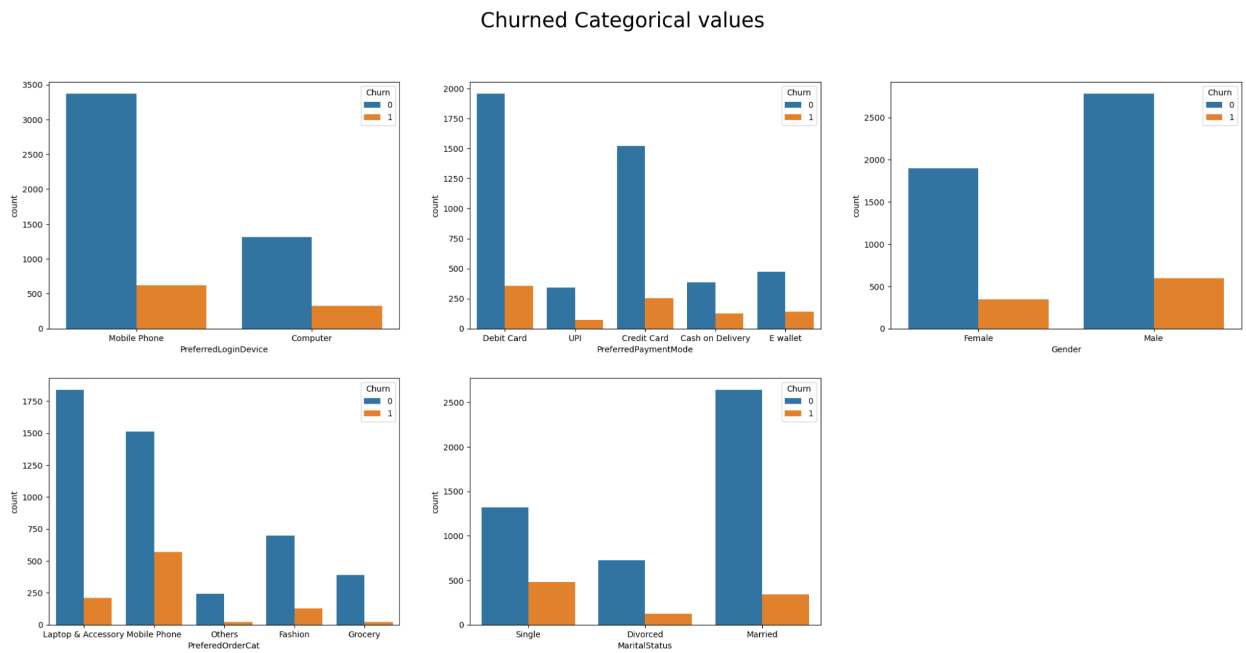


Fig 4.5. Bar Plot for Categorical Values wrt Churn Flag.

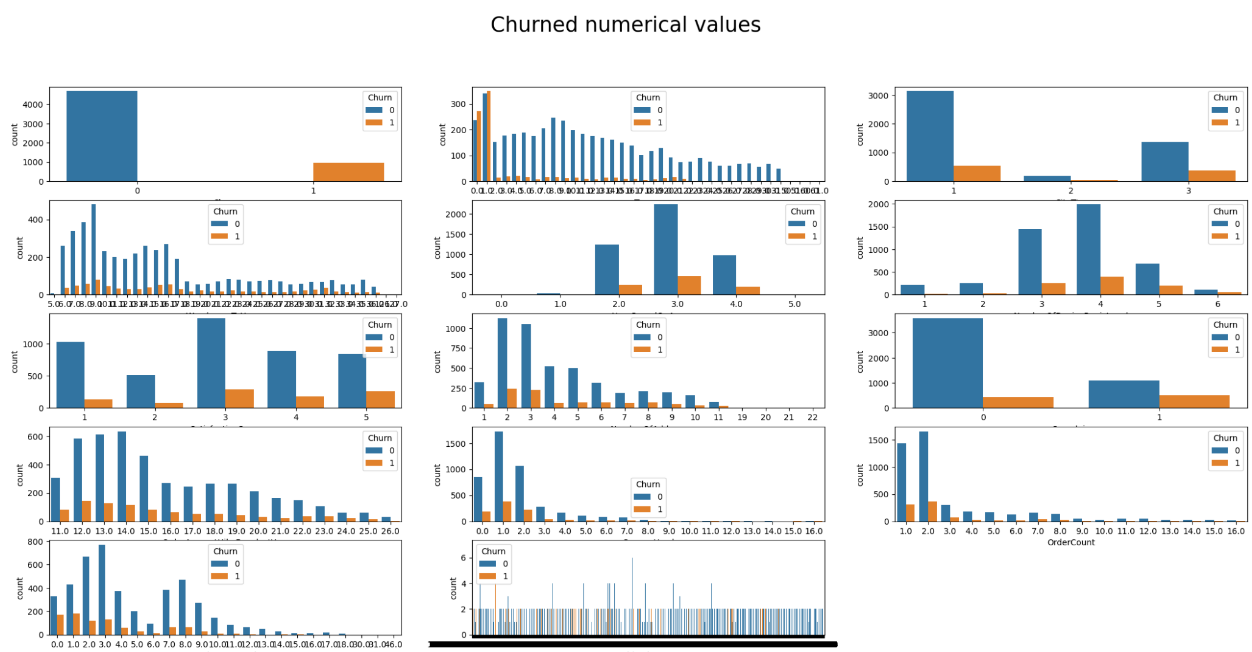


Fig 4.6. Bar Plot for Numerical Values wrt Churn Flag.

4.3 Balancing the Data using SMOTE method

Imbalanced classification involves developing predictive models on classification datasets that have a severe class imbalance.The challenge of working with imbalanced datasets is that most machine learning techniques will ignore, and in turn have poor performance on, the minority class, although typically it is performance on the minority class that is most important.

One approach to addressing imbalanced datasets is to oversample the minority class. The simplest approach involves duplicating examples in the minority class, although these examples don't add any new information to the model. Instead, new examples can be synthesised from the existing examples. This is a type of data augmentation for the minority class and is referred to as the Synthetic Minority Oversampling Technique, or SMOTE for short.

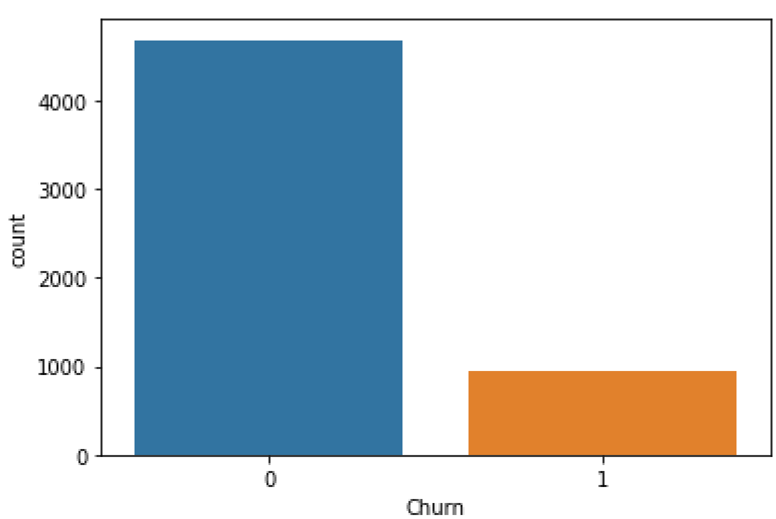


Fig 4.7. Bar Plot on Imbalanced Churn Flag data.

On plotting the countplot for the target variable ’Churn’ it is clear that the dataset is clearly imbalanced.After performing an oversampling on minority samples using SMOTE the dataset is perfectly balanced as below:

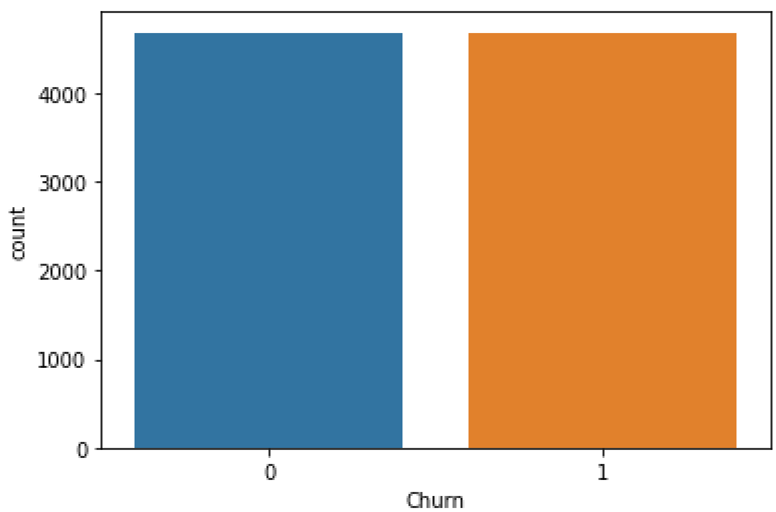


Fig 4.8. Bar Plot on Balanced Churn Flag data.

5. Data Preprocessing

5.1 Missing Value Detection and Handling

Missing Data can occur when no information is provided for one or more items or for a whole unit. Missing Data is a very big problem in real-life scenarios. Missing Data can also refer to as NA(Not Available) values in pandas. In DataFrame sometimes many datasets simply arrive with missing data, either because it exists and was not collected or it never existed.

In Pandas missing data is represented by two value:

- None: None is a Python singleton object that is often used for missing data in Python code.
- NaN : NaN (an acronym for Not a Number), is a special floating-point value recognized by all systems that use the standard IEEE floating-point representation.

In order to check null values in Pandas DataFrame, isna() function is used that can return a dataframe of Boolean values which are True for NaN values.

```
CustomerID                       0
Churn                            0
Tenure                         264
PreferredLoginDevice             0
CityTier                         0
WarehouseToHome                251
PreferredPaymentMode             0
Gender                           0
HourSpendOnApp                 255
NumberOfDeviceRegistered         0
PreferedOrderCat                 0
SatisfactionScore                0
MaritalStatus                    0
NumberOfAddress                  0
Complain                         0
OrderAmountHikeFromlastYear    265
CouponUsed                     256
OrderCount                     258
DaySinceLastOrder              307
CashbackAmount                   0
dtype: int64
```

Fig 5.1. Number of Missing Values in our Data.

Given that there are missing values in the numerical columns of the dataset, they can be replaced in this case by the median value due to the skewed nature of the numerical columns.

5.2 Encoding

- Encoding is an important step in data preprocessing, particularly when working with categorical data. Categorical data consists of values that are not numerical, columns which are categorical in the dataset are PreferredLoginDevice,
  PreferredPaymentMode,PreferedOrderCat,MaritalStatus,Gender . In order to use this data in machine learning models, it needs to be transformed into numerical values.There are two main types of encoding techniques used in data preprocessing:

1. Label Encoding: This technique involves assigning a unique integer value to each category. For example, if you have a categorical variable called "color" with categories "red", "green", and "blue", you could assign the values 0, 1, and 2 to these categories. This technique can be easily implemented using the LabelEncoder class from the scikit-learn library in Python.
2. One-Hot Encoding: This technique involves creating a new binary feature for each category in a categorical variable. For example, if you have a categorical variable called "color" with categories "red", "green", and "blue", you could create three new binary features called "is_red", "is_green", and "is_blue". Each feature would take on the value 1 if the original category matches the feature name, and 0 otherwise. This technique can be implemented using the OneHotEncoder class from the scikit-learn library in Python.

Using the one-hot encoding can increase the readability of the data by creating a special characteristic for each category. It may be simpler to decide which categories are most important for predicting the outcome variable as a result.

5.3 Scaling

Scaling is an important step in machine learning, particularly for algorithms that are sensitive to the scale of the input features. Scaling refers to the process of transforming the values of the input features to a standardized range.

There are several reasons why scaling is important:

1. Improves Performance: Some machine learning algorithms, such as k-nearest neighbors and support vector machines, are sensitive to the scale of the input features. Scaling can improve the performance of these algorithms by ensuring that each feature contributes equally to the model.
2. Speeds up Training: Scaling can also speed up the training of machine learning algorithms. Many optimization algorithms converge faster when the input features are on the same scale, which can result in faster training times.
3. Facilitates Interpretability: Scaling can make it easier to interpret the coefficients of a model. When the input features are on the same scale, the coefficients represent the relative importance of each feature in predicting the outcome variable.

There are several scaling techniques that can be used in machine learning:

1. Standardization: This technique transforms the input features so that they

have zero mean and unit variance. This is done by subtracting the mean of each feature and dividing by its standard deviation.

2. Min-Max Scaling: This technique transforms the input features so that they have a range between 0 and 1. This is done by subtracting the minimum value of each feature and dividing by the range (i.e., the difference between the maximum and minimum values).

5.4 Handling imbalanced dataset

Imbalanced data is a common problem in machine learning, where the target variable is significantly skewed towards one class. In such cases, traditional machine learning algorithms tend to perform poorly as they tend to prioritize the majority class and ignore the minority class. SMOTE (Synthetic Minority Over-sampling Technique) is an oversampling technique that can be used to address this issue.

SMOTE works by creating synthetic samples of the minority class by interpolating new observations between existing minority class observations. This is done by selecting one minority class observation at random and finding its k nearest neighbors. A new observation is then generated by randomly selecting one of the k neighbors and interpolating it with the original observation. The process is repeated until the desired number of new minority class observations is generated.

In this dataset only the balancing of the target column is done , which is the churn column.

5.5 Outliers

In statistics, an outlier is an observation that is significantly different from other observations in a dataset. Outliers can occur for a variety of reasons, including measurement error, data entry errors, or genuine differences in the underlying data generating process.

Outliers can have a large impact on statistical analyses, such as mean, variance, and correlation, and can lead to biased estimates and incorrect inferences. Therefore, it is important to detect and handle outliers appropriately.

There are several methods for identifying outliers, including visual inspection of data plots, statistical tests, and machine learning algorithms.

Some common statistical methods for identifying outliers include:

1. Z-score: This method involves calculating the number of standard deviations an observation is away from the mean of the dataset. Observations that fall outside a certain threshold (usually 2-3 standard deviations) are considered outliers.
2. Interquartile range (IQR): This method involves calculating the difference between the 75th and 25th percentile of the dataset. Observations that fall outside a certain threshold (usually 1.5 times the IQR) are considered outliers.
3. Boxplot: This method involves creating a graphical representation of the distribution of the dataset. Observations that fall outside the whiskers of the boxplot (usually 1.5 times the IQR) are considered outliers.

Once outliers have been identified, they can be handled in a variety of ways, such as removing them from the dataset using IQR, replacing them with more reasonable values using log. The appropriate method for handling outliers depends on the specific context and goals of the analysis. In the case of this dataset we have used log to replace the outliers.

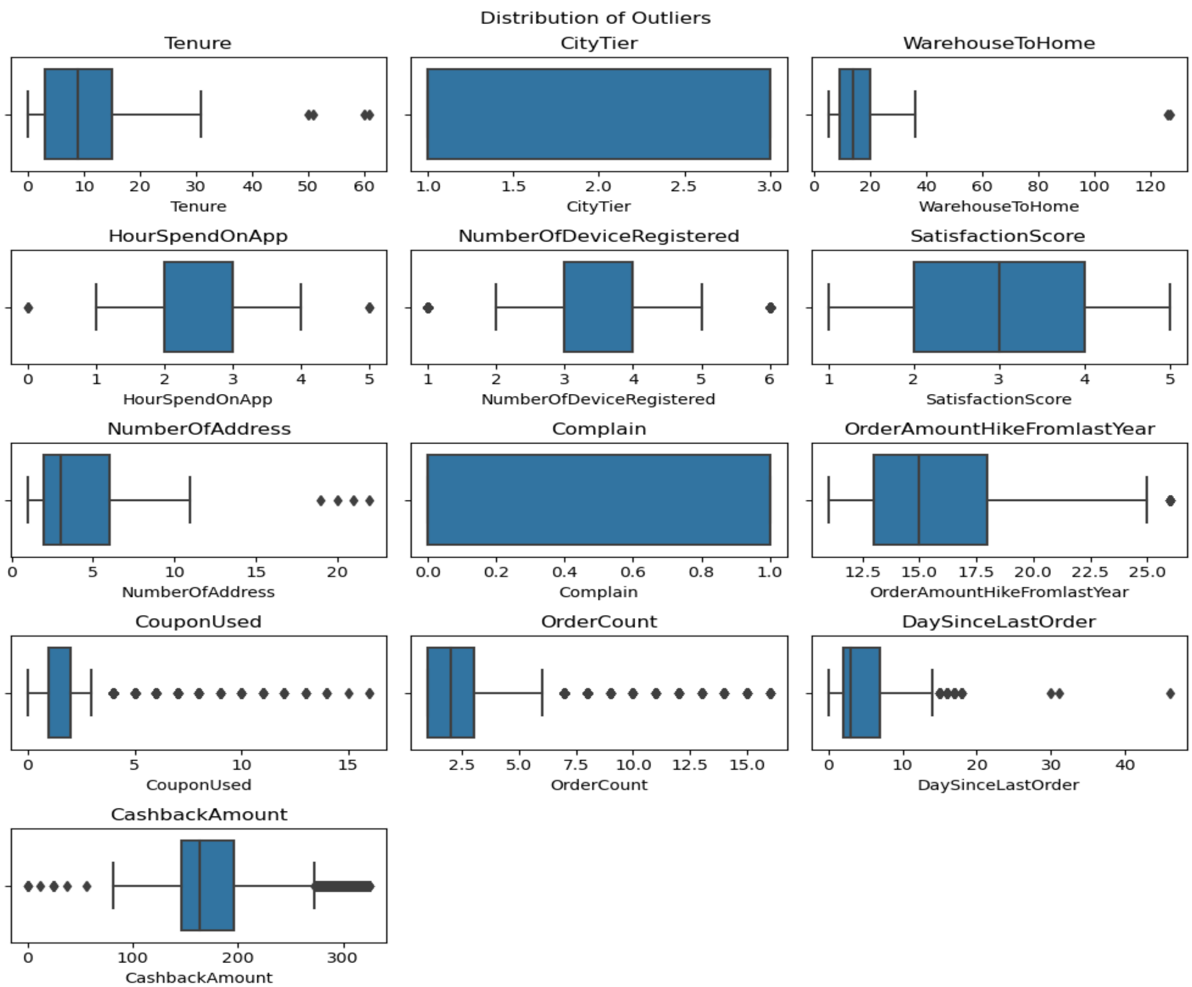


Fig 5.2. Boxplot showing Outliers.

5.6 Feature Engineering

Feature engineering is the pre-processing step of machine learning, which is used to transform raw data into features that can be used for creating a predictive model using Machine learning or statistical Modelling. Feature engineering in machine learning aims to improve the performance of models. Feature engineering extracts features from raw data. It helps to represent an underlying problem to predictive models in a better way, which as a result, improves the accuracy of the model for unseen data. The predictive model contains predictor variables and an outcome variable, and while the feature engineering process selects the most useful predictor variables for the model.

6.Model Generation and Training

Machine learning is an artificial intelligence (AI) branch which helps computers to learn and make data-based predictions. The origins of machine learning are rooted in pattern recognition and the idea that algorithms can learn without being trained to do so from historical data. Predictive modelling works by analysing current and past data to generate a model which can help predicting the future. In the HRM domain, these AI and ML models and future predictions can be used to make existing business more efficient. Once we've identified important features and data is ready to format, we can proceed with the project's predictive analysis section. The modelling begins when the information collected is divided into a training set and a test set. In the training dataset, algorithms are implemented and tested. The training set is assigned randomly 80% of the data and the test set is assigned 20% of the data.Since our target variable 'Churn' is binary, we will use classification algorithms.

There are various kinds of machine learning techniques available to learn from the given data which is called train data. When new or unseen data arises the learned model analyses and predicts desired class. In our project we have used the E Commerce Customer Churn Analytics data set to apply various machine learning algorithms to predict the chances of customers to churn. The machine learning algorithms for predicting the same are described :

6.1 Logistic Regression

One of the main problems in classification problems occurs when the algorithm never converges in the weight's updating, while being trained.This occurs when the classes aren't perfectly linear separable. So, to tackle binary classification problems , the Logistic Regression is one of the most used algorithms. Logistic regression is a simple but powerful classification algorithm (despite its name). This type of statistical model (also known as logit model) is often used for classification and predictive analytics. Logistic regression estimates the probability of an event occurring, such as voted or didn't vote, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. The odds ratio is one important concept in order to understand the idea behind logistic regression.The odds ratio is the probability that a certain event will occur. It can be written as:

$$\text{Odds Ratio} = \frac{P}{1 - P}$$

Where P stands for the probability of the positive event (the one that we are trying to predict).Derived from this we can define the logit function. Logit Function:

$$\text{Logit}(P) = \log \frac{P}{1 - P}$$

The logit function is simply the logarithm of the odds ratio (log-odds). This function takes input values in the range [0,1] and transforms them to values over the entire real-number range [-∞,∞].We will use it to express linear relationships between feature values and the log-odds.

Where P(y=1|x) is the conditional probability that a particular sample belongs to class 1 given its features x.

$$\text{logit}(P(y=1|x)) = W_0X_0 + \ldots + W_mW_m = \text{Sum}(W_iX_i) = W^TX$$

6.2 K-Nearest Neighbour

k-NN classifier is known as lazy learner in the machine learning community.They receive this name because they do not learn how to discriminate the dataset with an optimised function, but memorise the dataset instead. It never learns from the data and does not build any models. Rather, it finds out the examples from the train dataset which are closest to the unknown example. Based on the neighbour examples it will predict the new example. The value of 'k' determines the no. of closest data points or examples to be selected from the training example.The name 'lazy algorithm' also refers to the kind of algorithms called nonparametric. These are instance-based algorithms, they are characterised by memorising the training dataset, and lazy learning is a specific case of these algorithms, associated with zero computational cost during the learning. The overall process that the algorithm follows is:

- Choosing the number of k and the distance metric.
- Finding the k nearest neighbour of the sample to classify.
- Assigning the class label by majority vote.

The algorithm finds those k samples that are closest to the point to classify, basing its predictions on the distance metric.The main advantage is that as it is a memory-based algorithm, it adapts to new training data. The down-side is that the computational cost increases linearly with the size of the training data.

6.3 Support Vector Machine

A Support Vector Machine(SVM) is a kind of classification technique, where the data points are separated by a line in case of linear SVM, and a hyperplane in case of non-linear SVM. The separation is chosen in such a way that; the two sides of the hyperplane categorise the data set into two classes. When an unknown data comes it predicts which side/class it belongs to.In SVM, The optimization objective is to set a decision line that separates the classes by maximising the margin between this line and the sample points that are closest to this hyperplane. These points are called support vectors. The margin between the hyperplane and the support vectors are as large as possible to reduce the error in classification.

6.4 Decision Tree

Decision Tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions.

Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attribute.

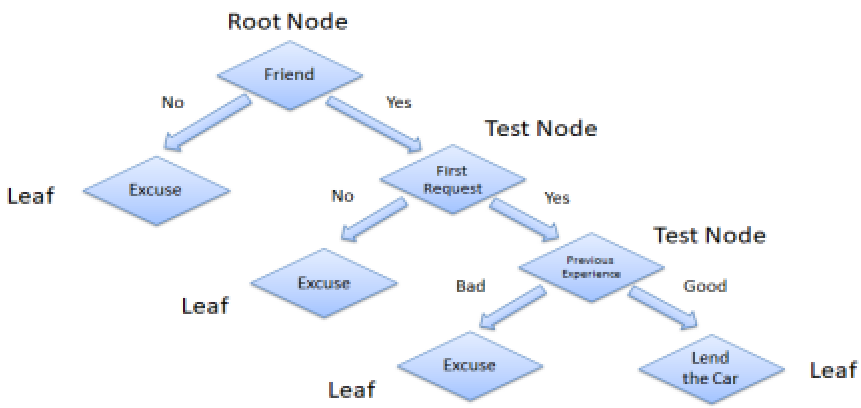


Fig 6.1 Decision tree analysis on an example of lending a car

Based on the features of the training set, the decision tree learns a series of questions to infer the class labels of the samples.The starting node is called the tree root, and the algorithm will split the dataset on the feature that contains the maximum Information Gain iteratively, until the leaves (the final nodes) are pure.

6.5 Random Forest

Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees. If we have a dataset with a lot of features (columns), the decision tree Algorithm usually tends to overfit, overcomplicating the model and the learning process.We can solve this issue by selecting each column randomly and making decision trees for each batch of columns. The main advantage of this method is that we usually won't need to prune the random forest (since the model is quite robust to noise). However, they are much less interpretable than decision trees.

The algorithm will perform the following steps:

● Drawing of a random bootstrap sample of size n.

● Growing a decision tree from the bootstrap sample. At each node, there will be randomly selected d features without replacement and the node will be splitted maximising the information gain.

● The previous process will be repeated k times.

● Aggregating the prediction done by each tree, assigning the class label by majority vote.

Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap. While classifying all the trees in the random forest gives a class to an unknown example and the class having maximum votes will be assigned to the unknown example. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

On classifying our model using different algorithms the classification reports shows variations in the accuracies,precision scores,recall scores,F1 scores etc.

The accuracy scores obtained using the above mentioned algorithms are as follows:

	Accuracy
LR	83.212
KNN	92.738
DT	92.952
RFA	97.480
SVC_rbf_ac	87.441
SVC_linear_ac	83.212
SVC_poly_ac	85.006

Fig 6.2:Comparison of accuracy using various algorithms

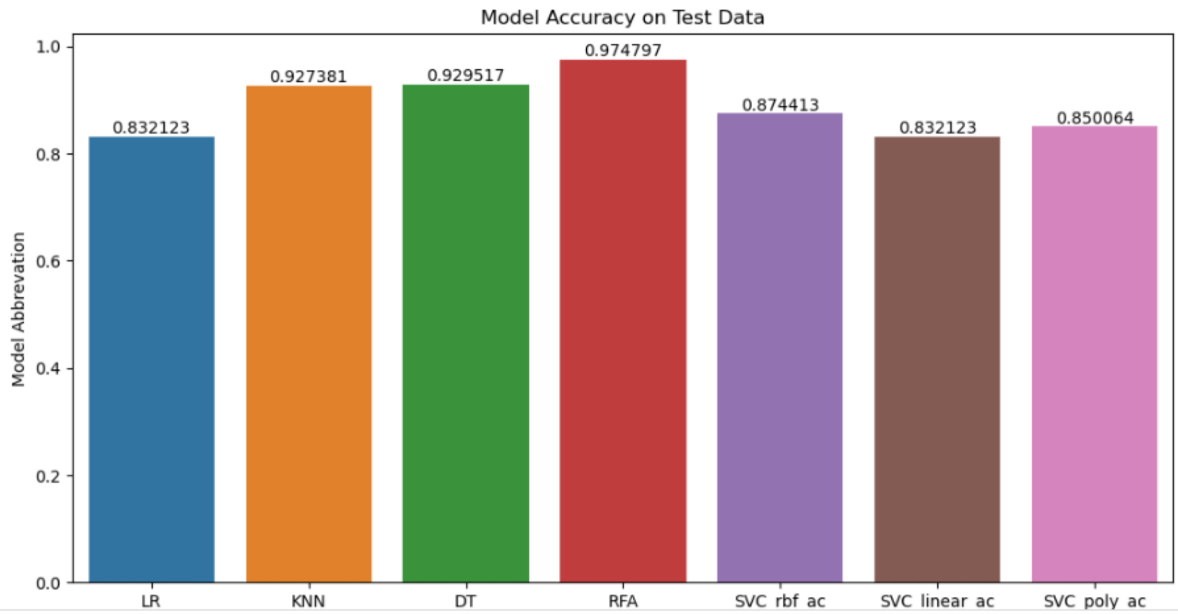


Fig 6.3 Barplot on accuracy comparison

Random Forest Algorithm delivers the best accuracy among the above.

7.Hyper Parameter Tuning

A Machine Learning model is defined as a mathematical model with a number of parameters that need to be learned from the data. By training a model with existing data, we are able to fit the model parameters. However, there is another kind of parameter, known as Hyperparameters, that cannot be directly learned from the regular training process. They are usually fixed before the actual training process begins. These parameters express important properties of the model such as its complexity or how fast it should learn. Models can have many hyperparameters and finding the best combination of parameters can be treated as a search problem. The two best strategies for Hyperparameter tuning are:

● GridSearchCV

● RandomizedSearchCV

7.1 GridSearchCV

In the GridSearchCV approach, the machine learning model is evaluated for a range of hyperparameter values. This approach is called GridSearchCV, because it searches for the best set of hyperparameters from a grid of hyperparameters values. For example, if we want to set two hyperparameters C and Alpha of the Logistic Regression Classifier model, with different sets of values. The grid search technique will construct many versions of the model with all possible combinations of hyperparameters and will return the best one.

7.2 Randomized Search CV

RandomizedSearchCV solves the drawbacks of GridSearchCV, as it goes through only a fixed number of hyperparameter settings. It moves within the grid in a random fashion to find the best set of hyperparameters. This approach reduces unnecessary computation.

● n_estimators: The number of trees in the forest. The default value is 100.

● max_features: The number of features to consider when looking for the best split:

➔ If int, then consider max_features features at each split.

➔ If float, then max_features is a fraction and max(1, int(max_features * n_features_in_)) features are considered at each split.

➔ If "auto", then max_features=sqrt(n_features).

➔ If "sqrt", then max_features=sqrt(n_features).

➔If "log2", then max_features=log2(n_features).

➔ If None, then max_features=n_features.

7.3Hyperparameter Tuning Result Summary

To determine whether we could get better results on accuracy for the model,we did hyperparameter tuning using Grid search CV and Randomized CV methods.Among the two,the accuracy score obtained on Randomized CV was good.

Hyperparameter tuning summary on Grid search CV is:

Best hyperparameters are: {'max_depth': None, 'max_features': 1, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}

Best score on Random forest Grid search CV is: 0.9798424776398345

Hyperparameter tuning summary on Randomized search CV is:

Best random search hyperparameters are: {'criterion': 'entropy', 'max_depth': 40, 'max_features': 1, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 360}

Best random search score is: 0.9823788546255506

8.Model Deployment

Model deployment is the process of implementing a fully functioning machine learning model into production where it can make predictions based on data. Users, developers, and systems then use these predictions to make practical business decisions.

Models can be deployed in multiple ways, but they’re usually integrated with apps through APIs so that all users can gain access to them.

Model deployment is the fourth stage in the model development life cycle after planning, data preparation, and model development. However, this stage is usually the most cumbersome for data scientists as it takes time and resources. Any given model typically undergoes countless modifications until it is ready to be released into a production environment. Also, some models don’t even make it to the deployment stage if they don’t meet the desired objectives.

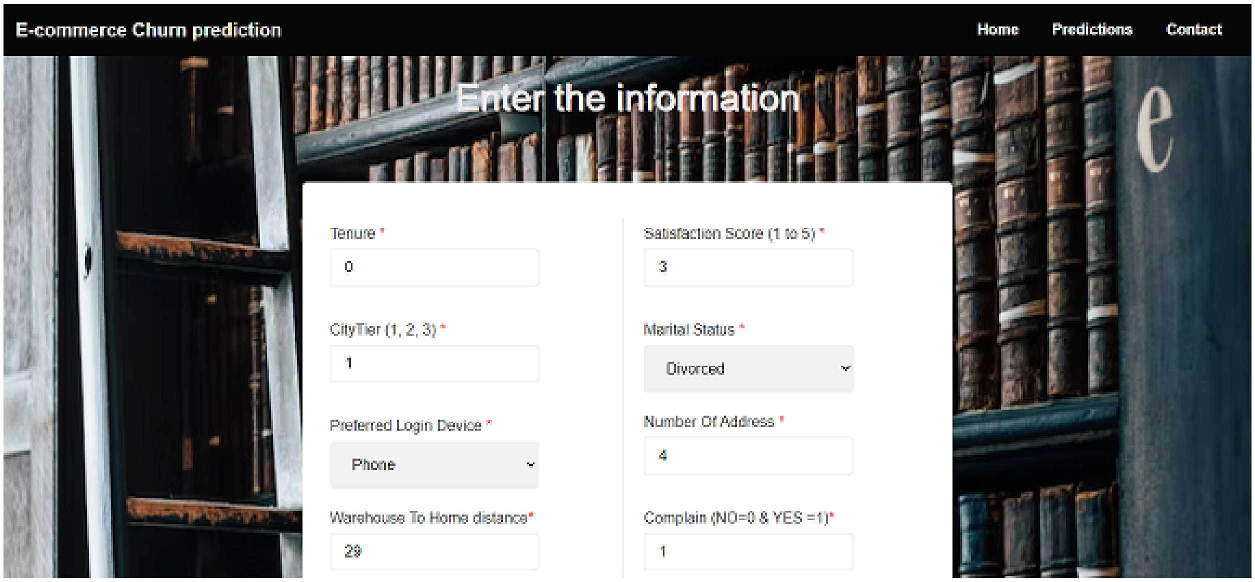
Once the validation team chooses the best operational model, it finally moves into the hands of the deployment team. The process of actually deploying the machine learning model consists of several steps:

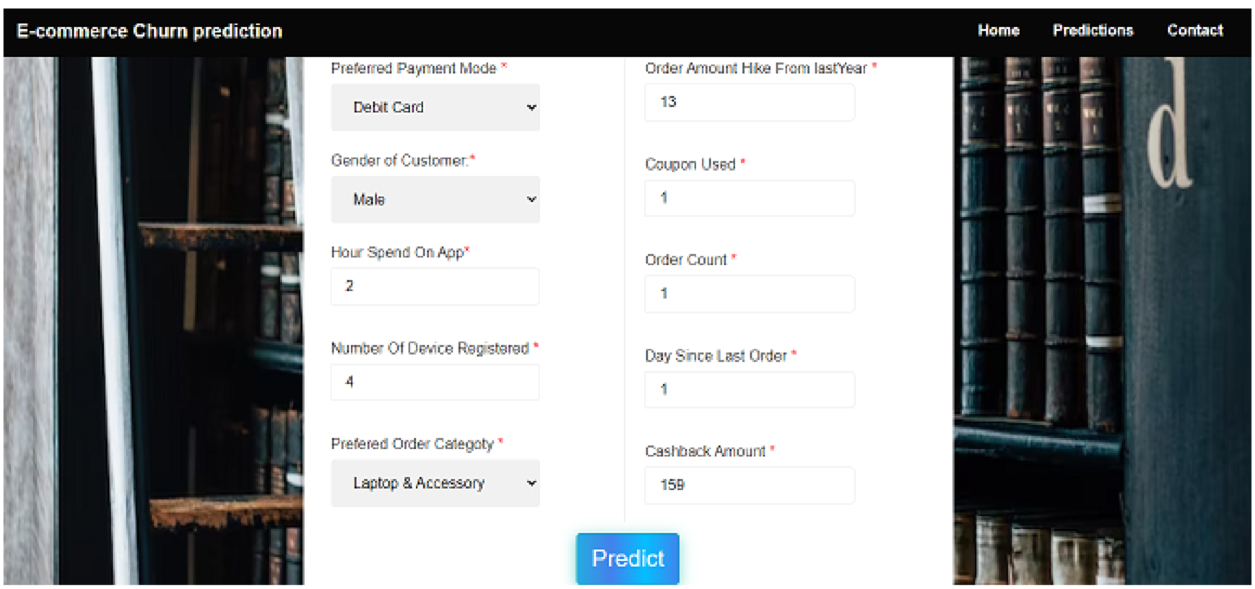
- First, The deployment team moves the model to a deployment environment, which usually includes specific servers or middleware. Here the model can gain access to the hardware and other resources it needs to draw data from.

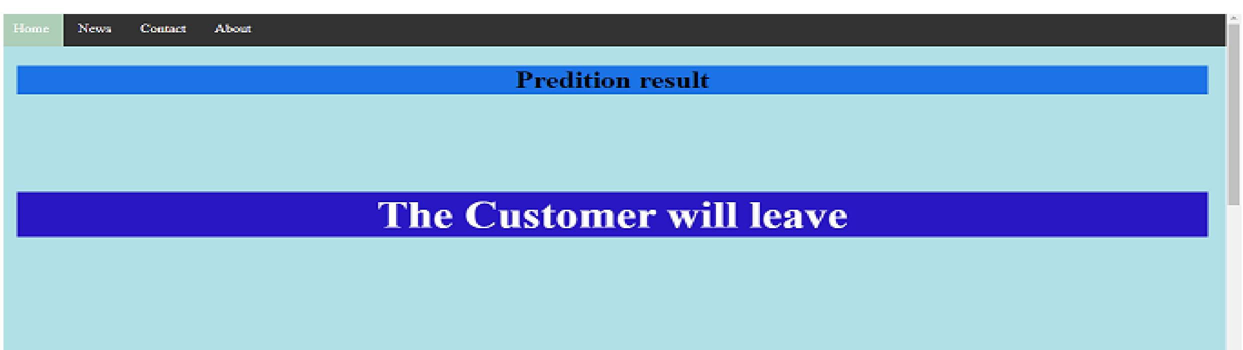
- Then integrate the model into a process. One way of doing this is by connecting it to the end user's computer via APIs or incorporating it into the user's software.

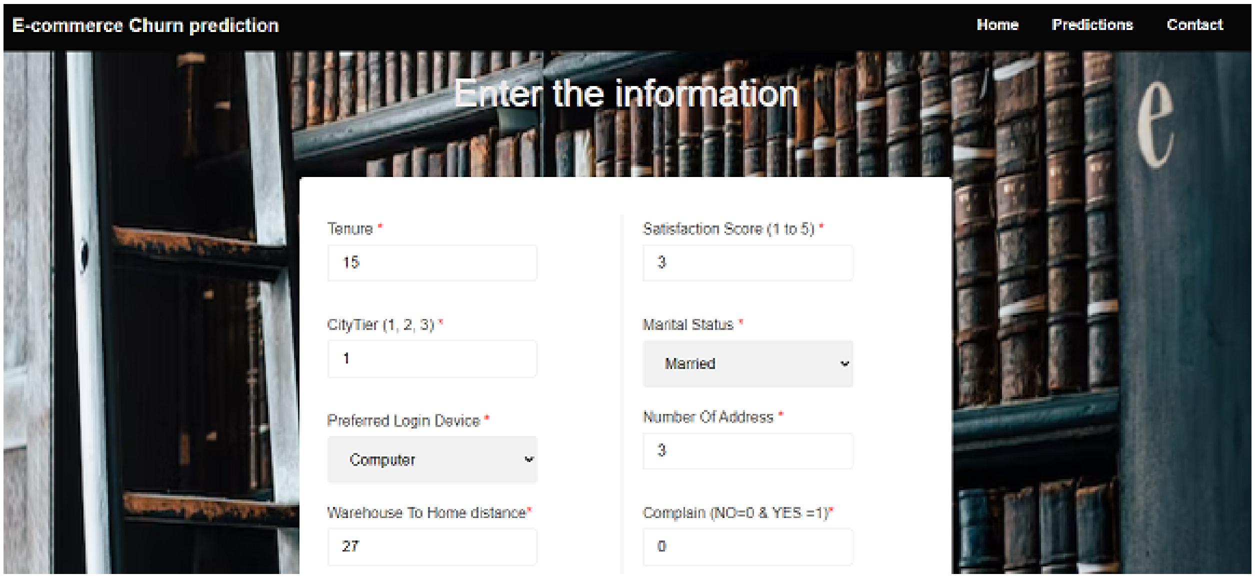
  Finally, the end users activate the model, access its data, and implement its output.Flask is a Python-based micro framework used for developing small-scale websites. Flask is very easy to make Restful APIs using python.Here ,in this project we had used Python Flask Web Framework to deploy our ML model.

8.1 Results on Model Deployment using Flask

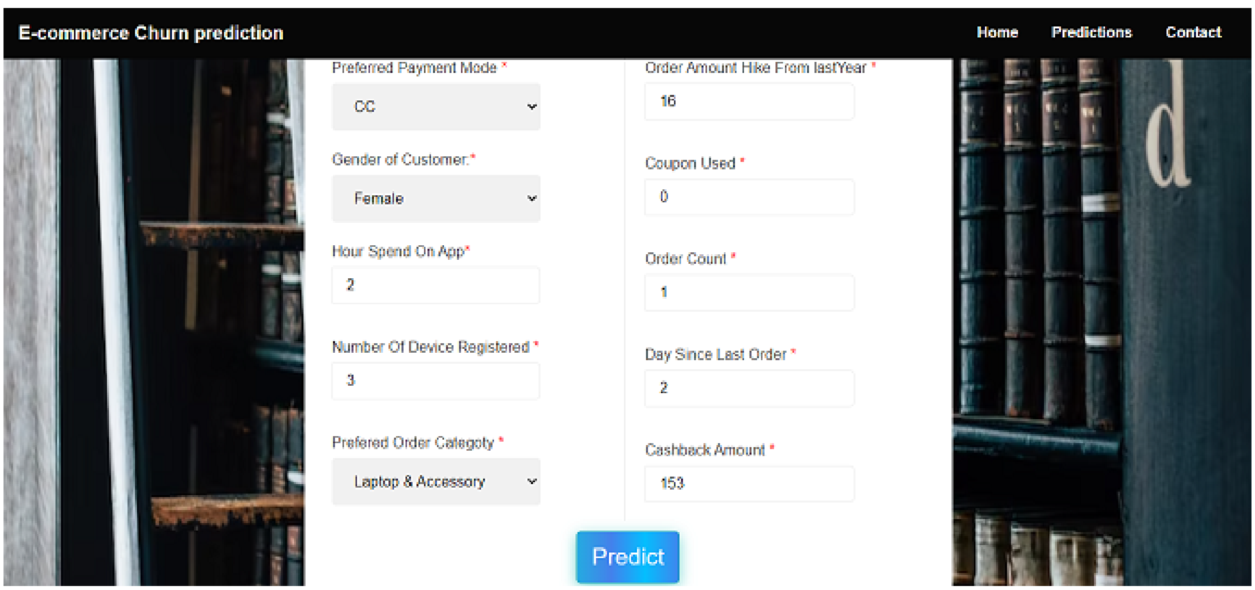




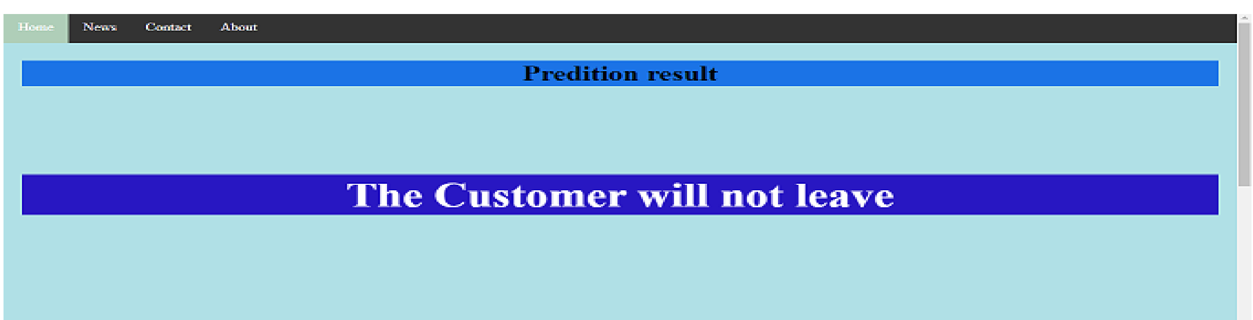




E-commerce Churn prediction
Home
Predictions
Contact
Enter the information
Tenure *
15
Satisfaction Score (1 to 5) *
3
CityTier (1, 2, 3) *
1
Marital Status *
Married
Preferred Login Device *
Computer
Number Of Address *
3
Warehouse To Home distance*
27
Complain (NO=0 & YES =1)*
0



E-commerce Churn prediction
Home
Predictions
Contact
Preferred Payment Mode *
CC
Order Amount Hike From lastYear *
16
Gender of Customer:*
Female
Coupon Used *
0
Hour Spend On App*
2
Order Count *
1
Number Of Device Registered *
3
Day Since Last Order *
2
Prefered Order Category *
Laptop & Accessory
Cashback Amount *
153
Predict



Home
News
Contact
About
Predition result
The Customer will not leave

9.CONCLUSION

In this project we applied some machine learning techniques in order to identify the factors that may contribute to a customer leaving an E-commerce company and, above all, to predict the likelihood of individual customers to churn. To conclude, Artificial Intelligence in combination with Human Resources have an effective impact on different functions of Business analytics. It will be more productive when there will be a balance between technological advancements and human applications. This project finds out which machine learning algorithm is performing well in predicting the customers, those are likely to churn from being a good customer in the company based on their demography,behaviour and review parameters.. From the experimental results, Random Forest clearly outperformed all other classifiers as obtained in evaluation criteria. This project might help the company to identify the factors causing the customers leaving the company and how they can take appropriate steps to minimise that. This project requires further exploration to minimise the prediction error rate and to improve performance of Machine Learning algorithms used. It may assist enterprises in implementing better retention strategies, decreasing turnover costs.

References

1. Zhang, D. (2015). Establishment and application of customer churn prediction model. Beijing Institute of Technology.

2. Saghir, M., Bibi, Z., Bashir, S., & Khan, F. H. (2019, January). Churn prediction using neural network-based individual and ensemble models. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 634-639). IEEE.

3. Wu, X. J., & Meng, S. S. (2017). Research on e-commerce customer churn prediction based on customer segmentation and Ada-Boost. Industrial Engineering, 20(02), 99- 107.

4. Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications–a holistic extension to the CRISP-DM model. Procedia Cirp, 79, 403-408.

5. Shao, D. (2016). Analysis and prediction of insurance company's customer loss based on BP neural network. Lanzhou University

6. Lu, N., Liu, X. W., & Lee, L. (2018). Research on customer value segmentation of online shops based on RFM. Computer Knowledge and Technology, 14(18), 275-276, 284.

7. Huang, J. (2018). A Comparative Study of Social E-Commerce and Traditional Ecommerce. Economic and Trade Practice, (23), 188-189.

8. Feng, X., Wang, C., Liu, Y., Yang, Y., & An, H. G. (2018). Research on customer churn prediction based on comment emotional tendency and neural network. Journal of China Academy of Electronics Science, 13(03), 340-345

9. Sun, J., Li, H., Fujita, H., Fu, B., & Ai, W. (2020). Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. Information Fusion, 54, 128-144.