



**ISSN Online 2617-3573**

 **Stratford**  
Peer Reviewed Journals & books

**A Robust Predictive Model for Customer Churn at City  
Breweries Using Machine Learning Techniques**

**Michael Abednico Chirume & Barbara Zanele Muchenje &  
Aaron Nyoni**

**ISSN: 2617-3573**

# A Robust Predictive Model for Customer Churn at City Breweries Using Machine Learning Techniques

<sup>\*1</sup>Michael Abednico Chirume, <sup>2</sup>Barbara Zanele Muchenje & <sup>3</sup>Aaron Nyoni

<sup>1</sup>Department of Operations Research and Statistics, National University of Science and Technology (NUST), Bulawayo, Zimbabwe

<sup>2</sup>Department of Operations Research and Statistics, National University of Science and Technology (NUST), Bulawayo, Zimbabwe

<sup>3</sup>Department of Operations Research and Statistics, National University of Science and Technology (NUST), Bulawayo, Zimbabwe

\*Email of the Corresponding Author: [michael.chirume@nust.ac.zw](mailto:michael.chirume@nust.ac.zw)

How to cite this article: Chirume, M., A., Muchenje, B., Z. & Nyoni, A. (2026). A Robust Predictive Model for Customer Churn at City Breweries Using Machine Learning Techniques. *Journal of Information, Technology and Data Science*, 10 (1), 118-133. <https://doi.org/10.53819/81018102t3189>

## Abstract

Customer churn presents a significant threat to business sustainability, particularly in the beverage industry where competition and shifting consumer preferences are intensifying. This research investigates customer churn at City Breweries, a brewery in Zimbabwe, with the aim of developing a robust predictive model using machine learning techniques. The study utilizes historical sales and operational data such as dispatch volumes, returns, mileage, and delivery routes to identify patterns associated with customer churn. Multiple classification algorithms, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), AdaBoost, Extra Trees, and XGBoost, were implemented and compared using key performance metrics, accuracy, precision, recall, and F1-score. The results highlight XGBoost as the most effective model for churn prediction due to its ability to handle class imbalance, capture non-linear relationships, and deliver superior predictive performance. While models like Logistic Regression and Decision Trees offer simplicity and interpretability, they fall short in capturing complex interactions within the data. XGBoost's regularization mechanisms and gradient boosting framework enhance its generalization capabilities, making it well-suited for real-world deployment. This study provides actionable insights into customer retention by identifying customers with a higher risk of churning with high accuracy. It also contributes to the existing body of knowledge by demonstrating the efficiency of ensemble machine learning models in the beverage sector, an area which is under explored in churn prediction research.

**Keywords:** Robust Predictive, Model, Customer Churn, City Breweries, Machine Learning Techniques

## **1.0 Introduction**

For many businesses, client attrition or churn is a major problem. When a consumer cancels their subscription and moves to a rival, this is known as customer churn. The decision of a consumer to switch to a different service provider is influenced by numerous factors. These elements typically have to do with the price and calibre of the service. When customer churn reaches specific thresholds, it results in significant losses in profits and companies know that maintaining existing customers is significantly less expensive than attracting new ones. The purpose of customer churn prediction (CCP) models is to determine which customers are most likely to attrition [8]. By doing this, these models allow companies to lower churn costs and increase the efficacy of client retention programs [28]. Machine learning is a branch of artificial intelligence that allows systems to identify patterns, learn from data, and make judgements with little or no human involvement [17]. It entails the application of algorithms capable of analysing “big” data sets, identifying patterns, and forecasting results using past data. Support Vector Machines and Random Forest are some techniques that can be used to do this. Businesses can utilize this to identify the main elements affecting client loyalty and put tailored marketing plans into place, which will enhance their efforts to retain customers. In addition to improving consumer engagement, this strategy will increase the company’s overall profitability and overcome a major knowledge gap in beverage business churn forecasting.

The beverage industry is a dynamic and quickly changing field. It includes a broad variety of goods, such as soft drinks, bottled water, alcoholic beverages, and drinks with a health focus. Customers are an organization’s most precious asset in the highly competitive business world of today [7]. In addition to drawing in visitors, businesses also need to interact with them, turn them into clients, and keep them around. Developing a solid connection with clients is essential to keep them from leaving [11]. This entails providing goods and services that are suited to their preferences and continuously working to meet their needs [10]. In recent years, City Breweries experienced some significant transformation in sales. This was due to increased competition and changing customer preferences [15].

These preferences have recently been increased by a lot of factors such as sustainability concerns and consumers being health-conscious [26]. Many brewers and beverage firms, like City brewers, struggle to keep their current clientele while also recruiting new ones since consumers are becoming pickier and demanding a greater range of products. Businesses risk losing market share if they don’t adjust to these expectations [22]. This issue has grown to be a major worry since it can significantly impact a business’s profitability and standing in the market [11]. To sustain profitability and market share, the industry must comprehend the elements that lead to customer attrition when faced with intensified competition and constantly changing consumer preferences. Through the creation of a predictive model especially for City Breweries, this study seeks to solve the problem of customer attrition, which is especially prevalent in the beverage business. The researcher intends to use machine learning techniques to examine the company’s historical data, pinpoint the primary reasons for staff loss, and implement targeted retention strategies. By correctly identifying which consumers are most likely to leave, City Breweries may adjust their product offerings and proactively interact with them through targeted marketing campaigns. In the end, this will increase client loyalty and boost total profitability. This project not only seeks to provide actionable insights for City Breweries but to also add to the literature on customer retention strategies within the beverage sector.

## **1.1 Problem Statement**

When consumers or subscribers abandon a business or service, this is known as customer churn [3]. It happens when clients leave a business because they are unhappy or because of competitive circumstances, including the launch of new goods. For businesses to assess, this problem may even be exacerbated by a decrease in the frequency of customer purchases [24]. Intentional churners are customers who choose to discontinue purchases. While unintentional churn happens when a company or service terminates a customer's access for reasons such as late payment, nonpayment for goods, or improper use of the business's data and information, clients that intentionally sever connection are known as purposeful churners [12]. Customer attrition has had a major impact on City Breweries and presented a big obstacle to sustaining the business's profitability. It's possible that outside variables like the COVID-19 outbreak made consumer attrition worse [13]. This is also true in the case of City Breweries as COVID-19 altered customer preferences and behaviours. City Breweries must comprehend the root reasons of client attrition since high churn rates not only reduce revenue but also hinder brand loyalty. Predicting which consumers are most likely to leave is one way to deal with this problem. This paper aims to develop a robust predictive model for customer churn at City Breweries using machine learning techniques.

The aims of the study are to identify the key factors contributing to customer churn in City Breweries through comprehensive data analysis, while one of the central goals is to develop a predictive model that the company can implement to support proactive customer retention strategies. In addition, the research purposes include comparing, analyzing, and assessing the effectiveness of various machine learning algorithms, such as Random Forest and Support Vector Machines, in predicting customer churn for City Breweries. These desired outcomes will provide the company with data-driven insights and practical tools to enhance customer loyalty and long-term growth. This study is essential because it may help City Breweries and the beverage sector as a whole deal with serious problems related to customer retention. The possibility that customers will switch companies and stop buying their goods or services is investigated to analyse the customer churn rate [19]. By offering insights into consumer needs and behaviours, CCP seeks to increase satisfaction for businesses and consumers alike [9]. Building, maintaining, and strengthening long-lasting relationships with clients is another objective of CCP. When a current company satisfies the needs of its clients, it retains them. Every firm wants to increase customer retention, upsell and cross-sell to existing clients, and bring on new ones. It is more economical to keep existing clients [24] because it can cost businesses more than twenty times as much to acquire new clients as it does to keep existing ones in today's crowded and fiercely competitive marketplaces.

## **2.0 Literature Review**

In the brewery sector, churn is a major obstacle that can significantly affect how well the company performs. If rate of churn is low, a company has large, stable customers, when it is not, the company may experience issues with its clientele. This means a business should strive for the lowest churn rate. As the business becomes more competitive, breweries need to come up with solutions to keep their customers and figure out why they are leaving. This makes predicting client attrition essential. This is made feasible by using machine learning techniques to develop a good model, which will ultimately assist the firm make better decisions and boost income.

The success of rivals, price increases, or poor customer service are just a few of the many reasons why customers leave. Calculated mathematically, the customer churn rate is the proportion of

consumers who discontinue using a business's goods or services within a specified time frame. Since a high customer turnover rate implies that current customers are leaving the organization, it is a bad sign for the enterprise. A churned customer may be an unhappy client who could harm a business's standing. In addition to losing the money spent by those departing consumers, the business may receive negative publicity that harms its reputation.

Customer turnover is particularly significant in the brewery sector, this is because brand loyalty plays a big impact in the sales of a firm and its market position. Changes in consumer preferences, economic situations, bad marketing, product quality, and product availability are all factors that contribute to churn and recognizing these trends can assist breweries in creating focused plans to reduce revenue loss. One popular strategy to maintain a low customer churn rate is to develop a prediction model to find probable churning consumers. Once the prediction model has identified those customers, the business may offer better customer service to keep those customers. The business can reduce the operational costs associated with acquiring new clients as a result. Data analytics is essential for anticipating and reducing client attrition to successfully handle this difficulty [16]. The systematic analysis and interpretation of data to provide significant insights is known as data analytics. When used in e-commerce, it can provide invaluable insights about customer behaviour, preferences, and churn-influencing factors. Enterprises must comprehend the role that data analytics plays in forecasting client attrition. Consumers progressively go from being devoted to regular clients over time rather than abruptly. Thus it is important to identify trends in customer behaviour. Traditional methods of addressing churn, such as surveys and historical sales analysis are often narrow focused. However, predictive analysis enables businesses to identify customer departures by identifying patterns.

Businesses must understand the underlying reasons of churn to find churners. It is even more important if a business wishes to go one step further and anticipate potential future churners. The application of machine learning techniques is essential since these characteristics are typically complicated, multifaceted, and sector-dependent. Therefore, it should come as no surprise that churn analysis techniques based on machine learning are used in many industries, such as e-commerce and banking [18]. Decision trees and other machine learning techniques have revolutionized churn prediction by providing advanced data analysis capabilities. These methods extract valuable insights from customer interactions, purchasing habits and demographics. Unlike traditional statistical models, machine learning adapts to evolving consumer trends, making churn prediction more dynamic and actionable. We may find ourselves unable to decipher the data after examining it, when that occurs, we use machine learning. Because there are so many data sets available, machine learning is becoming a necessity. To extract pertinent data, machine learning is used in many sectors. Using data to learn is the goal of machine learning. By using machine learning based models, breweries can optimize customer retention efforts and gain a competitive advantage.

Notable studies have applied Machine Learning (ML) algorithms, with particular focus on ensemble learning models, to tackle the CCP issue. [34] examined consumer churn prediction using machine learning models, including ensemble learning models with a focus on boosting approaches, Decision Trees (DT), Artificial Neural Networks (ANN), and Logistic Regression (LR). The study examined 47,000 local users' data, considering factors including credit, billing, and consumption history. The study's findings demonstrated that the ANN model outperformed DT and LR in terms of subscriber churn predicting [16] carried out a thorough analysis of feature selection techniques to improve CCP solutions. The study examined popular approaches, such as

<https://doi.org/10.53819/81018102t3189>

artificial neural networks, decision trees, and regression analysis. It identified three main advantages of feature selection, compatibility with lower-dimensional datasets, which is especially advantageous for neural networks; decreased processing time; and improved classification accuracy through the selection of relevant features. Additionally, the researchers emphasized the value of optimization methods such as genetic algorithms in enhancing CCP performance and recommended further research into fuzzy inference systems and Naive Bayes for churn prediction [10] provided a thorough analysis of SVM models in Customer Churn Prediction, contrasting grid-search with cross-validation to find the best parameter tuning.

In terms of prediction accuracy, the study contrasted Support Vector Machine models with Random Forest (RF) and Linear Regression techniques. Although Random Forest fared better than Support Vector Machine, the results demonstrated Support Vector Machine's resilience to noisy data and its superiority over LR when parameters were properly adjusted. [9] examined different dataset balancing tactics, such as random and complex under-sampling techniques, to address class imbalance in Customer Churn Prediction datasets. Weighted RF and gradient boosting models were used, and results evaluated using AUC and Lift. Advanced under-sampling algorithms performed better than random ones, according to the data [35] presented two ANN models that use the Genetic Algorithm (GA) to improve the accuracy of churn predictions. When applied to real-world cellular service datasets, the models demonstrated higher performance measures including accuracy, lift decile over 10%, and AUC. One model was based on a cross-entropy criterion for CCP, while the other model used GA to directly improve accuracy. Additionally, the study discovered that the cross entropy-based model had greater resilience to overfitting, and that medium-sized neural networks produced the greatest outcomes [20] shown that ensemble models can solve the CCP problem more effectively, especially those that use boosting techniques and Decision Trees.

However, churn datasets provide difficulties due to their size, imbalance, and complex feature sets. To overcome them, the study suggested combining genetic programming with Adaboost, which demonstrated better prediction accuracy than KNN and RF techniques [38] have out a thorough comparison of ML techniques for CCP. Using Cross-validation on a publicly available data set, the study implemented and evaluated popular models, like DT, ANN, Naive Bayes, SVM, and LR. The study also investigated how using ensemble learning with boosting approaches can improve performance. [20] presented FW-ECP, an intelligent forecasting system that combines ensemble learning approaches with wrapper and filter feature selection methods. The system used the mRMR approach for feature selection, the Particle Swarm Optimization for data set balancing, and the GA to remove redundant and unnecessary features. The capacity of the meta-heuristic algorithms to handle unbalanced and extensive training sets allowed for the higher performance of two ensemble learning-based predictors that were proposed and evaluated on two public datasets. [21] To solve the CCP problem, a thorough evaluation of seven classifiers as well as three distinct sampling strategies was carried out.

Using telecom datasets and four assessment measures, the study investigated the effectiveness of sample strategies and hyperparameter tweaking. According to the results, CatBoost performs better than its machine learning peers, achieving an astounding F1-score of 93% following the use of Optuna hyperparameter tuning. In terms of ROC AUC metric evaluation, XGBoost and CatBoost both achieve remarkable 91%. While CatBoost achieves this high-performance record with implementing Optuna hyperparameter tuning, XGBoost's achievement is reached after integrating SMOTE with Tomek Links [22] examined how well machine learning algorithms predicted

<https://doi.org/10.53819/81018102t3189>

telecom churn. Using a publicly available dataset, it examines algorithms such as Random Forests, XGBoost, LightGBM, and CatBoost and assesses their performance using measures like ROC AUC, F1-score, Accuracy, Precision, and Recall. The study shows how well SMOTE and ADASYN upsampling methods handle dataset imbalance, with ADASYN outperforming SMOTE in terms of performance improvement. With an F1-score of 89% and ROC AUC of 95%, LightGBM stands out after ADASYN, demonstrating the effectiveness of sophisticated boosting algorithms and smart upsampling in tackling the difficulties of customer churn prediction.

[5] presented a two-layer soft voting model and examined its performance on both balanced and imbalanced public telecom datasets to meet the CCP challenge. The model, which outperformed all base learners and included ensemble classifiers in the second layer and traditional machine learning techniques in the first, showed a notable improvement in the F1-score, particularly on balanced datasets. [37] offered a comparison of machine learning methods for predicting customer attrition, using logistic regression, support vector machines, decision trees, and naive bayes. They then looked at how boosting methods affected the categorization accuracy. SVM-POLY with AdaBoost outperformed the others in the results that were achieved. However, using feature selection techniques like univariate selection and others can further increase the classification accuracy. [7] employed three machine learning methods Bayesian networks, support vector machines, and neural networks to forecast client attrition. Feature selection considers principle component analysis (PCA), which shrinks the dimensions of the data. Nonetheless, the feature selection procedure might be improved by utilizing optimization strategies that raise classification accuracy. The performance evaluation made use of the gain measure and ROC curve.

### **3.0 Methodology**

This chapter sets out our project's methodological framework. This section will describe a few techniques that City Breweries can employ to apply machine learning algorithms to forecast client attrition. It will offer a thorough algorithm for analysing and characterizing customer churn patterns using the approaches. It is structured into several components, data collection, data preprocessing, model development, evaluation, and implementation. It will preview models that will be used to predict customer churn. The researcher will use logistic regression, Random Forest, and Support Vector Machines etc. Logistic regression will be used to determine the relationship between churn and its independent variables. This will result in a regression logistic equation with one dependent variable (customer churn), with some sets of independent variables such as ullages, routes, dispatch volumes, returns, time, distance. This study was conducted at City Breweries, which primarily operates within Zimbabwe. It is a commercialised business entity owned by the City of Bulawayo and specialises in traditional beer and other products. City products are available in the Bulawayo Metropolitan, Matabeleland, Midlands, and Harare Metropolitan provinces. This study will encompass various geographic regions where City distributes its products, such as Mberengwa, Tsholotsho, City Centre, Liquor Hub.

#### **3.1 Data Preprocessing**

Data preprocessing is crucial to ensure the quality of the data before analysis. It transforms raw data into a clean and usable format. This ensures that data is suitable for analysis and modelling. While cleaning data, some missing values are encountered, and missing data can significantly affect the performance of machine learning models. Techniques such as mean/mode imputation or removal of records with missing values will be employed [23]. Feature scaling is essential to ensure that all features contribute equally to a model's performance. This can be done by standardizing

numerical values to a common scale without distorting differences in the ranges of values.

### 3.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis frequently uses data visualization techniques for analysing, researching, and summarizing data collections [1]. Finding patterns and identifying abnormalities is made simpler with EDA, which helps decide how to best alter data sources to acquire the answers you need [25]. To ensure correct predictions, EDA can be used in the following ways: Using histograms to visualise the distribution of numerical features such as Ullages, Returns, and Mileage and Hours to identify skewness or outliers. Visualising categorical variables such as Zones and Regions using bar charts helps to understand their impact on churn rates. Using group by operations to calculate average churn rates for different categories e.g. by Dispatch or Zone. This helps to identify segments that may be at higher risk of churn. As a result, the variables in the data set and their relationships are better understood.

### 3.3 Feature Selection

Feature selection is the process of isolating the most consistent, non-redundant, and relevant features to use in model construction [4]. Identifying relevant features that significantly influence customer churn is important for improving model performance, and we can use the correlation analysis. Correlation analysis is used for spotting patterns within datasets. Calculating the correlation coefficient assesses the linear relationship between features (Ullages, returns, etc) and the target variable(churn). The correlations series will show the Pearson correlation coefficient between the churn variable and each numerical feature. A negative correlation (-1) indicates that as one variable falls, the other rises, whereas a positive correlation (+1) indicates that both variables rise in proportion to one another.

### 3.4 Select Kbest Feature

Select KBest is a univariate method that retains only the specified number of highest scoring features. It employs statistical tests such as the chi-squared test, ANOVA F-test, or mutual information score to rank the features based on their relationship with the output variable. Then, it selects the top K features with the highest scores for inclusion in the final feature subset. Select KBest is user-friendly and efficiently reduces the feature set to a manageable number, which is particularly crucial when dealing with large datasets. Data splitting is the process of separating a dataset into discrete subsets to make machine learning model testing, evaluation, and training easier [20]. Through data separation, we guarantee that the model is evaluated on a final, independent set after being trained on one and validated on another. Accurately assessing model performance is its purpose. The data is split into training and testing sets, with 80% of the data allocated for training and 20% for testing. Decision Tree, Random Forest, Support Vector Machine, Logistic Regression XGBoost, AdaBoost, and Extra Trees were chosen as the seven robust and practical classifiers. To identify the most influential features in the datasets, feature selection methods were performed. SelectKBest feature selection method was also utilized.

### 3.5 Decision Tree

A decision tree is a tree-shaped structure that uses inductive learning to extract classification rules from a training dataset. It sequentially divides a big dataset into smaller sets [9]. Class labels like "churner" or "non-churner" are represented by the leaves in this structure, while the relationships between the variables that result in these class labels are shown by the branches. A decision tree frequently achieves excellent accuracy in churn prediction issues, despite its incapacity to capture

<https://doi.org/10.53819/81018102t3189>

complex and nonlinear interactions among factors [9]. The two primary steps in the decision tree improvement process are construction and pruning. The dataset is recursively divided during the building phase until every sample in each partition has the same values. To improve the tree structure, some branches are cut off during the pruning phase. Usually, these are the branches with the highest computed error rate or that include noisy data [21].

### 3.6 Random Forest

The subspaces and bagging components of decision trees are randomly combined in Random Forest, which is made up of several tree topologies. Random Forest chooses a selection of variables at each node of the tree, then determines the optimal split between those variables for that node. Random Forest builds the training dataset for every tree by fitting them to bootstrap resampled datasets and de-correlating them using the bagging approach. After splitting each tree for the next partition, this approach chooses a random sample of variables. Consequently, the variance component of the output error is decreased if many comparable data sets are produced via bootstrapping, or resampling with replacement, and regression trees are developed without pruning and averaged [6]. The majority vote of each individual tree serves as the foundation for the findings derived from Random Forest. Random Forest overcomes some of the drawbacks of non-ensemble Decision Trees, like overfitting and lack of resilience, by using the bagging technique with a huge number of trees [5].

### 3.7 Support Vector Machine

A support vector machine uses an N-dimensional space to identify the best line or hyperplane for classifying data by maximizing the distance between each class [27]. Churn statistics are inherently unbalanced, with the class of interest that is the quantity of churning customers appearing infrequently across the dataset in comparison to the other class. As a result, inappropriate evaluation measures may be used. Compared to other conventional classification and prediction methods, SVM is a potent prediction method with numerous benefits [2]. One of the most significant benefits is that SVM has a significant computational advantage because it only needs a tiny portion of the dataset to solve the problem. SVM offers several unique benefits, including high accuracy (globally optimal), speed (it can operate quickly), strong generalization, and less susceptibility to overfitting issues.

### 3.8 XGBoost

A supervised classifier made for tabular or structured datasets is called XGBoost. It uses a gradient-boosting Decision Tree to increase accuracy, decrease memory usage, and speed up computation time [14]. Gradient boosting is a technique that effectively calculates mistakes by using residuals and new algorithm models. Together, these elements make the most use of the resources already available for training the prediction classifier. Furthermore, XGBoost reduces or minimizes the loss function by using gradient descent [18].

### 3.9 Adaptive boosting (AdaBoost)

AdaBoost is an adaptive supervised classifier that combines several weak or basic learners, like Decision Trees, to create a powerful learner for data classification. AdaBoost has shown successful in combining strong base learners, producing a more accurate model, despite being usually used to combine weak base learners. In order for AdaBoost to work, instances in the training dataset are weighted according to the precision of earlier classifications. AdaBoost can iteratively enhance performance by concentrating on misclassified instances in later iterations because to its adaptive

<https://doi.org/10.53819/81018102t3189>

strategy [29].

**3.10 Extremely randomized trees (Extra Trees)**

Decision Trees serve as the primary foundation for the classifier Extra Trees. Extra Trees, like Random Forest, uses randomization to reduce overfitting and overlearning from the data by varying some decisions and data subsets. Extra Trees and Random Forest are similar in that they both use random subsets of characteristics for node splits and construct several trees, but they are different in two important ways. Its first samples without replacement because it does not bootstrap observations. Second, random splits, not the best splits, are used to divide nodes. Furthermore, Extra Trees just randomly selects features, as opposed to Random Forest, which randomly selects all samples for both features and observations. This random splitting technique frequently produces somewhat better results than Random Forest [9].

**4.0 Findings and Results**

**4.1 Performance Evaluation Matrix**

In this study, the proposed churn prediction model is evaluated using accuracy, precision, recall, F1-score, and ROC area.

Equation 6.1 calculates the accuracy metric. It identifies a few instances that were correctly classified.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$
6.1

Here “TN” stands for True Negative, “TP” stands for True Positive, “FN” stands for False Negative and “FP” stands for False Positive. TP Rate is also known as sensitivity. It tells us what portion of the data is correctly classified as positive.

For any classifier, the TP rate must be high. TP rate is calculated by using Equation 6.2.

$$\text{TP Rate} = \frac{\text{True Positives}}{\text{Actual Positives}}$$
6.2

FP Rate tells us which part of the data are incorrectly classified as positive. The result of the FP rate must be low for any classifier. It is calculated by using Equation 6.3.

$$\text{FP Rate} = \frac{\text{False Positives}}{\text{Actual Negatives}}$$
6.3

Precision, also known as Positive Predictive Value (PPV), indicates which part of the prediction data is positive. It is calculated by using Equation 6.4.

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})}$$
6.4

The recall is another measure for completeness i.e. the true hit of the algorithm. It is the probability that all the relevant instances are selected by the system. The low value of recall means many false negatives. It is calculated by using Equation 6.5.

$$\text{Recall} = \frac{(\text{True Positive})}{(\text{True Positive} + \text{False Negative})}$$
6.5

The F1-Score value is a trade-off between correctly classifying all the data points and ensuring that each class contains points of only one class. It is calculated by using Equation 6.6.

$$F - \text{measure} = 2 * \left( \frac{Precision * Recall}{Precision + Recall} \right)$$

6.6

ROC area denotes the average performance against all possible cost ratios between FP and FN. If the ROC area value is equal to 1.0, this is a perfect prediction. Similarly, the values 0.5, 0.6, 0.7, 0.8 and 0.9 represent random prediction, bad, moderate, good, and superior respectively. Values of ROC areas other than these indicate something is wrong.

**Table 1:Accuracy on Various Churn Dataset**

| Method Used            | Accuracy | Precision | Recall | F1-Score |
|------------------------|----------|-----------|--------|----------|
| Logistic Regression    | 0.50     | 0.50      | 1.00   | 0.67     |
| Random Forest          | 0.45     | 0.47      | 0.80   | 0.59     |
| Support Vector Machine | 0.50     | 0.50      | 0.90   | 0.64     |
| AdaBoost               | 0.48     | 0.53      | 0.90   | 0.67     |
| Extra Trees            | 0.50     | 0.50      | 0.80   | 0.62     |
| XGBoost                | 0.60     | 0.56      | 0.90   | 0.69     |
| Decision Tree          | 0.55     | 0.54      | 0.70   | 0.61     |

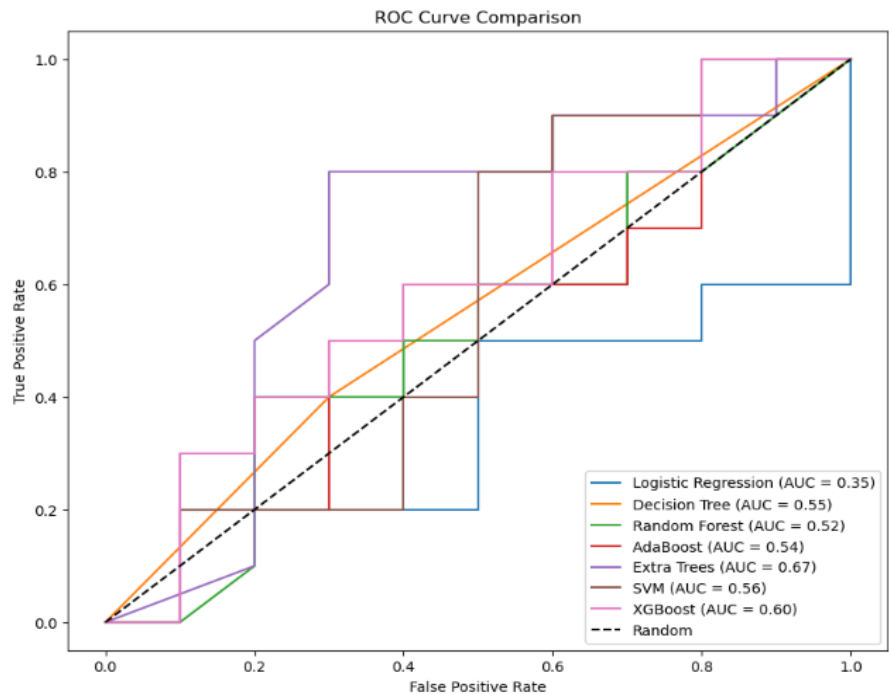


Figure 1: Feature Selection Scores

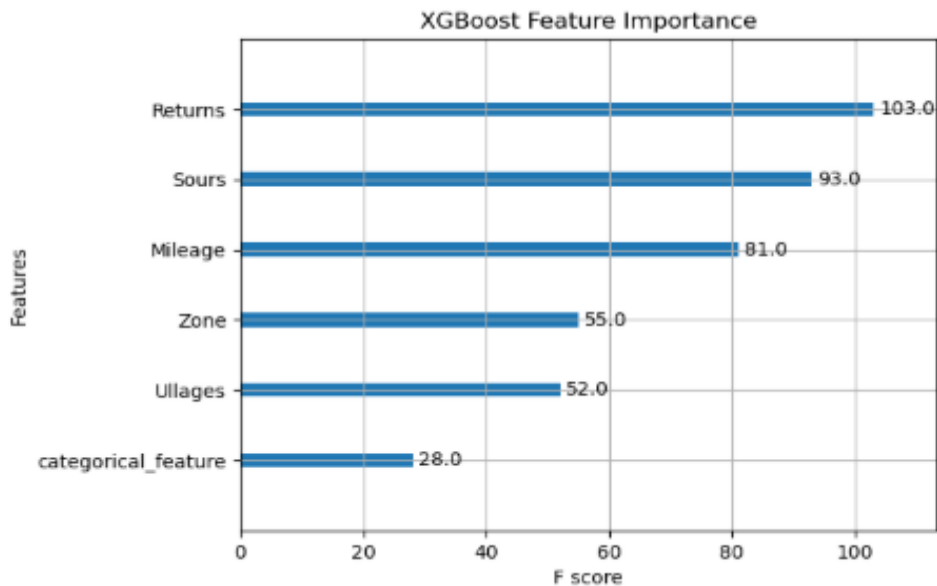


Figure 2: XGBoost Feature Importance

The best model is XGBoost with an accuracy of 60% and AUC-ROC of 60%. The Feature Selection Scores for the XGBoost model revealed that Returns (103), Sours (93) and Mileage (81) are the most influential features, and significantly contributed to the model’s predictions. Zone (55) and Ullages (52) had moderate importance, while categorical feature (28) had minimal influence. This indicated that the model heavily relied on Returns, Sours and Mileage for making accurate predictions, whereas Zone and the Ullages played a lesser role. These insights can help prioritize features for future model improvements or further analysis. From this pattern and

<https://doi.org/10.53819/81018102t3189>

behaviour, we make rules for the recommendation of only similar customers in the future.

## **4.2 Customer Retention**

The impact of retention is positive on a company profile that ultimately enhances the retention and marketing performance of the companies. Through this strategy, the company can understand the behaviour of customers and retain the customers by offering suitable packages to a specific group of customers. This research work indicates the important factors of customers, their behaviour and addresses some important issues which are valuable for companies. These issues are described below.

The research identified the key factors contributing to customer churn in City breweries through comprehensive data analysis.

It created a predictive model that can be implemented by City breweries to facilitate proactive customer retention strategies.

The research compared, analysed and assessed the effectiveness and the performance of various machine learning algorithms in predicting customer churn for City Breweries.

From results and customer behaviours, we can also achieve “one-to-one marketing” which emphasizes marketing with individual customers.

To fulfil the needs of customers, their unique behaviour and preferences are identified to offer them a variety of packages and services.

City Breweries needs to build a specific relationship with customers considering their identified needs and behaviour to build a strong relationship with them.

A system should be integrated between customers and operational front-line system to effectively manage customer’s needs. For retaining customers, introducing an integrated system could be effective across the channel for offering different services such as productivity and efficiency services, customized services, and price control services. In this context, City Breweries is concerned with personalizing the relationship with the customers and organization response to the customer’s satisfaction such as developing a collaborative communication system, extending front offices to incorporate all employees, partners, and suppliers to interact with customers through email, telephone, web pages, contacts, text, etc. Through segmentation, the customer’s behaviour is monitored and tracked to identify the patterns and usage. A direct marketing method is a retaining process through which customers are retained, and a variety of services are offered through many channels. Direct methods are more effective which can decrease the cost and increase productivity by offering different services and personalized offers directly. In the direct interaction with customer’s, the organization gets a better response and can immediately take a decision. From the customer’s response, the organization can also develop a model that can estimate likelihood of churn, identify similar groups of customers based on their responses and is more beneficial for decision makers.

## **5.0 Conclusion**

In the present beverage competitive market, churn prediction is a significant issue to retain valuable customers by identifying a similar group of customers and providing competitive offers/services to the respective groups. Therefore, in this domain, the researcher looked at the key factors of churn to retain customers and solve the problems of City Breweries and decision maker

of a company. In this study, a customer churn model is provided for data analytics and validated through standard evaluation metrics. The study shows that XGBoost performs better in predicting customer churn problems as it shows better F1 metric and mitigates the issue of misjudging the choices of customers who should have exited made by the other models. In addition, this paper also studies the three most important factors in the XGBoost, these are, returns, sours and mileage. The results show that these factors are the reason more people are likely to churn, which can be used to provide some advice for City Breweries. Overall, this research exemplifies the advantages of XGBoost in predicting customer churn and, based on the results, provides three user characteristics that can be used for more in-depth research.

In future study, researchers can focus on analysing questions like why some specific features are of high importance in prediction and why the mechanism of operation of decision trees makes it more effective in predicting churn. The project emphasizes that the effective solution lies in retaining current customers using customer churn prediction models. Machine Learning techniques are highlighted as instrumental in this regard, as they enable the identification of customers that are on the verge of churning, allowing for targeted retention strategies. The project begins by underscoring the severe threat of customer churn to breweries, supported by statistical evidence. It provides an extensive review of the customer churn problem, both in a broad context and then more specifically within the beverage sector. A thorough comparison of selected articles is then presented. The research findings contribute valuable insights to the field of customer churn predictive modelling in the beverage sector. By doing so, the project delineates a clear pathway for researchers, offering an overview and accumulating trends about machine learning applications in the industry. This not only aids in understanding the current landscape but also guides future research endeavours in this critical area.

## References

- [1] ABBAS, W., USMAN, M., AND QAMAR, U. (2022). Churn prediction of customers in a retail business using exploratory data analysis. In 2022 International Conference on Frontiers of Information Technology (FIT). IEEE, pp. 130–135.
- [2] ABDULLAH, D. M. AND ABDULAZEEZ, A. M. (2021). Machine learning applications based on some classification a review. Qubahan Academic Journal, 1 (2), 81–90.
- [3] AHN, J., HWANG, J., KIM, D., CHOI, H., AND KANG, S. (2020). A survey on churn analysis in various business domains. IEEE access, 8, 220816–220839.
- [4] BARRAZA, N., MORO, S., FERREYRA, M., AND DE LA PENA~, A. (2019). Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study. Journal of Information Science, 45 (1), 53–67.
- [5] BEEHARRY, Y. AND TSOKIZEP FOKONE, R. (2022). Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry. Concurrency and Computation: Practice and Experience, 34 (4), e6627.
- [6] BOGAERT, M. AND DELAERE, L. (2023). Ensemble methods in customer churn prediction: a comparative analysis of the state-of-the-art. Mathematics, 11 (5), 1137.
- [7] BRANDUSOIU, I., TODEREAN, G., AND BELEIU, H. (2016). Methods for churn prediction in the pre-paid mobile telecommunications industry. In 2016 International conference on communications (COMM). IEEE, pp. 97–100.
- [8] BUHLMANN ~, P. AND YU, B. (2002). Analysing bagging. The annals of Statistics, 30 (4), 927–961.
- [9] BUREZ, J. AND VAN DEN POEL, D. (2009). Handling class imbalance in customer churn prediction. Expert Systems with Applications, 36 (3), 4626–4636.
- [10] COUSSEMENT, K., BENOIT, D. F., AND VAN DEN POEL, D. (2010). Improved marketing decision making in a customer churn prediction context using generalized additive models. Expert systems with Applications, 37 (3), 2132–2143.
- [11] DAS, G. (2024). BUSINESS ENVIRONMENT. kitab writing publication.
- [12] DEVRIENDT, F., BERREVOETS, J., AND VERBEKE, W. (2021). Why you should stop predicting customer churn and start using uplift models. Information Sciences, 548, 497–515.
- [13] EHSANI, F. AND HOSSEINI, M. (2024). Customer churn prediction using a novel meta-classifier: an investigation on transaction, telecommunication and customer churn datasets. Journal of Combinatorial Optimization, 48 (1), 7.
- [14] EHSANI, F. AND HOSSEINI, M. (2025a). Consumer segmentation based on location and timing dimensions using big data from business-to-customer retailing marketplaces. Big Data, 13 (2), 111–126.
- [15] EHSANI, F. AND HOSSEINI, M. (2025b). Customer churn analysis using feature optimization methods and tree-based classifiers. Journal of Services Marketing, 39 (1), 20–35.
- [16] HADDEN, J. (2008). A customer profiling methodology for churn prediction.
- [17] HAENLEIN, M. (2013). Social interactions in customer churn decisions: The impact of

<https://doi.org/10.53819/81018102t3189>

relationship directionality. *International Journal of Research in Marketing*, 30 (3), 236–248.

[18] HALTEH, K., ALKHOURY, R., ZIADAT, S. A., GEPP, A., AND KUMAR, K. (2024). Using machine learning techniques to assess the financial impact of the covid-19 pandemic on the global aviation industry. *Transportation Research Interdisciplinary Perspectives*, 24, 101043.

[19] HANIF, I. (2020). Implementing extreme gradient boosting (xgboost) classifier to improve customer churn prediction.

[20] IDRIS, A. AND KHAN, A. (2014). Ensemble based efficient churn prediction model for telecom. In *2014 12th International Conference on Frontiers of Information Technology*. IEEE, pp. 238–244.

[21] IMANI, M. AND ARABNIA, H. R. (2023). Hyperparameter optimization and combined data sampling techniques in machine learning for customer churn prediction: a comparative analysis. *Technologies*, 11 (6), 167.

[22] IMANI, M., JOUDAKI, M., BEIKMOHAMADI, A., AND ARABNIA, H. R. (2025). Customer churn prediction: A review of recent advances, trends, and challenges in conventional machine learning and deep learning.

[23] JONES, M. E., LOGAN, T. D., ROSE´, D., AND COOK, L. D. (2025). Competition and consumer discrimination. Technical report, National Bureau of Economic Research.

[24] KHODABANDEHLOU, S. AND ZIVARI RAHMAN, M. (2017). Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behaviour. *Journal of Systems and Information Technology*, 19 (1/2), 65– 93.

[25] KUHL ¨, N., SCHEMMER, M., GOUTIER, M., AND SATZGER, G. (2022). Artificial intelligence and machine learning. *Electronic Markets*, 32 (4), 2235–2244.

[26] LALWANI, P., MISHRA, M. K., CHADHA, J. S., AND SETHI, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, 104 (2), 271– 294.

[27] LARIVIERE B. AND VAN DEN POEL, D. (2004). Investigating the role of product features in preventing customer churn, by using survival analysis and choice modelling: The case of financial services. *Expert Systems with Applications*, 27 (2), 277–285.

[28] LEE, C. S., CHEANG, P. Y. S., AND MOSLEHPOUR, M. (2022). Predictive analytics in business analytics: decision tree. *Advances in Decision Sciences*, 26 (1), 1–29.

[29] LEMMENS, A. AND GUPTA, S. (2020). Managing churn to maximize profits. *Marketing Science*, 39 (5), 956–973.

[30] LE NGUYEN, K., TRINH, H. T., NGUYEN, T. T., AND NGUYEN, H. D. (2023). Comparative study on the performance of different machine learning techniques to predict the shear strength of deep beams: Model selection and industry implications. *Expert Systems with Applications*, 230, 120649.

[31] LIN, W.-C. AND TSAI, C.-F. (2020). Missing value imputation: a review and analysis of the literature (2006–2017). *Artificial Intelligence Review*, 53, 1487–1509.

[32] MANZOOR, A., QURESHI, M. A., KIDNEY, E., AND LONGO, L. (2024). A review on machine learning methods for customer churn prediction and recommendations for business practitioners. *IEEE access*.

<https://doi.org/10.53819/81018102t3189>

- [33] MARTINEZ, W. L., MARTINEZ, A. R., AND SOLKA, J. (2017). Exploratory data analysis with MATLAB. Chapman and Hall/CRC.
- [34] MOZER, M. C., WOLNIEWICZ, R., GRIMES, D., JOHNSON, E., AND KAUSHANSKY, H. (1999). Churn reduction in the wireless industry. *Advances in Neural Information Processing Systems*, 12.
- [35] PENDHARKAR, P. C. (2009). Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. *Expert Systems with Applications*, 36 (3), 6714–6720.
- [36] QI, X., YU, H., AND PLOEGER, A. (2020). Exploring influential factors including covid-19 on green food purchase intentions and the intention–behaviour gap: A qualitative study among consumers in a Chinese context. *International journal of environmental research and public health*, 17 (19), 7106.
- [37] SENTHILSELVI, A., KANISHK, V., VINEESH, K., AND RAJ, A. P. (2024). A novel approach to customer churn prediction in telecom. In *2024 International Conference on Advances in Computing, Communication and Applied Informatics (AC-CAI)*. IEEE, pp. 1–7.
- [38] VAFEIADIS, T., DIAMANTARAS, K. I., SARIGIANNIDIS, G., AND CHATZISAVVAS, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1–9.
- [39] VEISI, H. (2023). Introduction to svm. In *Learning with Fractional Orthogonal Kernel Classifiers in Support Vector Machines: Theory, Algorithms and Applications*. Springer, pp. 3–18.
- [40] VERBRAKEN, T., VERBEKE, W., AND BAESENS, B. (2012). A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE transactions on knowledge and data engineering*, 25 (5), 961–973.
- [41] WYNER, A. J., OLSON, M., BLEICH, J., AND MEASE, D. (2017). Explaining the success of adaboost and random forests as interpolating classifiers. *Journal of Machine Learning Research*, 18 (48), 1–33.