World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Churn prediction using machine learning: A coupon optimization technique

Nadeem Ahmed [1,*] and Muhammad Umair [2]

[1] *University of Nottingham, England.*
[2] *University of Stirling, Scotland.*

World Journal of Advanced Engineering Technology and Sciences, 2024, 12(02), 332–354

Publication history: Received on 09 June 2024; revised on 16 July 2024; accepted on 19 July 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.12.2.0310

Abstract

Customer retention has been identified as one of the most crucial difficulties in every Business particularly in the grocery retail industry. In this context, an accurate forecast of whether a client will leave the organisation, also known as churn prediction, is critical for businesses to undertake successful retention strategies. High churn rates result in massive losses for corporations because keeping existing customers is more profitable than getting new ones and getting a new customer costs five times as much as keeping an old one. As a result, firms should be able to track churn rates to calculate client churn. Also, if we know which clients are going to quit before they do, we may devise preventative measures. Also, current marketing strategies such as giving coupons to customers, the majority of whom do not use them, incur significant costs for marketing and sending. Knowing which customers are not going to use that coupon will assist organisations in devising alternative strategies to retain that customer rather than sending coupons. This paper studies Dunnhumby data and proposes 2 different models one for predicting the churn and the other for coupon redemption model and both uses XGBoost Classifier Model. When both models are used together, one will predict if the customer is going to churn, and to prevent churn, we use marketing techniques such as sending coupons, so the coupon redemption model will target whether the customer will use the coupon or not, so we do not send them those coupons and propose different retention methods for these customers. This can help businesses save money by reducing churners and saving money on marketing staff and sending promotions.

Keywords: Churn rate; Machine Learning; E-commerce; CRM; Data Analytics; Coupon Optimization

1. Introduction

Customer churn prediction is a kind of Customer Relationship Management (CRM) in which a corporation attempts to create a model that forecasts if a consumer intends to leave or reduces their purchases from a company. Customer churn prediction is widely explored in a variety of areas, including financial services, subscription management, telecommunications, retail marketplaces, and electronic commerce (Chen et al., 2012). This research focuses on churn prediction in the retail industry. In the retail industry, a consumer is said to be churned when a consumer stop making purchases or reduces purchases over time from the firm for an extended period of time.

The impetus for Customer churn prediction stems from the observation made in CRM is that organizations store important data about their consumers in their records (Jones et al., 2000; Herman, 1965; Thomas, 2001). The companies have demographics data such as age, gender and location of purchase also they have various kinds of data such as purchase history and coupon redemption data collected during a particular campaigns with offers. This information may be utilised to determine whether a client is departing and what the reasons are. Because keeping existing customers is more profitable than getting new ones and getting a new customer costs five times as much as keeping an old one. (Reinartz and Kumar, 2003), businesses should try to foresee leaving clients and stop them from going away or declining

* Corresponding author: Nadeem Ahmed

purchases. Loyal customers spend more and spread positive word of mouth over time (Gremler & Brown, 1999). As a result, forecasting customer churn is critical for retail businesses progress.

1.1. Research Purpose

Businesses tend to emphasize more on making strategies to deal with consumers, with CRM as the fundamental approach for handling, maintaining, and establishing new long-term connections with customers as a crucial stakeholder (Chorianopoulos, 2015). Machine learning models will help CRM to accomplish their objectives by creating models that lead to informed decisions, creating better, stronger, and long-lasting. The main principle behind most ML-applications is to divide the dataset into test and training data. The training data is then supplied into an ML-model that learns from the data. The model is then fed previously unknown test data from which it predicts results, which are then compared to actual values. Metrics of how good the model is calculated from the discrepancies between forecasts and true data (Louridas and Ebert, 2016). As the number of data continues to grow, the use of machine learning algorithms to generate predictions has grown in popularity (Louzada, Ara and Fernandes, 2016). Future predictions might be useful since they help businesses to better prepare for the probable future (Roos and Gustafsson, 2007).

This research studies the open-source household dataset provided by Dunnhumby. This dataset comprises household-level transactions from 2,500 regular consumers at a business over a two-year period. It includes all of a household's purchases, not just those from a few categories. Demographic information and direct marketing contact histories are given for specific homes. The dataset contains 8 tables but only 6 of them are used for this research which are relevant for this paper which are Campaign Descriptions, Campaigns, Coupons, Coupon Redemptions, Transactions, and Demographics and the other two are not used because they contain information like product descriptions and emailer, which are not crucial for forecasting churn as this research focuses on the amount of purchase, but they are important for learning about consumer purchases and targeting them with particular discounts that the customer is interested in.

This study focuses mostly on Customer churn prediction and coupon redemption using machine learning to identify a model that is best suited for forecasting the churn of a retail store from Dunnhumby dataset. Also, goal here is to build a classification model that predicts whether or not a client will redeem their coupons for the remaining five campaigns of the year. Aside from knowing which consumers will redeem their coupons, it may be more interesting for a firm to determine which customers would not redeem them in order to either decide on alternate marketing and communication efforts to reach them or to not give them coupons at all and save money. Following that, the emphasis is on empirical study and creating a classification model using the available data.

The average retail churn rate is five to seven percent every year. A churn rate of less than 5 percent is ideal, but anything beyond 10 percent is a reason for concern according to industry standards (Roos and Gustafsson, 2007). Even if you acquire more consumers, your firm will not develop unless the volume of entering customers exceeds the amount of existing customers. Lowering your turnover rate by 5 percent boosts profitability by up to 125 percent.

The definition of churn varies per organization; however, it is tough in retail because individuals buy food on a weekly, fortnightly, or monthly basis. So, for this dataset, churn is maintained for two weeks. If a consumer has not purchased from a store for two weeks or longer, they are deemed churned.

1.2. Research Objectives and goals

The primary goal of this research is to use machine learning algorithm to optimise the Dunnhumby dataset to prevent customers from churning and to use coupon optimization to reduce the churning rate which is one of the marketing strategies to prevent churn.

The successful execution of this research will result in increased revenue and client retention, as well as tailored marketing to retain customers. If a certain coupon is not being utilised by a consumer, the company should reconsider its marketing strategy in order to keep existing customers and save money on not offering coupons that are not being used. Coupons such as an offer on meat, if sent to a vegetarian will be useless and it will never be used. This technique will study historic data and classify if the consumer will use that coupon or not.

1.3. Methodology Approach

1.3.1. Baseline dataset

The Research is divided into two stages first is creating a churn prediction model and the other one is coupon classification model. The first model will help to classify all the churners and the second model will help to classify all the customers which will redeem the coupon before a campaign is performed. The first section of the research covers the Exploratory Data Analysis (EDA) which describes all the tables in the dataset for example in demographics data, age and income distribution of buyer. This will help understand company which factors drive the loss or profit of the company which is common for both the models. For example, age between 19-24 is most of the churn customers or consumers with incomes lower than 20000 pounds have the most churning rate. Studying this will help company to analyze which group are churning more and specifically create marketing strategies to deal with that.

1.3.2. Churn Dataset and ML Model

The next section is all about churn prediction from creating a new dataset to classify churn It compromises of defining the churn and creation of a new dataset with all the features that will go into the ML model. Churn is defined as 2 weeks for this model, and this means that if a consumer does not shop for more than 2 weeks it is said to be churn customer. The raw dataset sometimes not contain the exact columns which are needed to go through a model such as Amount of purchase made by each household as every household might have more than one purchase in a month so Data Engineering is done which is calculation of values with already given data as a new feature. Some of the new features are created which are passed to the churn prediction model.

Once the dataset is ready with all the features, the data-pre-processing/Cleaning is necessary because raw data cannot be processed by a model. Even if it does, it will perform poorly because raw data conatins variables not in proper data type, missing values, outliers, multicollinearity, imbalanced classes, scaling, and cardinality issues. So, Data cleaning is done such as removing nulls or outliers if present, converting their data types and most importantly converting the categorical features to numerical values (label encoder) and machine learning models does not read categorical values. Removing the multicollinearity by dropping the columns which are highly co-related. Scaling the data with power transform method. Next is exploration of all features with output variable which is churn which means visualization of all the features with the defined churn for each household which will help us know which common factor is producing the churn for example people with certain salary range are more churning or with specific age and many more which is discussed in more detail in methodology section. In the end creation of XG Boost model which classifies churn.

1.3.3. Coupon redemption dataset and ML Model

The second stage starts with the creation of dataset for coupon redemption model which is done using Data Engineering with all the six tables which were used for churn prediction and defining which customers are said to be sensible to coupons or not sensible to coupons which means classifying consumers which will use the coupon or which will not before a campaign is run this will help companies to switch marketing strategies to deal with customers which are not going to use that coupon. Customers are sensible to coupon if the have redeemed at least one coupon else they are defined as not sensible. Data Cleaning is done as they might contain null values or outliers. Data Exploration of features is done with comparison to the output variable sensibility using charts to find out which factors are affecting more to the output feature. In the end, ML model is created with XG-Boost which classifies which consumers will use coupon or not.

2. Literature Review

2.1. Methodologies used on Churn Prediction

A new method has been proposed for customer segmentation (Joy et al, 2018). Customers are divided into groups based on their recency, frequency, and monetary values. Segmentation improves understanding of the customer's needs while also increasing revenue. Customer churn can be calculated by identifying clusters of customers who have similar needs. Clusters are discovered using fuzzy and k-means algorithms. The k-means algorithm has been shown to improve customer segmentation.

A novel approach was suggested for forecasting consumer loyalty (Pee, 2019). The perceived utility of a website is measured to determine how longitudinal changes and satisfaction effect consumer satisfaction. Perceived Usefulness (PU) is particularly high during the initial transaction, and subsequent transaction loyalty is dependent on consumer happiness. When a customer's happiness level rises, so does the customer's loyalty. PU assists in determining the best long-term strategy for customer retention and satisfaction. Customer retention is just as crucial as gaining new consumers. As a result, the approaches described here are extremely beneficial to the business's long-term viability.

In this research (Bagul et al, 2021), clusters have been developed based on recency, frequency, and monetary values that were computed using the RFM Model and KMeans Clustering. the number of patrons in the Every cluster is computed. In the end, the clients which need the business's more attention because they are vulnerable are recognised. Using this identification, appropriate Techniques can be used to keep these clients.

Various algorithms are tested and contrasted in order to estimate customer churn for a retail firm, and recommendations are made depending on the cluster to which the consumer belongs (Shetty et al, 2019). The best prediction algorithm was determined after analysing and studying several prediction algorithms. After comparing each churn predicting algorithm, Pareto/NBD was determined as the best. The model's input parameters were fine-tuned to meet the model's requirements. The items are classified, and consumers are assigned to clusters based on their RMF ratings. The recommendation is based on the cluster to which they belong.

(Gu et al, 2022) suggested a GBDT-based method to predict tobacco store attrition in this research. To that purpose, the features of geographical and time series dataset were extracted using a sliding window across time and geographic raster's before training them together. Given that the no. of deserting vendors was considerably more than the number of retained retailers, we were able to improve model execution in the prediction of unbalanced tests by altering the weight of CART leaf nodes. Most machine learning systems could predict vendor attrition with greater than 90percent accuracy in the 14-day and 30-day study periods after learning crucial spatiotemporal variables. CatBoost, among others, fared best in prediction after tweaking the weight of CART leaf nodes. In a 14-day observation period, the accuracy and recall of the updated CatBoost were 0.09 percent and 2.07 percent higher, respectively, than those of RFM metrics, and 0.04 percent and 0.97 percent higher, respectively, in a 30-day observation period. CatBoost beat LightGBM in prediction but required longer training time.

Two prediction models were developed for partial client attrition in retail (Miguéis, 2012). Both models contain a series of first-category purchases as a proxy for a customer's level of trust and demand maturity toward a firm. We simulate the initial purchase succession in both chronological and reverse order, considering the influence of the first impression on the present condition of the relationship as well as the impact of the most recent risks made. Both first-category purchase successions are modelled using a variable length dictated by the models' accuracy. The results show that both proposed models outperform the typical RFM model.

2.2. Machine Learning Approaches

A Logistic Regression Model have been used to forecast the likely churners while also implementing an exploratory data analysis using visualisation, statistical tests for feature selection, and data mining techniques (Ahmad et al, 2019). It has been noted that the logistic regression model has successfully forecasted superior churn prediction outcomes. It will produce a better outcome by raising the threshold values and choosing the appropriate features in various combinations.

(Wu et al, 2010) creates an ecommerce customer churn model and examines the factors that affect customer retention by using numerous data mining techniques, including clustering analysis, decision trees, neural networks, etc. to study customer churn. The questionnaire covered five key aspects of trust: the website's dependability and security, the company's strength, the utility of the product advertisements, and the cost-performance of the products.

In order to assess the elements determining the expenses incurred by the organisation to attract new consumers as well as costs to keep existing customers, (Patil et al, 2017) conducted a comparison study employing three algorithms: ensemble techniques, support vector machines, and boosting. They have implied that lowering client churn is a key objective for all internet businesses. Based on the transactions with the client, the churn value is calculated.

A new method was proposed for predicting churn (Gordini and Veglio, 2017). Customer churn is predicted using an SVMMauc model. The parameters have been optimised, as have the marketing data, which is noisy, nonlinear, and unbalanced. This method outperforms other methods such as traditional SVM, neural networks, and logistic regression. Churn prediction is critical in the e-commerce industry. As a result, this model outperforms other traditional methods in determining churn rate.

(Guo and Qin, 2015) used decision tree algorithms in data mining techniques to examine the fundamental information of ecommerce company customers and discover the characteristics of customer churns.

(Huang et al, 2012) investigated how telecom subscriber churn records fared in seven models, including decision tree, multilayer neural network, and evolutionary algorithm.

2.3. Evaluation Metrics

In this research (Coussement ,2014), the significance of accessible data that supermarkets save, as well as the optimal factors for predicting churn. Based on the approaches used, it is clear that analysing customers' historical behavioural patterns is critical for identifying churn. This is owing to the fact that clients in the retail business turnover slowly. The algorithms produced respectable results, with RBM achieving 83 percent and CNN achieving 74 percent.

The most well-liked machine learning algorithms employed by researchers for churn prediction was examined not just in the banking industry but also in other industries that heavily rely on user engagement (Kumar, 2016). Each churn prediction model examined here has poor prediction and accuracy.

In this paper (Mahalakshmi et al., 2020), Cross-validation is used to choose and elect superior algorithms. The concluding result is the Ensemble algorithm, which offers relevant preventative actions for churners. The writers also investigated the current system. Predictions in present models are made using unstructured or semi-structured input. Associated works reported findings from a single machine learning algorithm. A comparative analysis of several algorithms yields the best model in terms of accuracy. The algorithm's accuracy varies based on the client data.

(Yanfang and Chen, 2017) AUC test technique was used to assess the EBURM model. The findings reveal that the EBURM model corresponds to actual user expectations for both active and churn users. The EBURM model provides a customised operational advice method based on the many contributing aspects of the user retention rate. In comparison to the approach of user type prediction, this model can more precisely forecast user behaviour, reducing user churn. The development of the EBURM model to predict ecommerce user churn behaviour assists e-commerce platforms in more precisely formulating operational strategy, providing users with personalised recommendations, increasing user activity, retaining users, and improving the economic effects of e-commerce platforms.

(Khodabandehlou and Rahman, 2017) findings are divided into four sections: the first is about selecting efficient and significant variables in customer churning, the second and third are about comparing the capability and prediction accuracy of supervised machine learning methods with their different versions based on RFM and augmented RFM (RFMITSDP), respectively, and the fourth is about comparing RFM and augmented RFM models.

Vector Machine, and Logistic Regression. Finally, we can state that the C5.0 and Support Vector Machine algorithms are suitable for analysing data on client purchasing behaviour. the ROC curve depicts the most effective Support Vector Machine model, which has a specificity of 0.96 and a sensitivity of 0.59. The most effective algorithms in similar works were XGBoost, RF, and CART.

2.4. Coupon optimization techniques

Coupons were proven to have a favourable influence on repeat purchases in several research. (Taylor, 2001) concluded that buyers who accepted a coupon were almost seven times more likely to make a purchase after the promotion. (Lattin and Bucklin, 1989) discovered a large rise in consumers' purchase likelihood following a promotional purchase without distinguishing between users and non-users, however (Ailawadi et al., 2001) discovered only a minimal influence on customer retention for the product.

(PUSZTOVÁ and Babic, 2020) research sought to study the efficacy of several categorization algorithms for digital coupon marketing. For studies of historical data indicating coupon redemption for three possible goods in one order, we employed the CRISP-DM approach. We used two data samples for this purpose, one from the Data Analysis Cup and the other from the data preparation phase. We used and compared the following machine learning algorithms in the modelling and evaluation: C4.5, C5.0, Random Forest, Naive Bayes, Support

(Song and Yang, 2018) employs the XGBoost method in machine learning with feature engineering to develop a Digital Coupon Use Prediction Model, which was found to be dependable throughout the experiment due to its high accuracy. When organisations opt to provide digital discounts, the features with the highest rankings should be prioritised. When organisations opt to provide digital discounts, the features with the highest rankings should be prioritised. This research provides a decision-making foundation for businesses to accomplish correct marketing.

To forecast user behaviour, (Seo et al., 2022) found prior e-commerce marketing tactics. A deep learning algorithm for predicting customer attrition in real time produced an accurate result. They used findings to an online shopping mall in order to increase conversion rates and revenues. They created a framework to assess sales amount when combined with segment model and customised advised digital discount to see if our experiment had monetary benefit. They discovered that the model (scenario1) produces competitive outcomes. They discovered that it is appropriate for e-commerce online shopping malls to increase conversion rate and sales. Study actually shown that by utilising big data and deep learning technologies, marketing, a management discipline, could be solved more efficiently and swiftly.

2.5. Gap in Literature Review

There was a gap discovered in the literature review; very little research has been done in this area that performed both Churn Prediction model and coupon optimization on a single dataset; therefore, this research will assist the company in identifying all churners and then targeting them with specific coupon strategies to reduce churn.

3. Methodology

The Dunnhumby data, which comprises of eight tables, of which only six are used which are described below because other two tables are details of email and product so for this study they are not relevant as the products which customer bought would be beneficial for retargeting them after identification of churner but at this stage the research is focusing if customer has made a purchase or not rather than what consumer bought. The following phase is data exploratory analysis, which tells us what data we have data is not cleaned yet because that will be once dataset has been created for both models with extra features required for each of them. The following process is divided into two sections, essentially producing two new data sets, one for churn prediction and the other for coupon redemption. Following the construction of both datasets for churn prediction and coupon redemption, the churn is defined for the first dataset because there is no column in the dataset which says if the customer has churned or not, which splits the data into two categories: churners and non-churners. Following that, data pre-processing was performed, which included converting categorical data to numerical data, checking and removing of null-values and outliers which are important to check and remove because if they are present, they will bias the result, and converting data types. The next step is to see the target variable churn along with all of the other features. Using charts will enable us identify which factors are producing churn This provides information on churners such as their salary, age, and so on.

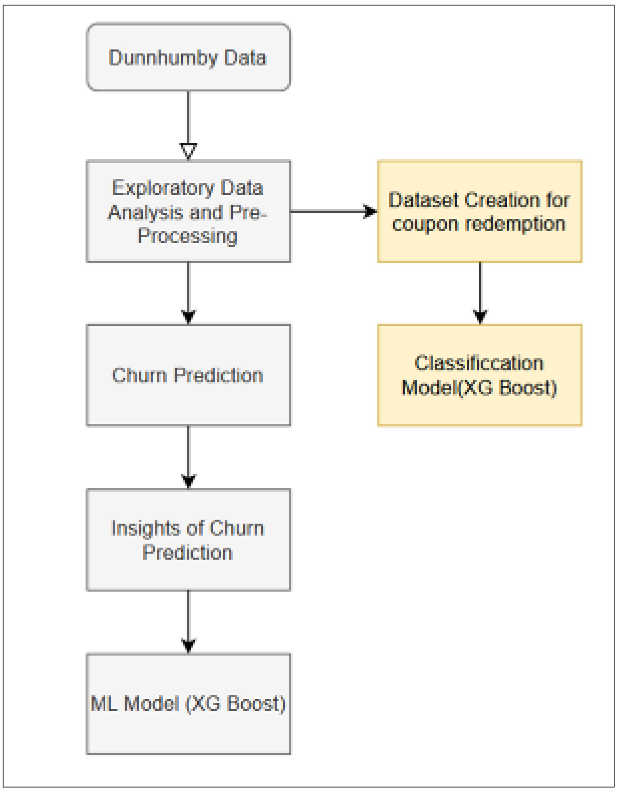


Figure 1 Methodology Flowchart

The final phase in churn prediction is to build an XG Boost classifier model that divides consumers into churners and non-churners. On the other side, after creating coupon redemption data with data engineering, the following stage was to determine whether or not customers are coupon sensitive which was if the customer has redeemed any coupon then they are said to be sensible else not. The following stage was data cleaning and display of the target variable coupon together with all of the other attributes same as performed for churn prediction model. Finally, the XG-Boost model was developed to differentiate between consumers who redeem coupons and those who do not.

3.1. Dunnhumby Dataset

Dataset taken is household dataset available on dunnhumby website, This dataset contains 8 tables but only 6 six of them are used because the other two tables are not required for this model as they contain information about the emailers (person who sent the email) and product description (name of the product) so this information is not required to calculate the churn prediction. This research focuses on calculating if customer has made a purchase or not rather than what customer has purchased. The tables which are used are described below:

- Campaign Descriptions (campaign_desc.csv) – This table contains details about start and end date of each campaign run with unique campaign id which has about 30 campaigns
- Campaigns (campaign_table.csv) – This contains information of each household participated in specific campaign from that 30 campaigns which were run.
- Coupons (coupon.csv) – This contains information on each individual coupon code, including the campaign id and the exact product purchased to use this coupon.
- Coupon Redemptions (coupon_redempt.csv) – It specifies which coupon was utilised on which day by which household and during which campaign.
- Transactions (transaction_data.csv) – This provides information on the number of transactions performed on which day and week, as well as the product id and quantity purchased.
- Demographics (hh_demographic.csv) – This describes all the demographic information related to each household such as their age, income, household size, etc.

3.1.1. Exploratory Data Analysis

Campaign description Table

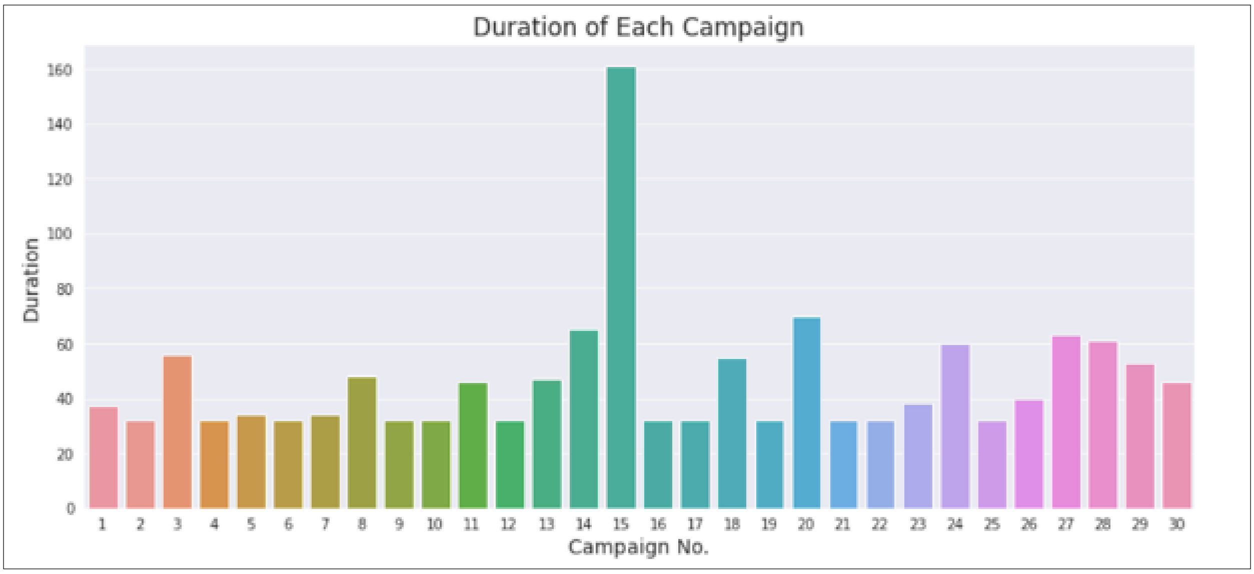


Figure 2 Each Campaign No. with duration of time

From figure 2 it can be seen that number of campaign are 30 and each of then with the duration of days. This is helpful information as it tells how many campaigns are present and from this only 25 campaigns will be used to train the model and test will be performed on the last 5 campaigns to classify the churners.

Campaign Table

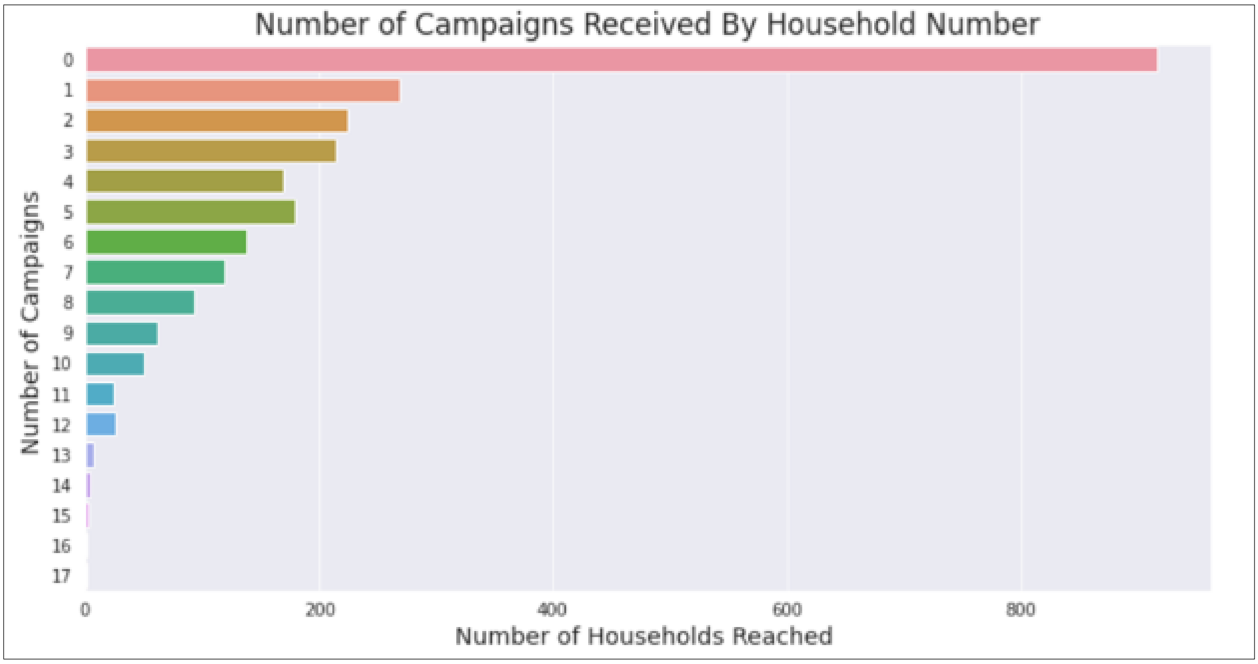


Figure 3 No. of Campaigns reached to each household

From the figure 3 campaign states to be reaching the highest number of households and it tends to decrease after that. This shows that out of 2500 homes, 1584 received a campaign once, while the remainder received no campaign at all. On the other hand, only a small percentage of homes participated in many campaigns. In the next sections, we will look into how this will affect churn rates.

Coupon redeemption Table

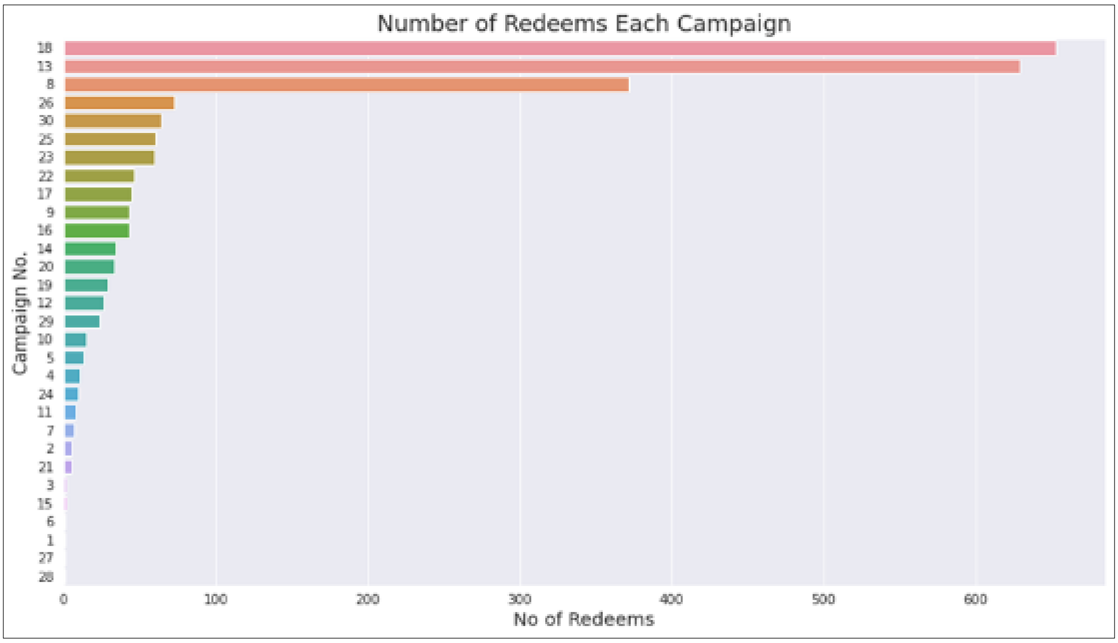


Figure 4 Coupons redeemed in each campaign

From Figure 4 number of redeems are highest for campaign 18 and 13. According to these findings, campaign organisers did well by advertising the appropriate campaigns (13 and 18) more frequently.

Transactional Table

Figure 5 Top performing stores

From Figure 5 it can be seen that sales of store 367 are highest with around 3600$. This information helps company identify which stores are performing the best based on their sales amount

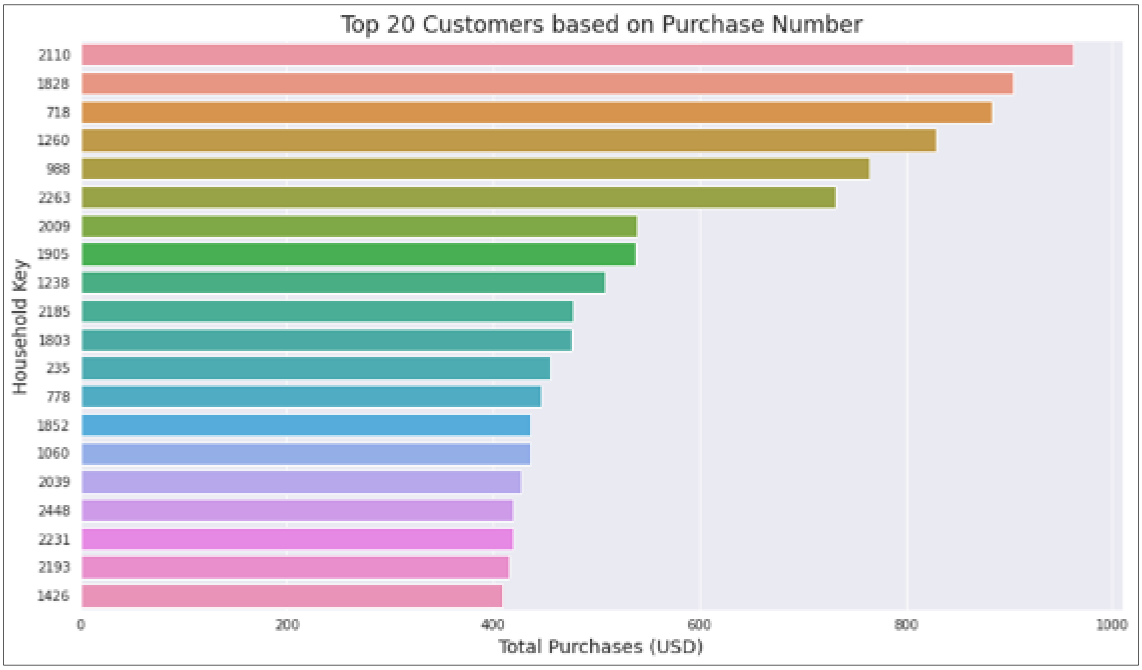


Figure 6 Top customers with purchases

Figure 6 shows the top households which are driving most revenue to the company and the highest being 980$. These are the loyal customers to the company so if any new products comes to the market these can be targeted by giving coupons to test the market of the product.

Demographics Table

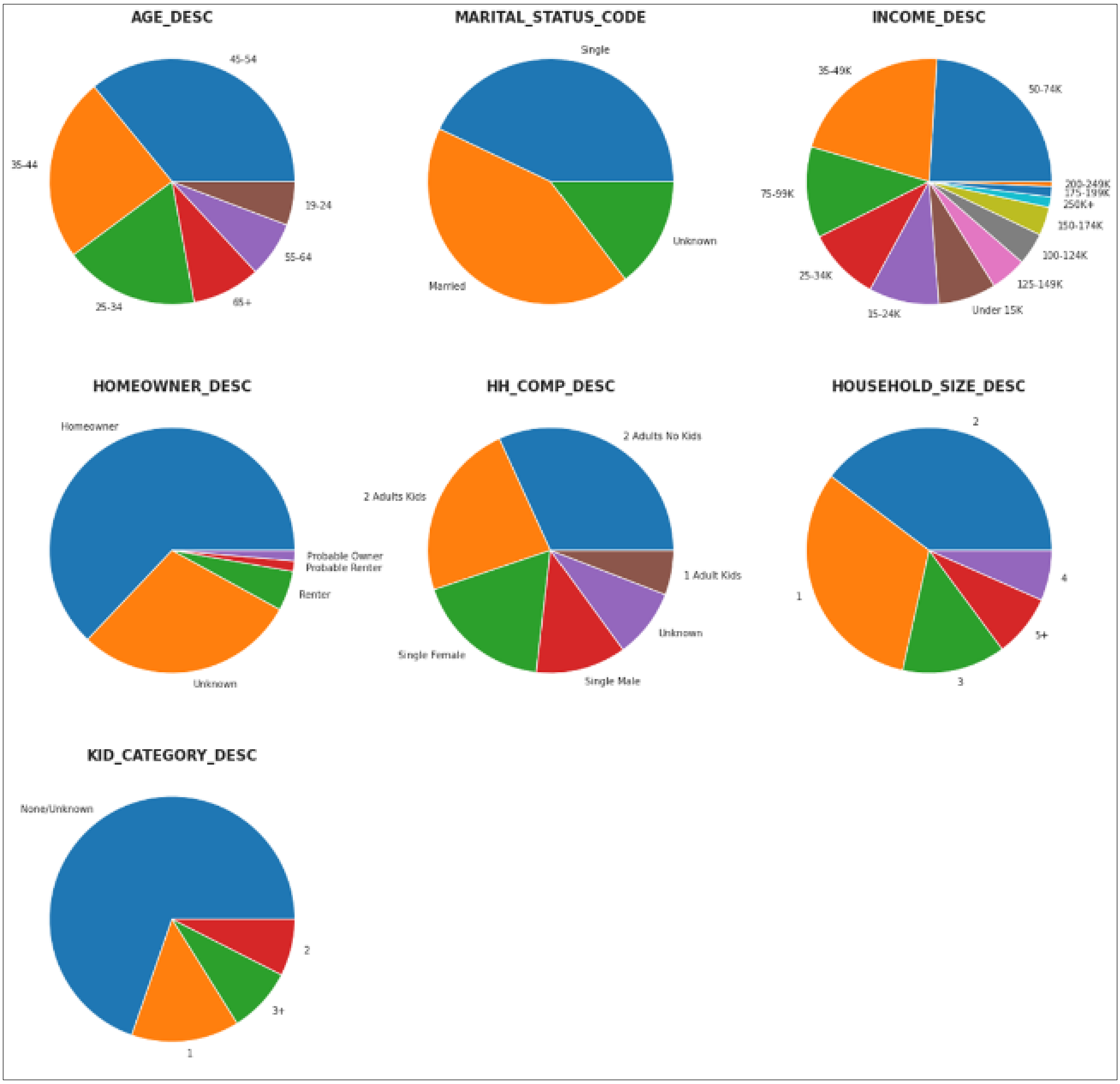


Figure 7 Demographics of customers

From figure 7 it can be seen that:

- The majority of clients 36 percent are between the ages of 45 and 54. This information is significant for the firm when compared to the churner age range which are visualized below once churn is defined because it can indicate if churners are of the same age range or churners are of an age range that makes up a small percentage of the company so that they may focus more on that age range with a variety of products.
- Married couples outnumber singles by nearly three to one. This shows that customers to this company are mostly married and can be compared with churn data if married couples are most in churning aswell.
- Almost half of the population earns between \$35 and \$74K per year. Most population is having a good salary which means that if they want to test a luxury product it can be a good market.
- The majority of clients 63 percent own a home. If most of them own their own houses they will definitely buy decoration products or furniture.
- The majority of clients 70 percent do not have children. This information says that most customers don't have babies so they should have much baby products in stores and all the factors discussed above will play a great

role in identifying churn segments which will be helpful for marketeers to target them later and build decisions based on certain cohorts.

3.2. Classifying Churn

Datasets were evaluated, and it was discovered that there is no column that shows if a client (household) is churned or not. As a result, churn definition was created before proceeding with the modelling. The average retail churn rate is five to seven percent every year. A churn rate of less than 5 percent is ideal, but anything beyond 10 percent is reason for concern according to industry standards. Even if you acquire more consumers, your firm will not grow until there are more incoming customers than existing ones.

A customer is considered as churn if they do not buy from store for 2 weeks. So, 2 weeks is our threshold for defining if the customer has churned or not. While defining this the split was 20 percent churned and 80 percent non-churned as seen in Figure 8. For retail business, 20 percent churn is a lot and this needs to be below 5 percent. So, predicting churn and retaining them is an essential part.

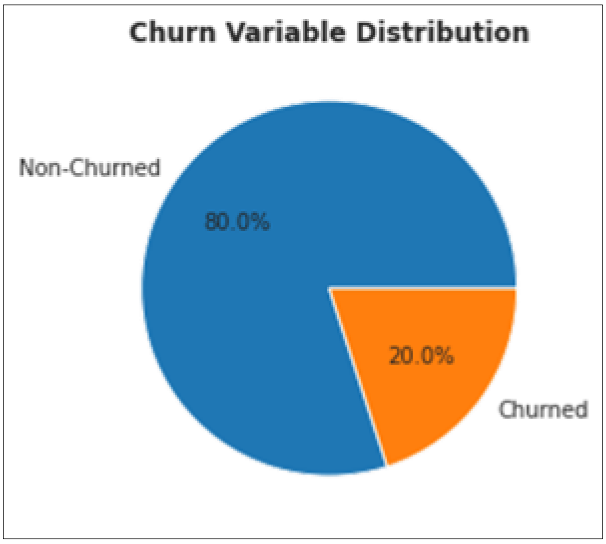


Figure 8 Split of Churner and Non-Churner

3.2.1. Feature Engineering and creation for Churn Dataset

Creation of Feature and merging them into single dataset used for Churn Prediction. These feature are created as the raw data has many transactions and coupon redemption for that single campaign for every household so these were grouped together for each household.

Features created –

- Feature 1: A list of the campaigns that each home received – Out of all 30 campaigns getting all campaigns which each household was part of. This will help in final prediction that if it has participated in the certain campaign or not. If it didn't participate for specific campaign and the result will be obviously churner so to avoid that this feature is created.
- Feature 2: Total number of campaigns received per household – For every household total number of campaigns were calculated if they have been part of how many campaigns so that the final prediction is not biased.
- Feature 3: Campaigns that resulted in voucher redemption – Calculating campaigns that resulted in coupon redemption will determine the campaign's success.
- Feature 4: The number of redemptions made by each family – Calculating total redemptions made by each home to determine whether or not they are interested in that voucher.
- Feature 5: Household purchase within two years – Calculating the total quantity of purchases made by each home over the course of two years in order to identify the top company participants. This will assist the firm in determining which homes are their most loyal clients. In final prediction, knowing the purchase history of every household will result in better classification.

In the final dataframe, we now have 91 features and a target variable.

3.2.2. Data Pre-Processing

Converting NaN values and data types

Replacing NaN values to 0 because features such as number of redeems and kid category has values NaN and None respectively. It was converted to 0 because if redeems are not known it cannot be replaced with mean or median as this will affect our final output by increasing the number of redeems. So, converting them to 0 was the only option. Converting Object to Integers as some of the features are in the wrong data type such as integers are in Objects. XG Boost can handle missing values and it also considers NaN values to zero but we have done it before. Household 0 was an outlier because the number of received campaigns was the highest more than triple the average so this houshold was removed.

Label Encoding

ML models only function with numerical data, they cannot be fed categorical values and must be transformed to numerical data. Label Encoder is used to convert categorical data into numerical which is available from the sklearn library in python programming. Label Encoding is a well-known encoding technique for dealing with categorical information. Based on alphabetical sorting, each label is issued a unique integer in this approach.

Standardization of data

The power transform approach is used to give the data a more Gaussian distribution. Power transformations are a set of techniques that employ a power function (such as a logarithm or exponent) to make a variable's probability distribution Gaussian or more-Gaussian-like. Power transform removes skewness from the data. The Box-Cox and Yeo-Johnson transformations are two distinct methods for making a continuous (numeric) variable seem more regularly distributed. They are frequently employed in feature engineering to eliminate skew in raw variables. Box-Cox works only with positive values while Yeo-Johnson has no restrictions and it deals better then 0 values for this reason Yeo-Johnson method is used on this dataset.

3.2.3. Data Exploration of features with target variable

Customers with age 45-54 and 65+ turn to churn less which can be seen in Figure 9 as discussed above while exploratory data analysis of raw data it will be useful to compare this information with the original data. In raw data almost half of customers are of this age range and the lowest churn rate is in this section which is good for the company and they can target more on the rest of the age range with specific coupons to each group.

Customers with salary 25-34k and paying rent tends to churn less which can be seen in Figure 10 which are in very small percentage of around 9.2 while customer having more salary tend to churn more as they are the ones who can afford more expensive products and half of the customers earn more than 50k. So, company needs to focus more on this section to improve their churn rate and specifically target them.

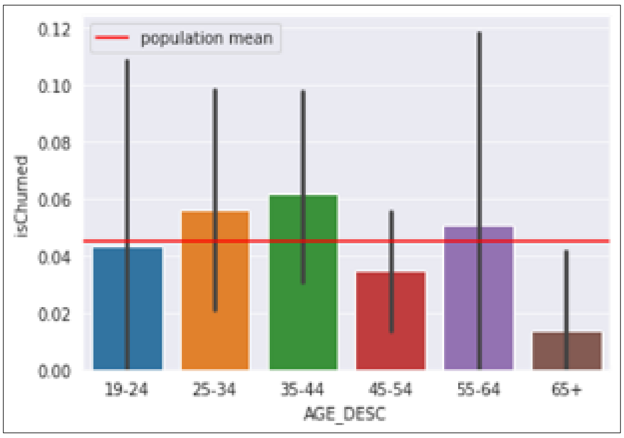


Figure 9 Comparison of Age with Churn

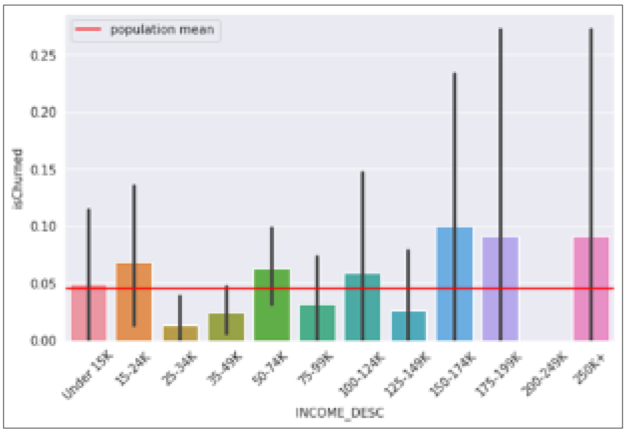


Figure 10 Comparison of Income with Churn

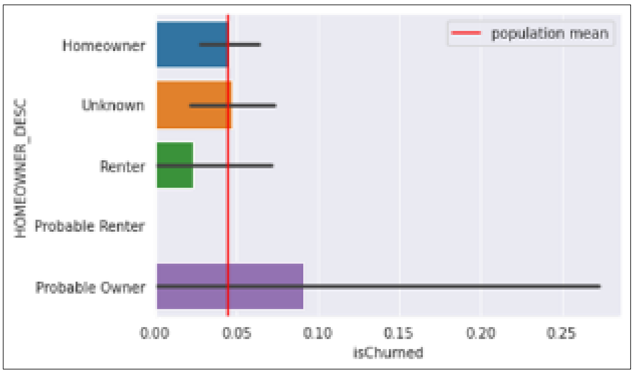


Figure 11 Comparison of Homeowner with Churn

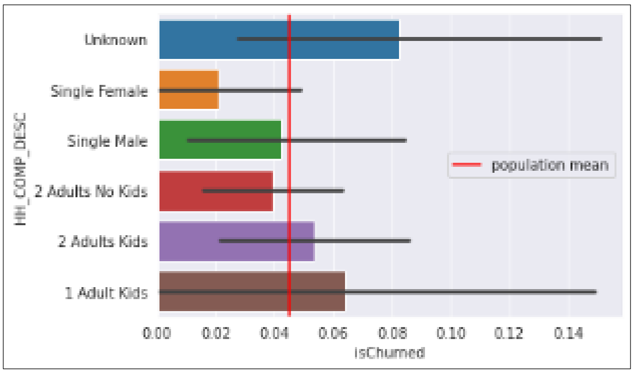
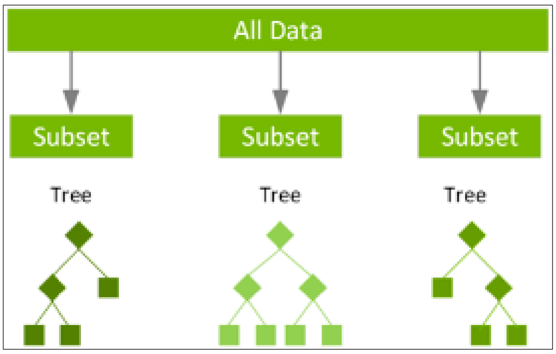


Figure 12 Comparison of Marital vs Churn

The following visualisations can help the organisation particularly target each cohort by looking at their churn rate and creating alternative marketing strategies to lower churn while taking into account all of the criteria indicated above. Also, when comparing this to exploratory data analysis of raw data, if the specific target is valuable for the firm to invest money in, if they are in tiny numbers, it will be futile to keep them rather than targeting that cohort first, which would bring in more revenue.

3.2.4. Model Creation

Splitting data into train and test data. 70percent of the data was used to train the data and 30percent was used to test the data. The machine learning model is XGBoost (Extreme Gradient Boosting), which has been shown to perform well with skewed datasets. One explanation is that it is a tree-based model, and trees are particularly resistant to skewed data and outliers. XGBoost is a method of ensemble learning. It is not always adequate to depend on the outcomes of a single machine learning model. Ensemble learning provides a methodical approach to combining the predictive capacity of numerous learners. The end result is a single model that aggregates the output of numerous models. Boosting trees are constructed consecutively, with each succeeding tree attempting to minimise the mistakes of the prior tree. Each tree learns from the trees that came before it and updates the residual mistakes. As a result, the following tree in the sequence will learn from an updated version of the residuals. The basic learners in boosting are weak learners with a strong bias and predictive power that is just slightly better than random guessing. Each of these weak learners gives some crucial information for prediction, allowing the boosting strategy to efficiently combine these weak learners to build a strong learner. The ultimate powerful learner reduces both the bias and the variation.



Source: https://www.nvidia.com/en-us/glossary/data-science/xgboost/

Figure 13 Decision Tree Model

The boosting ensemble approach is composed of three easy steps:

- Step 1 To forecast the target variable y, an initial model F0 is defined. A residual will be associated with this model (Y – F(0))
- Step 2 A new model h1 is fitted to the previous step's residuals. F0 and h1 are now joined to produce F1, the enhanced version of F0. F1's mean squared error will be less than that of F0.

$$F_1(X) = F_0(X) + H_2(X)$$

Step 3 To increase F1's performance, we might model after its residuals and develop a new model F2. This can be repeated for'm' iterations till the residuals are as low as possible. The additive learners in this case do not interfere with the functions generated in the preceding phases. Instead, individuals provide their own information to reduce the inaccuracies.

$$F_m(X) = F_{m-1} + H_m(X)$$

Gradient boosting entails the following steps:

- Step 1 F0(x), with which we will initialise the boosting process, must be defined as follows:

$$F_0(X) = argmin_{\gamma} \sum_{i=1}^{n} L(Y_i, \gamma)$$

- Step 2 Iteratively, the gradient of the loss function is computed:

$$r_{im} = -\alpha[\frac{\delta(L(y_i, F(x_i))}{\delta F(x_i)}]_{F(x)=F_{m-1}(x)}$$

- Step 3 Each hm(x) is fitted to the gradient at each step:

$$F_m(X) = F_{m-1}(X) + \gamma_m h_m(X)$$

ML model for churn prediction has been created once it is known which all customers are churners and which customers to target. Specific campaigns are introduced by marketers to retain these customers by various techniques on of which is providing discount coupons. Coupon redemption technique will be used to identify which all customers are not using the coupons beforehand so that before running a specific campaign if it is know that this customer will not redeem that coupon it will be waste of money to send them so different other strategies can be involved to target them. Next, dataset will be created for coupon redemption technique.

3.3. Dataset creation for Coupon Redemption

Only 4 tables have been used for creation of this dataset which are Demographic data, Campaign data, Coupon redemption data, Transaction data. Two tables are not used for this dataset because coupon description and campaign description tables are not needed as only focus is to find out if that coupon is redeemed by a household or not rather than what was that coupon.

Merging these datasets with following conditions and the reasons are given with specific conditions below:

- Keeping only customers whose demographics dataset is present and which took part in atleast one campaign to receive coupons as customers which did not take part in any campaign will be of no use and act as outliers.
- Calculating total sales, median basket spends and average product price from day 1 to 615 as as we are leaving last 5 campaigns to predict the output and start_day of last 5 campaigns were from day 615 so we exclude them from our data.
- Exclude transactions with a sales value and quantity that are less than or equal to zero because if a household has never done a transaction or never bought a product from that retail it will affect out end result so these are excluded.
- Keeping only the sales data which is before day 615 because only campaigns which are run until 615 are taken into account.
- Keeping only retrieved coupons before Day 615 because we are leaving last 5 campaigns.
- Dropping columns which are not used in the model such as 'DAY', 'COUPON_UPC' because on what date the coupon was redeemed is not needed and also the coupon id as which campaign the specific coupon was redeemed is important.

3.3.1. Defining Coupon Redemption

Coupon Redemption is defined as customers which have redeemed at least one coupon are said to be sensible or if not then they are said to be not sensible to coupons.

3.3.2. Data Exploration of Coupon Data

Visualizations below is helpful for the company for coupons with the exploratory data analysis of raw data in Chapter Examples are given below:

- Customers who are single and have an age of 45-54 are most non-sensible to coupons which are around 36 percent which are a lot of customers which are not reacting to the coupons which means they need to send them different coupons.
- People living on rent are more sensible to coupons which means customers not having their own home react more to discounts.
- Customers with 2 Adults and no kids are most likely to not use the coupons which means that they are not utilizing the coupons and company is bearing a loss in targeting them. If they are getting coupons which are kid products, and they don't have kids this means they need to change the strategy.

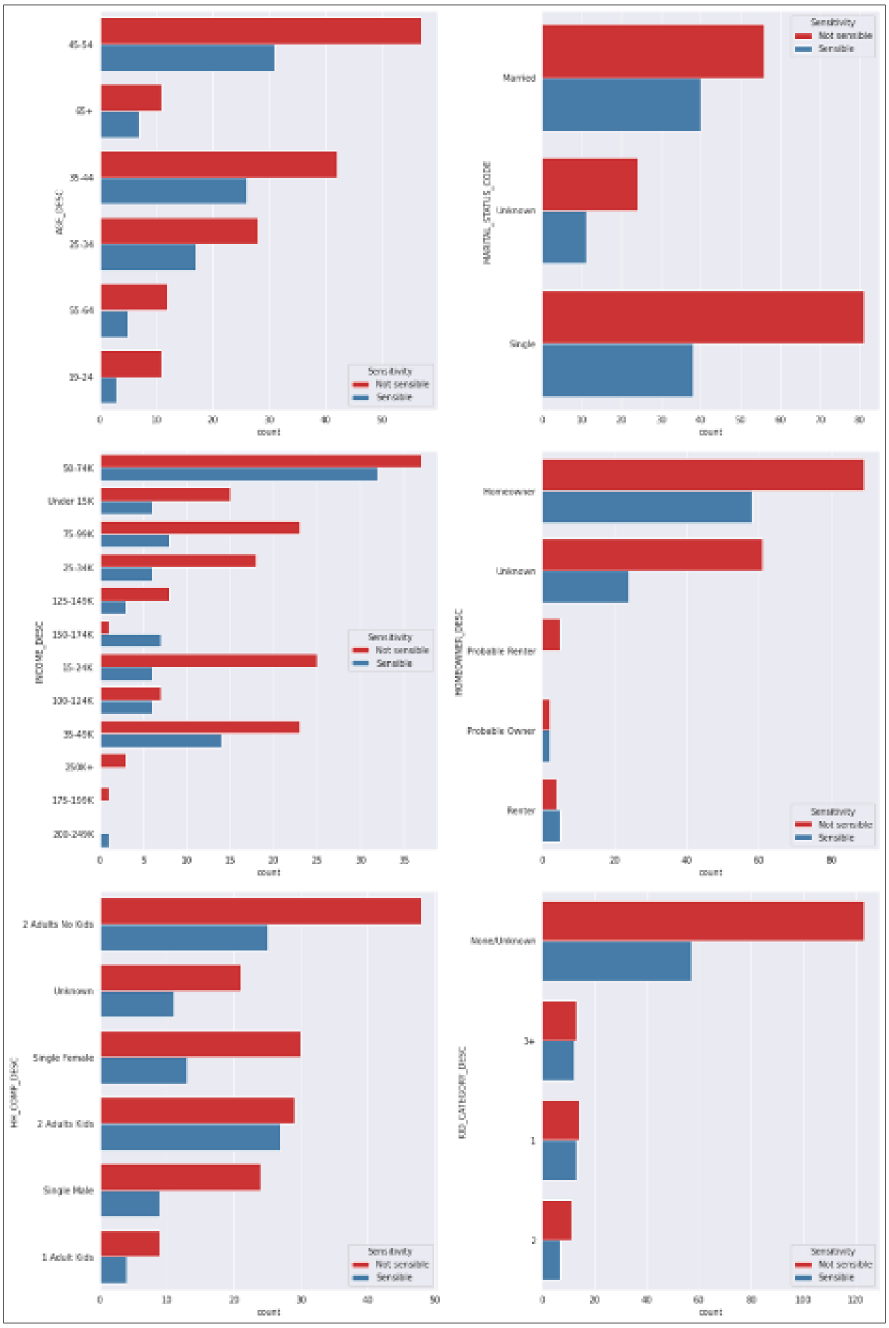


Figure 14 Comparison of Sensitivity of coupon with other features

3.3.3. Checking skewness of the data

64.4percent of our clients are unconcerned with coupons. As a result, our sample is slightly skewed, with more clients who aren't interested in discounts. From figure 15 we can observe that the variables have a Multinomial distribution and a Gaussian-like distribution with a lengthy right tail. This is supported by the positive skewness readings.

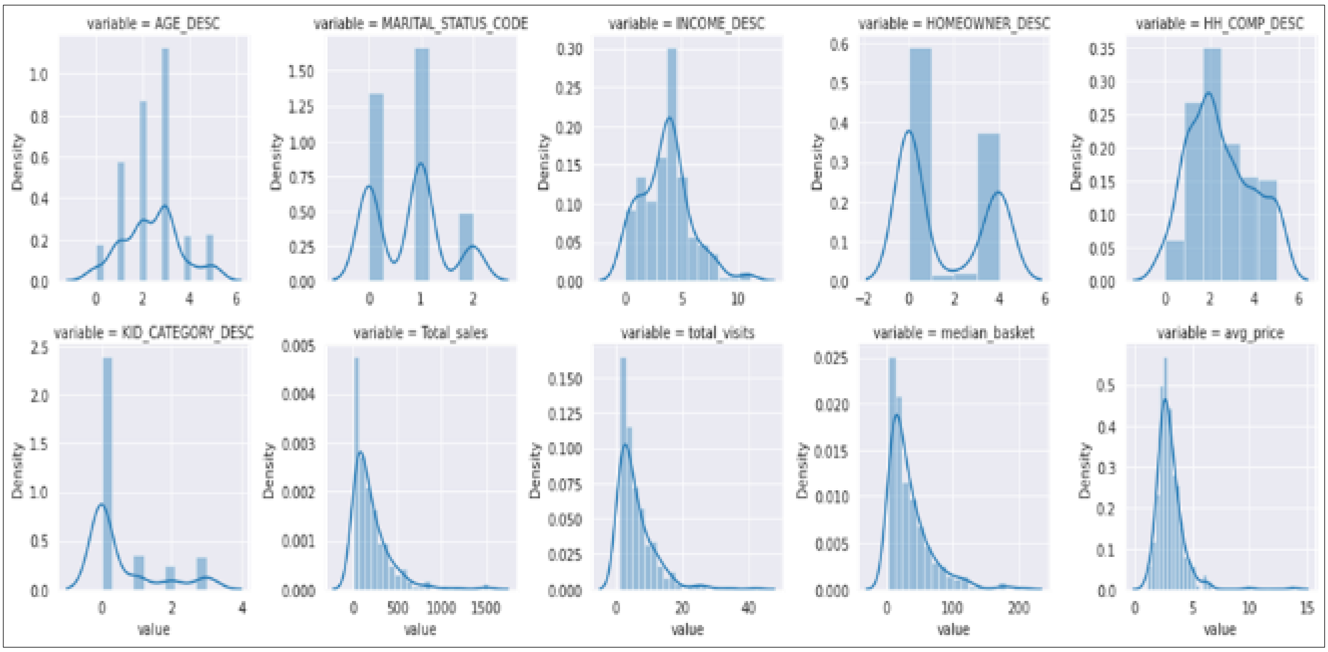


Figure 15A Skewness of Coupon Data

3.3.4. Checking Co-Relation

From Figure 16 Households SIZE DESC,KID CATEGORY DESC, and HH COMP DESC are all interrelated. Even though, as previously stated, XGboost handles correlated characteristics, we will delete HOUSEHOLD SIZE DESC.

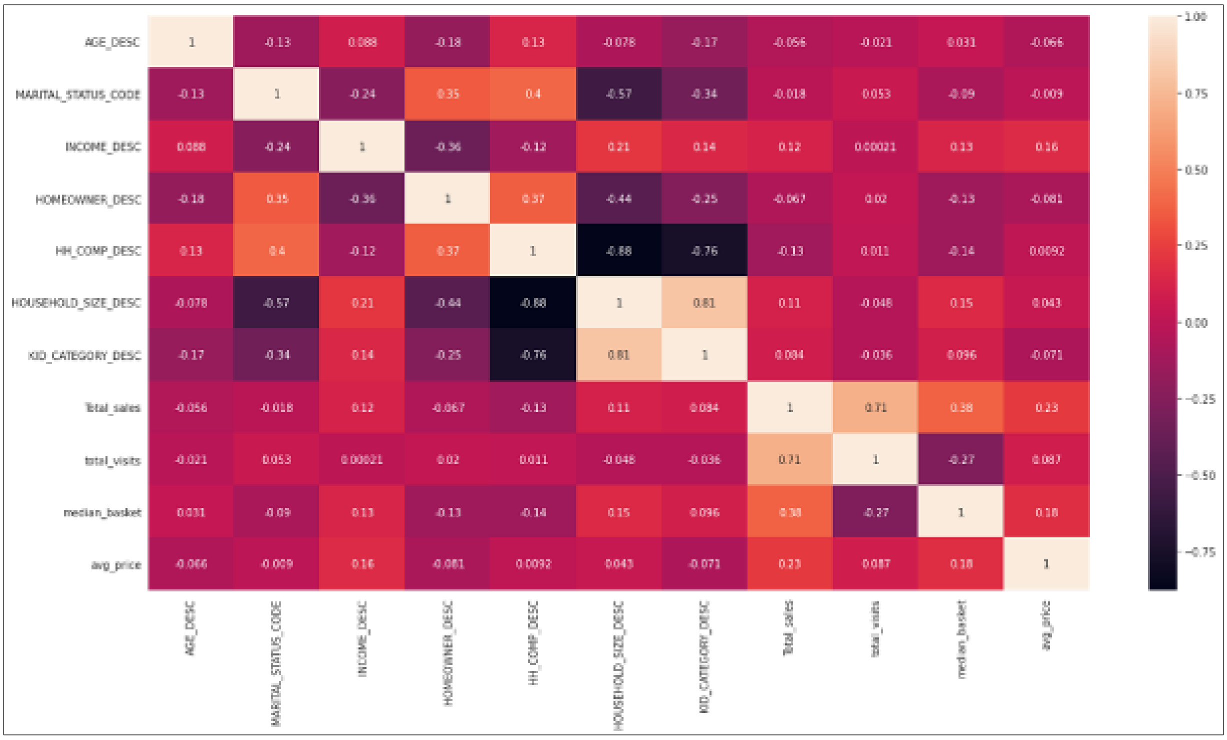


Figure 13 B Co-relation of coupon data

3.3.5. Converting Categorical data into numerical

Label Encoder is used to convert categorical data into numerical which is available from sklearn library. Label Encoding is a well-known encoding technique for dealing with categorical information. Based on alphabetical sorting, each label is issued a unique integer in this approach.

3.3.6. Standardisation and Feature Selection

The power transform approach will be used to give our data a more Gaussian distribution. Power transformations are a set of techniques that employ a power function (such as a logarithm or exponent) to make a variable's probability distribution Gaussian or more-Gaussian-like. Power transform basically removes skewness from the data. Yeo Johnson method is used on this dataset.

PCA transformation is used for feature selection. 7 components were taken that explains 90percent of the variance. The technique of principal component analysis (PCA) is used to identify a smaller number of uncorrelated variables known as principle components from a larger set of data. The method is commonly used to highlight variance and capture significant patterns in a data collection.

3.3.7. Splitting data into train and test

Splitting data into train and test data 70percent of the was used to train the model and 30percent was used to test the data. Sklearn was used to split the data.

3.3.8. Model Creation for Coupon Optimization

XG Boost model is used for coupon redemption because it works well with small to medium dataset as our dataset is small this is one of the best algorithms to use also it handles missing data very well.

3.4. Evaluation Metrics

Following the application of machine learning algorithms, certain tools are required to determine how successfully they fulfilled their tasks. These are known as performance evaluation metrics. A large variety of metrics have been introduced in research, each of which analyses different elements of an algorithm's performance. As a result, for each machine learning challenge, an adequate set of measures for performance evaluation is required. In this research, various standard metrics are employed for classification issues to gather useful information about algorithm performance and to conduct a comparison study. Precision, recall, f1-score , accuracy, confusion matrix, and ROC-AUC score are some of these measurements (Mahalakshmi et al., 2020). Accuracy is defined as the ratio of correct predictions and total predictions. Precision is defined as ratio of true positive and sum of true positive and false positive (1). In other words, How many of the observations predicted by an algorithm to be positive are really positive. The recall is defined as ratio of true positive and sum of true positive and false negative (2). That is, How many of the observations that are truly positive were predicted by the algorithm.

$$\text{Precision} = \frac{TP}{TP+FP} \text{ ……..(1) ,}$$

$$\text{Recall} = \frac{TP}{TP+FN} \text{ ……………(2)}$$

3.5. Hyperparameter Tuning

The parameters that determine the model architecture are known as hyperparameters, and the process of looking for the best model architecture is known as hyperparameter tuning. Hyper parameter tuning is performed with Cross-validation because data contains skewness and the data is skewed on output variable because 80 percent of the data is non-churner.

4. Results

4.1. Churn Prediction

4.1.1. Evaluation Metrics

The most important metric for this model will be recall and accuracy because recall identifies what percentage of true positives were successfully identified which is most important for this model because identifying actual number of churners is most important. The XG Boost model has an accuracy of about 98 percent. Even though the test set's accuracy is 98 percent, it is deceptive since our goal variable is skewed toward not churned (80 percent). Even a very simple model that always chooses the majority class will achieve 80 percent accuracy. As a result, concentration is required more on how well our model works on the minority population (churned households). Only 4 of the 236 samples on the test set are being churned according to Figure 16. Because our model was unable to detect any of them, the test set recall was calculated to be 0 percent as shown in Table 1. This is the area of improvement. Since there are various techniques to improve a model's performance one of the best is to feed more data but unfortunately data is limited so another method is to optimize hyperparameters with cross-validation to improve the performance. So, the goal is to find the best combination of parameters which is the best model for the data.

Table 1 Baseline model churn results

	precision	recall	F1-score
0	0.98	1	0.98
1	0	0	0
accuracy			0.98

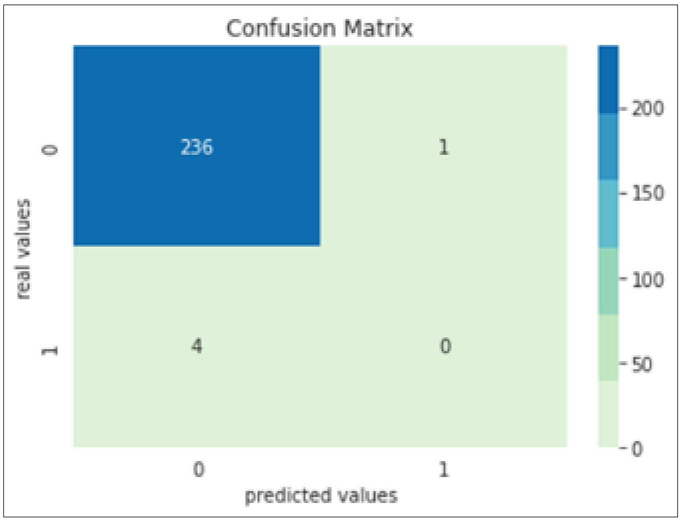


Figure 14 Base model Confusion matrix churn

4.1.2. Randomised Search CV with Cross-Validation

Random search is a method for selecting and training a model using random combinations of hyperparameters which utilise the best random hyperparameter combinations. Randomisedsearch CV is used to obtain the best hyperparameters which will be fed to the XG Boost model to get the best results. Best parameters obtained from the algorithm are shown in the table 2. Since the data is imbalanced due to skewness on output variable churn to deal with that problem Cross-Validation is performed. This will help to train and test dataset on multiple folds and validates it. K-Fold method is used to perform Cross-Validation because it will divide the data into k folds and k is defined by the user which is 5 in this case. This will divide the data into k equal parts and first set is used to train the model and remaining k-1 will be used to test the data. In the second iteration, the second set is chosen as a test set, while the remaining k-1 sets are utilised to train the data and determine the error. This method is repeated for all k sets.

Table 2 Hyperparameters Churn

XG Boost Classifier	{'subsample': 0.6000000000000001, 'silent': False, 'scale_pos_weight': 6.2, 'reg_lambda': 1000.0,

	'n_estimators': 98, 'min_child_weight': 0.4, 'max_depth': 48, 'max_delta_step': 1, 'learning_rate': 0.15, 'gamma': 0.8, 'colsample_bytree': 0.1, 'colsample_bylevel': 0.30000000000000004}

4.1.3. Optimised Model Results

The results shown in Table 3 are obtained from model with hyperparameter tuning with Cross-Validation. Despite having lower overall accuracy, we are now receiving substantially better outcomes for the minority class in terms of recall. The table 3 displays all metric results for both the test and train sets. The train and test results are similar, indicating that our model did not overfit. This will help the company to reduce churn while having a better accuracy.

Table 3 Optimized Churn model results

	Score
test_roc_auc	0.738727
train_roc_auc	0.810915
test_recall	0.779308
train_recall	0.850954
test_precision	0.387171
train_precision	0.408329
test_accuracy	0.639844
train_accuracy	0.762509
test_balanced_accuracy	0.698351
train_balanced_accuracy	0.741392

4.2. Coupon Redemption

4.2.1. Base model

We will focus on two assessment measures to analyse the performance of our model Accuracy and Precision. Precision is important in this case because we want to reduce Type 1 mistake, often known as false positive. Indeed, we would wish to prevent the situation in which a consumer is anticipated to be coupon sensitive but is not. This would result in a financial loss because we would print and mail him/her coupons for free.

Table 4 Coupon Model Results

	precision	**recall**
0	0.38	0.37
1	0.56	0.61
accuracy		0.57

Our base model has a 57.7 percent accuracy and a precision of 38.1 percent as seen in Table 4. This indicates that our model is right 57.7 percent of the time when it predicts that a client is coupon sensitive. We will endeavor to tune our model to increase its performance by hyperparameter tuning with Cross-Validation. Hyperparameter tuning will be performed to get the best parameters of the model which in turn will have a great impact on model. Cross Validation is performed because data is imbalanced because of skewness on output variable skewness.

4.2.2. Hyperparameter tuned model

Our accuracy increased to 61.7 percent after hyperparameter tuning, and our precision increased to 45 percent as seen in Table 5. The best parameters obtained are shown in Table 6. The model is not 100 percent accurate but this will help the company to reduce its costs by predicting the coupon sensitivity at least 60 percent of the time.

Table 5 Coupon Optimised Model Results

	precision	**recall**
0	0.45	0.57
1	0.48	0.64
accuracy		0.61

Table 6 Coupon Optimized Model Parameters

XG Boost Optimised	{'gamma': 0.7, 'learning_rate': 0.04, 'max_depth': 3, 'n_estimators': 198}

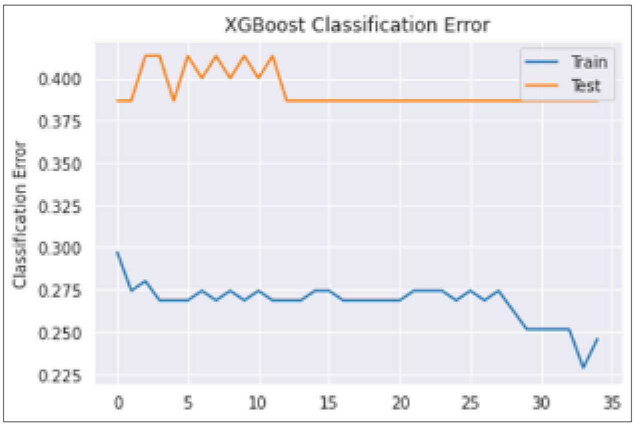


Figure 15 XG Boost Classification Error

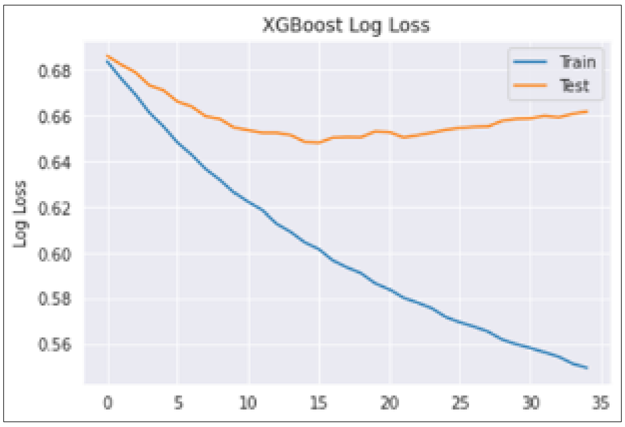


Figure 16 XG Boost Log Loss

From Figure 18 and 19 we can see from the graph that overfitting has been avoided thanks to early stopping. The best iteration has been found earlier at round 3 and this is confirmed graphically, we can see the test curve start increasing and decreasing and tends to be constant at point 11.

5. Discussion

5.1. Results against hypothesis

Results obtained from the model are good for the company If used both the models together. According to original hypothesis "The primary goal of this research is to use machine learning algorithm to optimise the Dunnhumby dataset to prevent customers from churning and to use coupon optimization to reduce the churning rate which is one of the marketing strategies to prevent churn". The model will assist the organisation in increasing profits by analysing data and providing information about churners, and the coupon optimization model will assist in developing alternative marketing techniques to minimise churn and keep consumers satisfied.

Recommendation

It is suggested that identifying all the churners before they are actually churned in the future will help the company retain their customers which will in turn increase the revenue f the company which can be identifying the Churn Prediction Model. The Coupon Redemption model will help the company identifying the customers who will not use the coupons which in turn helps save time and money sending coupons to churners. The company can use different methods of marketing to retain them such as giving discounts on specific products which they buy regularly. By identifying Customer churn, you know that the current method of marketing is not working, or the customer is not happy you can send them feedback forms to know what the reason and you can increase your customer service depending on the problem customer is facing. Increasing customer satisfaction means that word of mouth will increase which in turn will attract more new customers. Coupons may assist in the introduction of new product lines as well as encouraging clients to try a more profitable brand or service.

Limitations

The data was limited and very less to predict for customers in a larger company where the customers are in millions such as Sainsbury's and Tesco so the model might not perform well with larger customers as training data is less. The data contains skewness which led to the poor performance of the churn model. The model might not work on the current date as data is old and the purchasing behavior changes with time also the prices of products increase. Customers with no purchase were removed which might be new to the company and coupons are needed to attract them to buy the product.

6. Conclusion and Future Works

The original hypothesis stated that two models will be created to make the company profitable by reducing churn. Two models were suggested in this paper: one for churn prediction and the other for coupon redemption. As a result, it adds to the body of knowledge on churn prediction. It motivates marketing executives to develop a marketing strategy before launching a promotional campaign to attract clients. Both models, the first predicting which consumers will churn and the second predicting whether this coupon will be utilized by the customer or not, will help marketing executives improve their tactics for planning differently for different sorts of customers. Providing several sorts of discounts to various categories of people aids a firm in retaining more clients. This in turn will result in more profits for the company as more customers will be purchasing and more customers will take part in redeeming their coupons. Churn model has an accuracy of 63 percent and Coupon model has an accuracy of about 61 percent which means both models are not 100 percent accurate in classifying but it will definitely help the company grow once the more data is fed into algorithms to train the more the accuracy will increase with time.

Future Works

The dataset was not large enough so if this model runs on a large dataset, it's difficult to know how the model performance will be so in the future the model will be trained on a larger dataset.

Only XG- Boost model is used to classify churn and coupon and there are different other models which can be used so there is no comparison of which model performs better so more algorithms will be used to train on this dataset and test.

Two other datasets which were not used in this model which has the data about the history of product bought by customers can be used to create a model that can predict the best product purchased by the customer and a recommendation system can be built to create a specialized coupon to target them specifically.

References

[1] Ahmad, A.K., Jafar, A. & Aljoumaa, K. Customer churn prediction in telecom using machine learning in big data platform, J Big Data 6, 28 (2019). https://doi.org/10.1186/s40537- 019-0191-6.

[2] Ailawadi, K. L., Neslin, S. A., & Gedenk, K. (2001). Pursuing the valueconscious consumer: store brands versus national brand promotions. Journal of Marketing, 65(1), 71–89.

[3] Bagul, N., Berad, P., Surana, P. and Khachane, C., 2021. Retail customer churn analysis using rfm model and k-means clustering. Int J Eng Res Technol (IJERT), 10, pp.349-354.

[4] Chen, Z.-Y., Fan, Z.-P. and Sun, M. (2012) 'Decision support A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data', European Journal of Operational Research, 223, pp. 461–472. doi: 10.1016/j.ejor.2012.06.040. 64 Chollet, F. and others (2015) 'Keras'.

[5] Chorianopoulos, A. (2015). Effective CRM using Predictive Analytics. In Effective CRM using Predictive Analytics. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9781119011583.

[6] Coussement, K., 2014. Improving customer retention management through cost-sensitive learning. European Journal of Marketing.

[7] De la Llave Montiel, M.A. and López, F., 2020. Spatial models for online retail churn: Evidence from an online grocery delivery service in Madrid. Papers in Regional Science, 99(6), pp.1643-1665.

[8] Gordini, N. and Veglio, V., 2017. Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. Industrial Marketing Management, 62, pp.100-107.

[9] Gremler, D. D., & Brown, S. W. (1999). The loyalty ripple effect: Appreciating the full value of customers. International Journal of Service Industry Management, 10, 271–293.

[10] Gu, Q., Feng, M. and Lin, Y., 2022, April. Research on Retailer Churn Prediction Based on Spatial-Temporal Features. In 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP) (pp. 876-884). IEEE.

[11] Guo, F. and Qin, H.L., 2015, November. The Analysis of Customer Churns in e-Commerce Based on Decision Tree. In 2015 International Conference on Computer Science and Applications (CSA) (pp. 199-203). IEEE.

[12] H. -l. Wu, W. -w. Zhang and Y. -y. Zhang, An Empirical Study of Customer Churn in E-Commerce Based on Data Mining, 2010 International Conference on Management and Service Science, Wuhan, China, 2010, pp. 1-4, doi: 10.1109/ICMSS.2010.5576627

[13] Huang, B., Kechadi, M.T. and Buckley, B., 2012. Customer churn prediction in telecommunications. Expert Systems with Applications, 39(1), pp.1414-1425.

[14] Jones, M. A., Mothersbaugh, D. L. and Beatty, S. E. (2000) 'Switching barriers and repurchase intentions in services', Journal of Retailing, 76(2), pp. 259–274. doi: 10.1016/S0022-4359(00)00024-5.

[15] Joy, Christya, A.Umamakeswaria, L.Priyatharsinib, A.Neyaab RFM ranking – An effective approach to customer segmentation Journal of King Saud University - Computer and Information Sciences ,5 September 2018 9. LG Pee, JJ Jiang ,G Klein E-store loyalty: Longitudinal comparison of website usefulness and satisfaction,International Journal of Market Research 1–17

[16] Khodabandehlou, S. and Rahman, M.Z., 2017. Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. Journal of Systems and Information Technology.

[17] Kumar, A. S. and D. Chandrakala. A Survey on Customer Churn Prediction using Machine Learning Techniques. International Journal of Computer Applications 154 (2016): 13-16.

[18] Lattin, J. M., & Bucklin, R. E. (1989). Reference effects of price and promotion on brand choice behavior. Journal of Marketing Research, 26(August), 299–310.

[19] Louridas, P. and Ebert, C. (2016) 'Machine Learning', IEEE Software, 33(5), pp. 110–115. doi: 68 10.1109/MS.2016.114.

[20] Louzada, F., Ara, A. and Fernandes, G. B. (2016) 'Classification methods applied to credit scoring: Systematic review and overall comparison', Surveys in Operations Research and Management Science, pp. 117–134. doi: 10.1016/j.sorms.2016.10.001.

[21] Mahalakshmi, J., Gnanavel, D., K S, G., Phriyaa S, H. and Sree K, K., 2020. Prediction of Customer Churn on e-Retailing. International Journal of Recent Technology and Engineering (IJRTE), 8(6), pp.5541-5545.

[22] Miguéis, V.L., Van den Poel, D., Camanho, A.S. and e Cunha, J.F., 2012. Modeling partial customer churn: On the value of first product-category purchase sequences. Expert systems with applications, 39(12), pp.11250-11256.

[23] Patil, A.P., Deepshika, M.P., Mittal, S., Shetty, S., Hiremath, S.S. and Patil, Y.E., 2017, August. Customer churn prediction for retail business. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 845-851). IEEE.

[24] Pee, L.G., Jiang, J.J. and Klein, G., 2019. E-store loyalty: Longitudinal comparison of website usefulness and satisfaction. International Journal of Market Research, 61(2), pp.178-194.

[25] PUSZTOVÁ, Ľ. and Babic, F., 2020. Performance Assessment of Different Classification Methods for Coupon Marketing in E-Commerce. Acta Electrotechnica et Informatica, 20(3), pp.11-16.

[26] Reinartz, W. J. and Kumar, V. (2003) 'The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration', Journal of Marketing, 67(1), pp. 77–99. doi: 10.1509/jmkg.67.1.77.18589.

[27] Roos, I. and Gustafsson, A. (2007) 'Understanding frequent switching patterns', Journal of Service Research, 10(1), pp. 93–107. doi: 10.1177/1094670507303232.

[28] Seo, D., Choi, S. and Yoo, Y., 2022. Prevention of Customer Churn Due To Issuance of Real-Time Coupons Based on Deep Learning.

[29] Shetty, P.P., Varsha, C.M., Vadone, V.D., Sarode, S. and Kumar, D.P., 2019. Customers churn prediction with RFM model and building a recommendation system using semi-supervised learning in retail sector. Int. J. Recent Technol. Eng, 8(1), pp.3353-3358.

[30] Song, L. and Yang, W.S., 2018, May. A Novel Digital Coupon Use Prediction Model Based on XGBoost. In 4th Annual International Conference on Management, Economics and Social Development (ICMESD 2018) (pp. 775-782). Atlantis Press.

[31] Taylor, G. A. (2001). Coupon response in services. Journal of Retailing, 77(1), 139–151.

[32] Thomas, J. S. (2001) A methodology for linking customer acquisition to customer retention, Journal of Marketing Research. Available at: https://search.proquest.com.ezproxy.cc.lut.fi/docview/235217043/fulltextPDF/FCE0D9F7E5E64C59PQ/1?accountid=27292 (Accessed: 16 April 2019).

[33] Yanfang, Q. and Chen, L., 2017, December. Research on E-commerce user churn prediction based on logistic regression. In 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 87-91). IEEE.