

Research Article

Data-Driven Churn Prediction: Evaluating ML Models for Business Retention Strategies

Ramshankar Nagarajan¹ , Siva Subramanian Raju^{2*}  and Athiraja Atheeswaran¹ 

¹Department of Computer Science and Engineering, Saveetha Engineering College, India

²Department of Computer Science and Engineering, SRM Institute of Science and Technology-Tiruchirappalli, India

Abstract: Customer churn, which is the loss of customers to a service, is a major threat to business profits, especially in competitive industries like telecommunications, finance, and e-commerce. Identification of at-risk customers early enough helps organizations to take proactive retention efforts, thus minimizing loss of revenues. The research study will focus on forecasting customer churn through the application of a machine learning (ML) algorithm to provide strategies to enhance customer retention strategies. A comparative analysis was done using five ML models: logistic regression, K-Nearest Neighbors, support vector machine (SVM), decision tree, and random forest (RF). To have a better understanding of the performance of the models, several metrics were applied, namely, accuracy, sensitivity (recall), specificity, precision, and F1 score. In terms of performance, RF showed the highest performance and the best accuracy (80.74%) and F1 score (61.34%), and the specificity and precision were high. SVM also demonstrated stable results, especially in sensitivity, which makes it appropriate for detecting churn-prone customers. The results indicate the advantage of ensemble algorithms such as RF because of their strong nature and generalization on complex data forms. This study addresses existing research gaps by standardizing data preprocessing, evaluation metrics, and model benchmarking, while incorporating multiple performance indicators beyond accuracy to effectively handle class imbalance and real-world decision-making trade-offs. The study further provides insights into model interpretability, deployment feasibility, and computational cost, enabling organizations to select churn prediction models based on both predictive performance and practical applicability.

Keywords: customer churn prediction, machine learning, random forest, support vector machine, predictive analytics

1. Introduction

1.1. Context and importance

Customer behavior has become very important in contemporary marketing strategies. In the current competitive markets, it is much more expensive to get a new consumer than retain an old one [1]. The analysis of customer behavior enables businesses to derive information about customer preferences, purchasing behavior, loyalty, and, most importantly, the churn probability. With the help of the behavioral data obtained from digital channels, businesses can recognize the changes in the customers' behavior in real time, which will allow them to adjust their strategies in response. Such a capacity to individualize services and efforts of engagement assists businesses to eliminate attrition and promote total customer satisfaction, which directly influences long-term profitability [2].

1.1.1. Relevance of churn prediction in modern businesses

Customer attrition (loss of current customers) is an ongoing problem in different kinds of businesses like telecommunications,

banking, retail, and subscription services. The high churn rates not only cause the direct loss of revenue but also increase the cost of acquiring new consumers. Proper prediction of churn will enable businesses to take early actions to correct customer dissatisfaction before it translates to attrition. By identifying the risks early, organizations can provide a more specific retention offer, such as a personalized offer, customer support intervention program, and loyalty program, thereby alleviating the cost of the acquisition process [3]. In the business environment where companies are trying their best to stay competitive, a good churn prediction (CP) system is no longer an added benefit but a must.

1.1.2. Economic implications of customer retention vs. acquisition

Retaining an existing customer is generally more economical than bringing in a new one, with research indicating that it can be five to seven times less expensive. Repeat customers not only provide repeat revenues but also become brand ambassadors as they tend to influence other people to use the company by spreading positive word of mouth. With customer retention, the businesses are guaranteed continuous revenues, reduced marketing costs, and higher customer lifetime value (CLV) [4]. Moreover, the process of acquisition is usually expensive in terms of advertisement, induction, and promotions. Thus, the financial benefits

*Corresponding author: Siva Subramanian Raju, Department of Computer Science and Engineering, SRM Institute of Science and Technology-Tiruchirappalli, India. Email: sivasubr1@srmist.edu.in

of customer retention are obvious: lower customer churn increases profitability and saves resources [5].

1.2. Challenges and why churn is hard

1.2.1. Shift from intuition-based to data-driven decision-making

In the earlier days, marketing was based on guesses and simple statistical tools. The proliferation of customer data and innovation in analytics have however led companies shift to a data-driven decision-making process. Companies can now discover previously unseen patterns in consumer behavior using information in web logs, transaction histories, social media, and customer relationship management (CRM) systems. The machine learning (ML) models can be especially useful in analyzing this huge amount of data and uncovering the insights that could have been missed by the traditional approaches [6]. Using data as the basis of decision-making, businesses will be able to optimize their strategy to better engage customers, retain, and predict churn.

1.2.2. Motivation for ML in churn prediction

ML has an incomparable benefit in the prediction of churn since it reveals nonlinear relationships between various factors of customer behavior. The classical approaches fail to perform well in the interpretation of complex patterns in large datasets. However, ML models have the capacity to learn using previous customer data and generalize to new examples, providing useful predictive power. Such models automate feature selection (FS), model training, and validation, by far speeding up the implementation of CP systems [7]. The ability to use real-time information to continuously update predictions is what makes machine learning an essential tool for companies that want to stay ahead of customer churn.

1.2.3. Strategic role of ML in customer retention

By incorporating the use of ML-driven CP into the strategy of a business, companies can become proactive and specifically target the retention of customers [8]. Due to the identification of the customers with high churn probability, businesses may focus retention activities, including personalized discounts, communication, or loyalty programs on specific customers. Besides assisting in retaining high-value customers, this data-driven method enables businesses to maximize the use of resources [9]. When CP models are properly developed, organizations have an opportunity to match their retention efforts to the CLV so that their resources are used to invest in the most profitable segments.

1.3. Research gap

Although there is a growing volume of literature on CP, a significant demand exists for comparative analysis of basic ML models [10]. The majority of the works concentrate on single models or do not have a consistent benchmarking strategy of commonly used algorithms. Also, the trade-offs between model complexity, interpretability, and computational efficiency have not been fully explored and are important in terms of real-world business applications. This paper aims to resolve that shortcoming by providing an analytical discussion covering five of the most popular ML models, namely, logistic regression (LR), K-Nearest Neighbor (KNN), support vector machine (SVM), decision tree (DT), and random forest (RF), based on a common experimental setting. This will give more insights into the practical applicability of each of the models, and thus, businesses can be

able to make informed decisions on which of the models best suits their business CP.

1.3.1. Problem statement

Although customer CP has become a popular topic in the academic literature, the problem is choosing the most suitable ML model that would be a balance among accuracy, interpretability, and computational efficiency. Models that can be used in business must be interpretable, scalable, and easy to implement, besides being able to predict churn with high accuracy. This paper focuses on the issue of model selection, giving a detailed comparison of classical and ensemble ML models on a shared dataset, with standardized evaluation criteria. The study will help businesses and researchers to find the best CP models based on practical aspects like performance, explainability, and computational resources.

1.3.2. Research gap and motivation

Despite extensive research, several limitations persist:

- 1) Lack of reproducible pipelines comparing multiple classical models under unified preprocessing.
- 2) Limited focus on interpretability and business deployment feasibility.
- 3) Sparse discussion of trade-offs between accuracy and cost.

This study addresses these gaps by benchmarking five widely adopted ML algorithms and discussing their interpretability and operational constraints.

In particular, the limited availability of real-world labeled churn datasets has contributed to a research landscape overly dependent on public benchmarks, underscoring the need for future cross-organizational validation.

1.4. Research objectives and contributions of this work

This study is aimed at achieving the following main objectives:

To develop a churn prediction (CP) model using a dataset containing customer behavior variables.

To apply and test five ML algorithms: LR, KNN, SVM, DT, and RF.

To measure the effectiveness of these models with standard evaluation measures: accuracy, sensitivity (recall), specificity, precision, and F1 score. To examine the balance between the interpretability of the models and predictive accuracy.

To present suggestions to the businesses regarding the best-suited models for forecasting churn based on their strategic objectives, like accuracy, transparency, and resource efficiency.

The outcome of the study will have a direct effect on the business decision-making process because it will enable the company to make feasible decisions concerning the advantages and limitations of each algorithm. Learning the real-world relevance of various CP models, the business can use the most appropriate one to enhance its customer retention strategy, better distribute resources, and increase profitability.

1.4.1. Contribution of the study

This research presents several contributions to both scholarly research and business practice:

An extensive overview of basic ML models in predicting customer churn.

Single pipeline to preprocess data, train the model, evaluate it, and make it reproducible.

The multi-metric evaluation framework, which gives a more comprehensive knowledge of the model performance.

Useful guidelines for businesses on the type of CP models to implement according to the computational cost, interpretability, and scalability.

A basis for future work that can investigate real-time updates, incremental learning, and behavioral drift detection on the CP systems.

2. Literature Survey

Customer churn prediction (CCP) poses a significant challenge in the telecommunications sector, and recent advancements in ML and artificial intelligence (AI) have introduced fresh possibilities. The suggested methodology involves six stages, among which are data preprocessing, feature analysis, FS, splitting data into the train and test sets, and the use of popular predictive models. Hyperparameter tuning is done by using k-fold cross-validation. Adaboost and XGBoost classifiers have the highest accuracy and Area Under the Curve (AUC), and thus, they perform better than others [11]. This paper examines the ML methods applied in the prediction of churn with 38 primary studies of 420 published between 2018 and 2021. The review points out limitations of churn models, namely, the inability to generalize to other areas of the industry, dependence on demographic, product-usage, and revenue characteristics, and absence of studies of information-rich customer–company interaction. It also brings to light the need for hybrid sampling approaches and a formal guideline for selecting appropriate evaluation metrics for models trained on imbalanced churn data. The study notes that there is a need to carry out more research in this sector [12]. With this study, it is proven that Deep-BP-ANN is an effective model for forecasting customer attrition in the telecom industry. The model employs Variance Thresholding and Lasso Regression in the FS, as well as early stopping to avoid overfitting. The model works in cases where the numbers of neurons are large and the regularization of the activities is done to reduce overfitting. It shows improved performance compared to existing ML techniques, and it performs better than existing deep learning methods based on holdout or 10-fold cross-validation on the same datasets [13]. This paper suggests a loss prediction model of B2C e-commerce business with the help of k-means customer segmentation and SVM prediction. The model enhances the indexes of prediction following the segmentation, which shows the importance of segmentation by k-means clustering. SVM prediction is more accurate than LR; thus, it is important in CRM [14]. This article aims at creating a credit card CCP model based on FS and five ML algorithms. Three models were applied to choose independent variables, whereas five models were chosen, namely, Bayesian networks, C5 trees, CHAID trees, CR trees, and neural networks. All the models were capable of forecasting the churn model, and the C5 tree model was the best. The three most important attributes required in the model were the number of transactions, revolving balance, and change in the number of transactions [15]. The study employs data mining and ML techniques to forecast and examine the CC based on a customer bank dataset. Heatmap analysis was used to identify factors that influence churn, and ML algorithms that were compared include RF, LR, Adaboost, and XGBoost. The best algorithm that was discovered to have a high accuracy in customer churn prediction is XGBoost. The research offers certain information to companies on how to come up with effective customer retention policies, and it has limitations, including the fact that the dataset is restricted to a customer bank and the possible factors

are not addressed in the research [16]. AI and ML technologies have enhanced the performance of a company, especially in the telecommunications sector. This research has come up with an AI-driven customer churn prediction (CCP) model of Telecommunication Business Markets (AICCP-TBM) to manage churners and non-churners within the industry. The model employs an FS method, which is based on Chaotic Salp Swarm Optimization and Fuzzy Rule-based Classifier to differentiate between churners and non-churners. The performance of the model was tested on a benchmark CCP dataset and proved to perform better in terms of predictions [17]. This study suggests a CC forecasting model based on unstructured data of conversations on the phone. With a big dataset of a call center, the model is able to predict risks of client churn and produce meaningful insights with interpretable ML. Managers can use this information to come up with retention strategies to suit the various customer segments, which will lead to the enhancement of customer retention in the financial services industry [18]. The business and financial term CC is used to refer to the gradual reduction of customer retention with time. ML algorithms such as LR and Naive Bayes (NB) can be used to determine the presence of potential customers who might be thinking of moving to a different bank. Such models are able to predict future customer attrition based on such information as age, location, gender, and credit card details. The study indicates that NB is effective compared to LR [19]. Churning of customers is also a very serious problem in the telecommunications industry, which influences the capacity of the businesses to retain as well as acquire new customers. The limitations of traditional solutions, such as rule-based and ML, include scalability and robustness. This research developed smart decision forest (DF) models to make CPs with emphasis on logistic model tree, RF, and Functional Trees. Experiments demonstrated that these models were able to distinguish between churn and non-churn consumers very well and outperformed the baseline and current ML-based approaches to CCP. Thus, CCP and ML tasks should be performed with the help of DF-based models [20]. The paper deals with CC in the telecommunications industry, where early prediction of churn is significant. It proposes the use of explainer models such as gradient boosting machine (GBM), which proved to be better with an accuracy of 81%. The paper comes up with a conclusion that GBM is superior in CP and visualization methods are effective in explaining important churn factors in the telecommunications industry [21]. The telecommunication sector is affected by customer churn, which can be enhanced through techniques of data transformation and an ML model. Research was conducted to study different models and a way of optimizing prediction models. The authors chosen univariate technique and grid search to pick the most appropriate hyperparameters. Test on publicly available TCI datasets indicated that the proposed technique enhanced the performance of prediction by up to 26.2% and 17% in AUC and F-measure, respectively [22]. In the given study, data science and AI methods are applied to forecast CC in the telecommunications sector. It uses past data, emphasizing the usage and preferences of services. A stacking ensemble is applied, and CatBoost performs better than a single algorithm. This assists in formulating good retention policies and increases customer satisfaction and loyalty, which is part of the comprehension of ensemble models [23]. XAI-Churn TriBoost is a data-driven model that was made to help the telecommunications industry retain customers. The model is based on the dataset of more than 2 million records, and it is a combination of extreme gradient boosting, categorical boosting, and light GBM. It has good predictive performance, with an accuracy of 96.44%, precision of 92.82%, recall of 87.82%, and

F1 score of 90.25. The model applies the explainable AI (XAI) methodology to explain personal predictions and determine the most important aspects that influence churn [24]. Churn among customers is a serious matter in the telecommunications sector, as advance warning of angry subscribers allows companies to determine churn causes and adopt new policies. The problem is critical to control using efficient CCP solutions. A number of approaches, such as rule-based and ML procedures, have been suggested to manage CCP. Nevertheless, CCP solutions that are based on rules are not flexible and robust, and the uneven distribution of churn datasets may affect the effectiveness of such solutions. The study examines the performance of various tree-based classifiers in CCP and the impact of issues related to data quality such as the class imbalance problem (CIP). In the findings, CIP does not improve the performance of tree-based classifiers as compared to other classifiers [25].

2.1. Summary of the literature survey

The available literature on CP shows the growing use of ML models in resolving the issues of consumer attrition. Although the effectiveness of separate models such as RF and SVM has been proven by numerous studies, there is still no study that compares these models in terms of their performance on different datasets and industries. The research is an extension of the current studies since it directly compares several models (LR, KNN, SVM, DT, and RF) on the same evaluation metrics. Also, our study is not only about predictive accuracy but also about interpretability and computational efficiency of these models, which is an important gap in the literature. This method will give a practical understanding that can be used by a business that is interested in

implementing a CP model, and it will give them an idea of which model can be applied to their operations.

3. Methodology

This section describes the methodology that will be used in predicting CC based on ML techniques. It gives a clear picture of the data collection process, preprocessing, model building, and evaluation procedures in this paper. The overall methodology is represented in Figure 1.

3.1. Overview of predictive analytics workflow

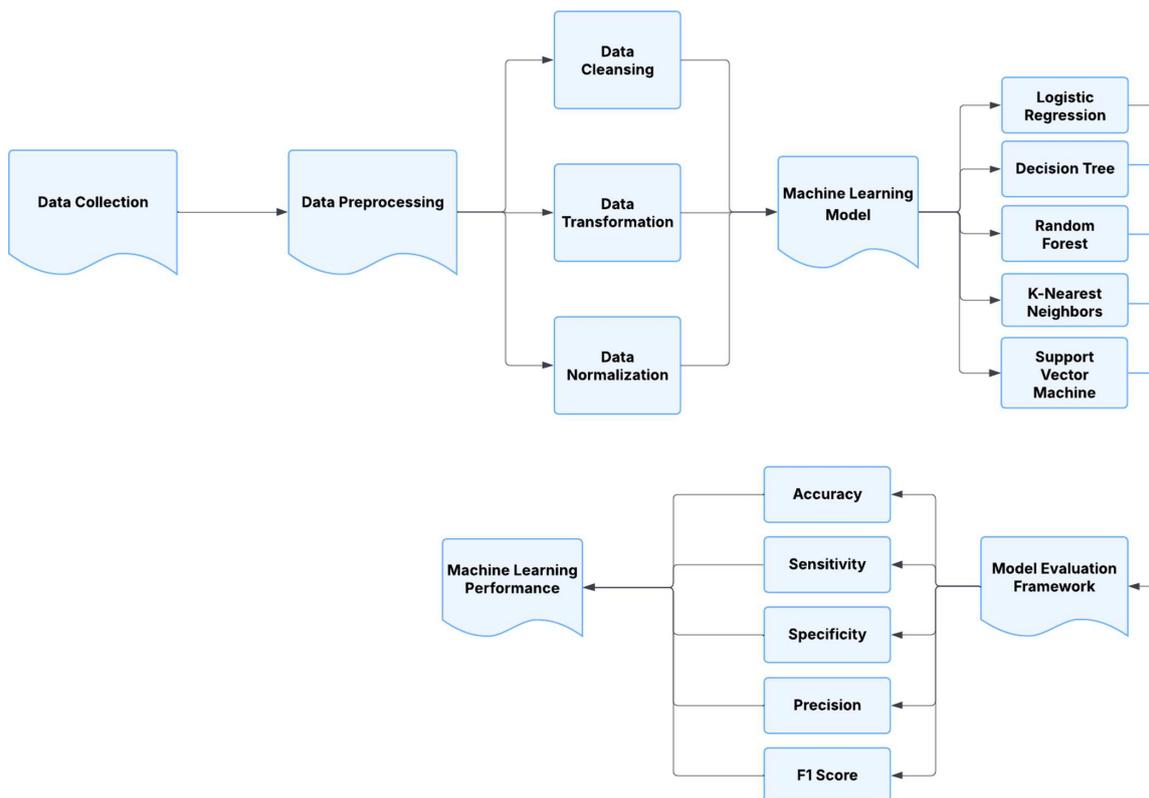
The predictive analytics process starts with the definition of the underlying problem of predicting churn of a customer. This task is presented as a binary classification task. It goes through a few main steps: data collection, data exploration, preprocessing, model development, model evaluation, and deployment. The purpose is to develop a predictive model that will have the ability of predicting customers who are at a high risk of churning so that the business can proactively take retention measures.

The dataset was selected due to its accessibility, structured labeling, and frequent adoption in churn research, enabling reproducible comparative analysis. However, findings should be validated on additional datasets to extend external validity.

3.2. Data collection, exploration, and preprocessing

The data of this research was sourced from Kaggle and contains 7043 customer records and 21 attributes, which include demographic information, account information, service usage

Figure 1 Overall methodology



patterns, and a target variable. The target variable is a binary classification problem because it determines whether a customer has stopped the service in the past month or not.

3.2.1. Dataset overview

The dataset records an overall picture of the customer behavior and service interaction and is categorized into the following categories:

Customer churn data: The Churn column (target) tells whether the customer has left the service (Yes/No).

Services subscribed by the customer: The services are PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, and StreamingMovies.

Account details: It comprises tenure (the duration the customer has been in the organization), Contract, PaymentMethod, PaperlessBilling, MonthlyCharges, and TotalCharges.

Demographics: These include gender, SeniorCitizen, Partner, and Dependents.

The data is moderately unbalanced, where 5174 non-churned (No) and 1869 churned (Yes) customers exist. This relates to an even distribution of classes as 73.5% and 26.5%, respectively. There are 21 features, 3 of which are numerical (tenure, MonthlyCharges, and TotalCharges) and 18 of which are categorical (both binary (e.g., Partner, Dependents) and multi-class (e.g., InternetService, PaymentMethod)).

3.2.2. Data cleansing

Data cleansing is the initial procedure of the preprocessing pipeline. This includes the detection and processing of inconsistencies, which may include missing values, irrelevant, or duplicate data entries. The column customerID is not informative to the prediction task as it is removed from the dataset. There are also missing values in the column of TotalCharges. Such missing values are substituted with the median of the column in order to avoid bias and maintain the integrity of data distribution. By doing this, the imputation will not have much effect on the learning of the model. In case duplicate records exist, row-level checks are used to identify and eliminate them to avoid redundancy. These data cleaning processes play a crucial role in enhancing the robustness and accuracy of models.

3.2.3. Data transformation

ML algorithms have to be fed with numerical data; therefore, the categorical features are transformed into numeric ones. Categorical variables on a binary basis like gender, Partner, Dependents and PhoneService are coded through label encoding, where a value is assigned to 0 or 1. As an example, the gender is coded into 0 and 1, whereby 0 represents female and 1 represents male. In case of more than two categories of categorical variables like InternetService, PaymentMethod, and Contract, one-hot encoding is used. The method produces a separate binary column per category so that the model does not presuppose any ordinality between the categories. This transformation does not introduce bias and increases the feature space, retaining all the categorical information. This transformation makes the input data in a structured form that can be used in ML algorithms and also removes any reliance on the fact that raw attributes are categorical.

3.2.4. Data normalization

In order to have uniformity in scale among the numerical features, the features are normalized. The features are different

in scale and units, and this may have a disproportionate effect on model training, particularly in algorithms that are sensitive to feature scales (e.g., KNN, SVM). Normalization brings the data to a standard scale, usually the range [0, 1], by the min-max scaling method. This not only speeds up the convergence in training but also increases the stability and correctness of the learning.

3.2.5. Dataset splitting

Following the preprocessing, a training/validation and testing set were separated with a ratio of 80/20. The data was split into 80% and 20%, whereby 80% was utilized in model development and hyperparameter optimization. The rest 20% was withheld as a test to determine the performance of the final model on unseen data.

3.2.6. Cross-validation strategy

To ensure model stability, stratified 10-fold cross-validation was performed on the training set to minimize performance variance due to random sampling. This technique repeatedly partitions the data, evaluates the model on multiple folds, and averages the scores, providing a more reliable estimate of generalization performance compared to a single train-test split. However, due to dataset availability constraints, validation was limited to this benchmark dataset.

3.3. Machine learning model development

ML has become a fundamental aspect of contemporary customer churn prediction because it is able to identify patterns and associations in large and complex data. Compared to the traditional statistical or rule-based systems, ML models do not require any prior knowledge, but they learn directly out of the historical data and update their predictions automatically as new patterns emerge. The model development process is the process of choosing the right algorithms, training them on the historical data about customers, optimizing the parameters of the model, and checking its predictive power [26].

Let us consider a dataset:

$$D = \{(x^{(i)}, y^{(i)})\}_{i=1}^N \quad (1)$$

Where:

$x^{(i)} \in R^d$ represents the feature vector for the i^{th} customer.

$y^{(i)} \in \{0, 1\}$ indicates the target label : 1 for churn, 0 for nonchurn

The objective is to learn a function f that maps $x^{(i)}$ and $y^{(i)}$, such that

$$f(x^{(i)}) \approx y^{(i)} \quad \forall i \quad (2)$$

Below is a general overview of the types of models and their advantages in the context of CP.

3.3.1. Logistic regression

LR is a common model applied to predict CC because it is easy to interpret and understand. It is used to model the likelihood that a customer will churn (label = 1) in terms of features like tenure, contract type, and payment method. Model coefficients can help businesses comprehend the effect of each variable, and this can be useful in targeting retention strategies. It presupposes

the linear connection between the input characteristics and the log-odds of the result. This model is especially helpful in cases where interpretation and feature importance are vital. It performs on large data and categorical data once encoded. LR is a good baseline to use in churn classification [27].

$$P(y = 1|x) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (3)$$

$$\hat{y} = \begin{cases} 1 & \text{if } P(y = 1|x) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$L_{Logistic} = -\frac{1}{N} \sum_{i=1}^N [y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})] \quad (5)$$

Where:

w is a weight vector,

b is a bias term,

$\sigma(\cdot)$ squashes values to (0, 1).

A customer is predicted to churn if $P > 0.5$.

3.3.2. Decision tree

The DTs are simple and efficient in customer churn prediction since they divide the customers into categories regarding their monthly bills or the type of internet service. Every internal node in the tree presents a yes/no query of a feature of the customer, and the final prediction is made at a leaf node. This method allows one to simply explain the churn reasons to non-technical stakeholders. To give an example, it may demonstrate that customers with short tenure and month-to-month contracts are more prone to churn out. Trees are naturally adequate to deal with numeric and categorical data. They, however, are prone to overfitting on the noisy data. They are commonly applied as a basis of ensemble techniques.

Gini Index:

$$\text{Gini}(S) = 1 - \sum_{k=1}^k p_k^2 \quad (6)$$

Entropy

$$\text{Entropy}(S) = - \sum_{k=1}^k p_k \log_2(p_k) \quad (7)$$

where p_k is the proportion of samples belonging to class k . Information Gain measures the reduction in impurity after splitting on attribute A :

$$\text{IG}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (8)$$

3.3.3. Random forest

RF improves customer churn prediction by averaging the outputs of a number of DTs. It randomly selects features at each split, and it creates a forest of trees through bootstrap sampling to achieve model diversity. This minimizes the chances of overfitting compared to a single tree and enhances generalization to unseen customers. It is able to detect complicated relationships among features, like the combination of tenure and billing type, and their impact on the churn risk. The summation of the voting of all trees stabilizes it and makes it accurate. It also provides feature importance ranking that assists businesses to find out important

churn drivers. RF is perfect when interpretability is not of primary concern.

$$f_{RF} = \text{majority_vote}(f_1(x), f_2(x), \dots, f_T(x)) \quad (9)$$

Each $f_t(x)$ is DT trained on a bootstrapped subset of data:

$$f_t(x) : \text{trained on } D_t \subset D \quad (10)$$

3.3.4. Support vector machine (SVM)

In CP, SVM is used to identify an optimal hyperplane that would best distinguish between the churners and the non-churners, using feature vectors. It is particularly efficient when the data is of high dimension, that is, when numerous behavioral features are coded. SVM is able to maximize the margin between the two classes, thus providing a strong decision boundary. It is suitable for complex data patterns and can process non-separable data nonlinearly with the help of kernels. SVM assists in the identification of minute differences between the customers who remain and those who churn in CP. It is not easily interpretable but has a high accuracy in balanced or moderately imbalanced datasets. It is frequently applied in situations where the boundaries of decisions are complicated, and this necessitates high generalization.

Linear SVM decision function:

$$f(x) = \text{sign}(w^T x + b) \quad (11)$$

Optimization problem:

$$\min_{w, b} \frac{1}{2} \|w\|^2 \text{ subject to } y^{(i)} (w^T x^{(i)} + b) \geq 1 \quad (12)$$

Kernel trick (for nonlinear cases);

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (13)$$

3.3.5. K-Nearest Neighbors (KNN)

KNN uses similarity between a new customer and the previous customer to predict the churn of the customer. It computes distances between feature vectors (like total charges, type of contract, or usage of the services) and finds the k nearest customers. The most prevalent of these neighbors (churn or not) becomes the prediction. KNN is simple and efficient, whereby similar customers show similar results. It does not need any training but may be slow when making a prediction on large datasets. The accuracy of the model is strongly reliant on the value of k and the good scaling of the data. It works optimally on low-dimensional or already processed feature spaces with little noise [28].

Euclidean distance:

$$\text{dist}(x, x') = \sqrt{\sum_{j=1}^d (x_j - x'_j)^2} \quad (14)$$

Prediction function:

$$f(x) = \text{majority_vote}(\{y^{(j)} : x^{(j)} \in N_k(x)\}) \quad (15)$$

3.3.6. Benefits of using machine learning to predict churn

1) High accuracy: ML models are able to recognize nonlinear patterns and hidden signals observed by the traditional methods, which are not recognized, and hence improve accuracy in prediction.

- 2) Scalability: The models of ML are able to process huge amounts of data about customers and allow enterprises to forecast churn among millions of customers.
- 3) Automation: Trained models can then be used to make real-time predictions of churn and can be used to automate customer retention efforts such as trigger-based campaigns.
- 4) Proactive retention: ML can be used to ensure that the companies know who is going to churn out so that it can be handled in time through offers, support, or engagement.
- 5) Cost-efficiency: Since the retention efforts are only directed to the customers who are likely to churn, it will save the businesses cost in marketing and increase Return on Investment (ROI).
- 6) Adaptability: The models can be retrained on a regular basis since the customer behavior evolves over a period of time, and hence, the predictions remain relevant [29].

3.3.7. Hyperparameter sensitivity and selection rationale

To ensure fair comparison and avoid suboptimal configurations, hyperparameters for KNN and RF were tuned using grid search combined with stratified 10-fold cross-validation. For KNN, *k* values were tested in the range of 1–15. The results indicated that extremely small *k* values (*k* = 1–3) produced high variance and were sensitive to noisy samples, while larger values (*k* > 11) blurred local neighborhood structure and reduced sensitivity to subtle churn patterns. Optimal performance was observed around *k* = 7, which provided a balanced trade-off between stability and local pattern fidelity.

For RF, the number of DTs (estimators) was evaluated from 50 to 300 in increments of 50. Configurations with fewer than 100 trees tended to underfit and exhibited noticeable fluctuation across folds, whereas increasing the number of trees beyond 200 resulted in only marginal improvements (< 0.5% accuracy gain) while increasing training time by approximately 18%. The final choice of 200 trees was therefore selected to maximize generalization without incurring unnecessary computational cost. These observations are consistent with the bias–variance dynamics of ensemble models, and the selected hyperparameters reflect a performance–efficiency balance suitable for real-world deployment environments. Key hyperparameters applied in the model are represented in Table 1.

3.4. Model evaluation framework

After training models, they are evaluated with the help of an extensive assessment framework. The main goal is to establish the extent to which the model performs well on the new data, which has not been seen before, and how well it can distinguish between churners and non-churners.

Model predictions are tested on a holdout set or cross-validation, and a number of metrics are computed.

3.4.1. Accuracy

Accuracy determines the general accuracy of the predictions. Nevertheless, accuracy is not the only metric that can be misleading since churn datasets are typically skewed (there are more non-churners than churners).

3.4.2. Sensitivity (recall)

Sensitivity or recall is the ratio of the number of actual churners correctly classified. It is essential in the case of churn since the inability to predict potential churners may lead to a loss of revenue.

3.4.3. Specificity

Specificity is the identification of non-churners by the model. The high specificity will make sure that non-at-risk customers who are loyal will not be marked incorrectly as being at-risk and subject to intervention.

3.4.4. Precision

Precision is the ratio of the number of customers that are predicted to churn and the ones that actually churn. It will assist in reducing false positives, so retention efforts will be done to the right audience.

3.4.5. F1 score

The F1 score is a trade-off between precision and recall. When the F1 score is high, this means that the model has a good balance between detecting churners and false alarms [30].

Additional measures such as AUC-ROC and confusion matrices are also obtained in order to get a more detailed insight into the model performance. This approach is an end-to-end methodology on how to build CP models based on ML. Starting with the structured data collection and preprocessing, proceeding with the intelligent modeling, and through an intensive evaluation, the process is both technically sound and business relevant. ML is not just about better prediction but also about enabling organizations to take proactive measures to minimize churn, better allocate resources, and boost customer satisfaction.

4. Results

The results part gives a detailed analysis of the five ML models, namely, LR, KNN, SVM, DT, and RF, using several measures of performance. These models were run on the CC

Table 1
Key hyperparameters applied in the model

Model	Key hyperparameters	Values used
Logistic regression	Regularization, Solver, Penalty	<i>C</i> = 1.0, liblinear, L2
KNN	<i>k</i> -Neighbors, Distance Metric	<i>k</i> = 7, Euclidean
SVM	Kernel, <i>C</i> , Gamma	RBF, <i>C</i> = 1.0, gamma = scale
Decision Tree	Criterion, max_depth, min_samples_split	Gini, max_depth = 10, min_samples_split = 2
Random Forest	n_estimators, max_features, max_depth, bootstrap	200, sqrt, None, True

dataset, and their performances were determined with accuracy, sensitivity, specificity, precision, and F1 score.

The performance of five ML models in CCP is indicated in Figure 2. RF model had the highest accuracy of 80.74 compared to other models, and this shows that it had a better capacity of classifying the churn and non-churn customers. SVM comes next with a high score of 79.96, indicating that it performs well, probably because of its ability to work with data in high dimensions. LR also achieved 79.53, which is an indication that it is reliable in carrying out binary classification. DT scored 78.89, which is a little less than the rest, and it could be because of overfitting. KNN showed the least accuracy of 77.9, which could be affected by noise and selection of neighbors. On the whole, ensemble models such as RF have a tendency to generalize better than single models. The findings indicate that the complexity of models and the treatment of nonlinearity have implications for the predictive performance. The features were selected properly, and the data was preprocessed, which could have led to high values of accuracy. Thus, RF is the most suitable model among those compared for CP.

In Figure 3, sensitivity (recall) scores of different models in predicting CC are presented. SVM has the greatest level of sensitivity of 57.89, which means that it is highly able to recognize the real churn instances. LR comes right behind with 57.38, and thus, it is a reliable model to detect churn. RF attained 56.14, which is a decent variance/bias trade-off in the identification of churners. KNN possesses a moderate sensitivity of 54.38, which can be influenced by the selection of the neighbors and data

distribution. The weakest performance is achieved by DT and is equal to 41.88, which is probably caused by overfitting and inability to generalize. Sensitivity plays an important role in CP since it indicates the success of a model in identifying the customers who are likely to churn. The sensitivity values are relatively low in comparison to accuracy, and this implies that there is a possibility of class imbalance. The higher the sensitivity of the models, the more useful in the retention strategy since they aid in the identification of at-risk customers early enough. Thus, SVM is the most sensitive and then comes the LR and RF.

Figure 4 shows the specificity of different models in CCP, which is the proportion of non-churn customers that are correctly identified by the model. The DT model has the greatest specificity of 92.68, which implies that the model is best in identifying the customers who are not likely to churn. RF comes next with 89.94, as it is an ensemble method, which is performing well. SVM has 88.12, which shows that it is balanced in its classification. LR achieves an 87.32 score, which indicates its regular performance in identifying stable customers. The lowest specificity is 86.5 of KNN, which might be associated with its sensitivity to the local patterns and noise. It is indicated by the overall high values of specificity that each of the models is more efficient in detecting non-churners than churners. Such an imbalance can be attributed to biased data, in which the majority of cases are non-churn. Although high specificity can decrease false positives, it can result in an increased number of missed cases of churn unless balanced with sensitivity. Therefore, in real scenarios, a model that has a decent balance between sensitivity and specificity, such as RF or SVM, will be preferable.

The precision scores of the various models in the prediction of CC are shown in Figure 5, and this is the measure of the proportion of the correctly identified churners against the no of all the churn cases that were predicted. The precision of DT is 68.09,

Figure 2
CCP using different models using the accuracy parameter

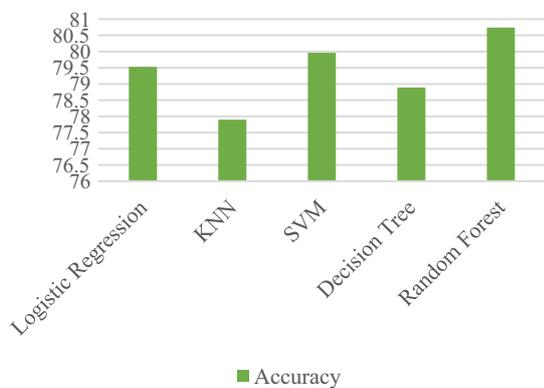


Figure 3
Customer churn prediction using different models using the sensitivity parameter

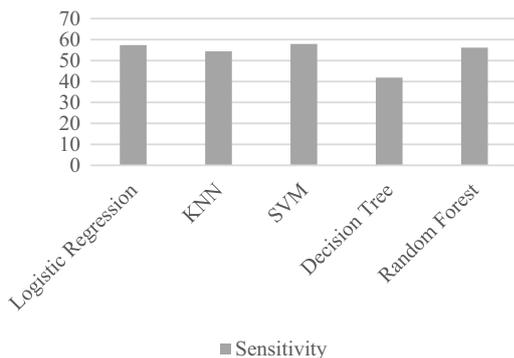


Figure 4
Customer churn prediction using different models using the specificity parameter

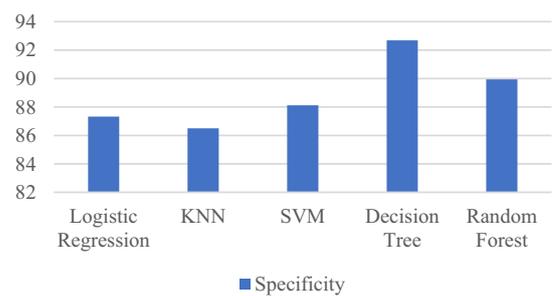


Figure 5
Customer churn prediction using different models using the precision parameter

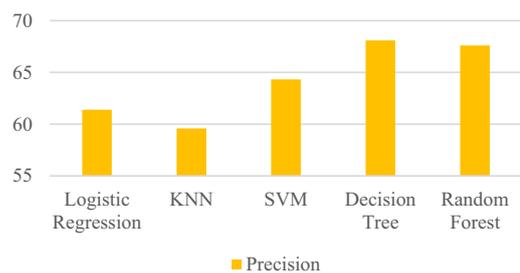
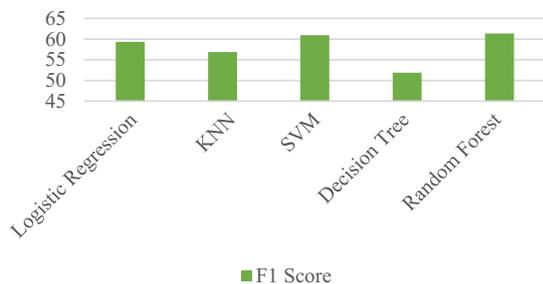


Figure 6
Customer churn prediction using different models using the F1 score parameter



and this shows that DT has a high power of making positive predictions. RF comes right behind with 67.61, indicating that it is effective in minimizing false positives using ensemble learning. SVM has a good value of 64.33, which implies that both the accuracy of the correct prediction of true churners and the reduction of errors are good. LR possesses a moderate precision of 61.4, with consistent but a bit lower accuracy in the CP. KNN has the lowest precision of 59.59, and this could be attributed to its local decision boundaries and sensitivity to noise in data. In CP, high precision is essential to ensure that no one targets customers who have little chance of leaving. It shows that tree-based models, in particular DT and RF, perform better when it comes to making confident CPs. Such models may be applicable in targeted retentions that do not require a lot of wastage of resources.

Figure 6 shows the F1 score of different models used to predict CC that balances the precision and the recall to give a single measure of the efficiency. RF model has the best F1 score of 61.34, which implies that it has a good trade-off between the number of churners that are correctly predicted and the number of false positives. SVM ranks closely behind with 60.94, which indicates its balanced performance in churn classification. The results of LR are also consistent with 59.32, and it works well on various metrics. KNN achieves a lower F1 score of 56.87, implying that it is performing moderately in managing class imbalance. The F1 score of DT is the lowest, 51.86, which could be explained by the sensitivity that is not high enough even though the precision is high. In CP, it is of particular importance to have a higher F1 score because false positives and false negatives both affect the business. These findings indicate that ensemble models such as RF are more reliable in churn tasks. In general, RF and SVM are the best models, which balance between accuracy of predictions and reliability.

4.1. Performance summary of each model

- 1) The overall performance of the LR was good (accuracy 79.53%), with moderate sensitivity (57.38%), high specificity (87.32%), decent precision (61.4%), and a good F1 score (59.32). It is interpretable and reliable and can also fail to handle complex patterns.
- 2) KNN performed the lowest in general, with the accuracy being 77.9 and the F1 score being 56.87. It obtained a good score in the sensitivity (54.38%) and specificity (86.5%), whereas the precision was a little lower (59.59%). It is sensitive to noise in the data and cannot handle high-dimensional data.
- 3) SVM was very consistent in all the metrics with an accuracy of 79.96%, the best sensitivity of 57.89%, good specificity of 88.12%, good precision of 64.33%, and solid F1 score

of 60.94%. It is good in nonlinear relationships; hence, it is appropriate in complex patterns.

- 4) DT performed rather mixed. Although it had the best specificity (92.68%) and precision (68.09%), it was not good in sensitivity (41.88%), and it had the lowest F1 score (51.86), signifying that it could not generalize and that it was prone to overfitting.
- 5) The RF model was the best one with an accuracy of 80.74% and an F1 score of 61.34%. It was also exhibiting balanced sensitivity (56.14%), high specificity (89.94%), and strong precision (67.61%). Its ensemble form improves its robustness and minimizes overfitting.

4.2. Highlighting best-performing models

- 1) RF is the most effective model in general, taking the first place in accuracy, F1 score, and precision, and having high specificity. It has a good balance between recall and precision, thus making it very appropriate in CP.
- 2) SVM is its close competitor, as it performs well on all measures, and sensitivity is the most important in churners identification as soon as possible.
- 3) Although DT was very precise and specific, its sensitivity is very low, as well as the F1 score, which makes it less reliable in the real world.

4.3. Generalization potential across customer-facing industries

The findings show that RF and SVM possess good generalization capabilities in different areas related to customers.

- 1) RF is able to generalize well because it is robust, able to deal with noisy data, and has less overfitting.
- 2) SVM can work well in the situation of high-dimensional data and nonlinear customer behaviors, which allows SVM to be used in fraud detection, credit scoring, and churn modeling.
- 3) LR is still a desirable baseline model because of its interpretability in the regulated industry such as finance and healthcare.

4.3.1. Dataset generalizability considerations

While the Kaggle telecommunications dataset is a widely adopted benchmark in churn research, its extensive reuse may lead to research saturation and potential model overfitting toward its specific feature distribution. The dataset does not capture emerging service offerings, customer–platform interactions, or market-specific churn drivers found in other industries. Therefore, the generalizability of the findings beyond similar telecom contexts should be interpreted cautiously.

4.3.2. Synergy between preprocessing steps and model performance

To evaluate the interaction between preprocessing operations and model performance, ablation experiments were conducted using configurations with and without normalization and categorical feature encoding. The results showed that normalization had a notable synergistic effect on algorithms that rely on distance or gradient calculations. KNN exhibited a performance gain of approximately + 3.5% in accuracy and + 2.8% in F1 score after applying min–max scaling, since unscaled numerical attributes such as MonthlyCharges otherwise dominate distance

computations. Similarly, SVM benefited from normalization due to improved geometric margin optimization and kernel stability. In contrast, tree-based models such as DT and RF demonstrated minimal sensitivity (< 0.7% variation) because their split decisions are based on relative thresholds rather than numerical magnitude. Furthermore, one-hot encoding significantly improved the performance of LR and SVM by preventing implicit ordinal relationships among multi-class categorical features (e.g., Contract type). Without encoding, these models prematurely assigned linear importance to categorical labels, which led to misinterpretation of churn-related patterns. These findings indicate that preprocessing and model selection must be co-designed to maximize predictive quality and avoid feature dominance or ordinal distortion.

5. Discussion

5.1. Strengths of machine learning

The use of ML as a tool to predict CC is quite powerful because ML can learn by itself, can discover non-obvious patterns in the data, and can keep getting better with each new data point. These findings on different models, such as LR, KNN, SVM, DT, and RF, reveal the specific advantages of ML in this field.

Among the most important benefits of ML, it can automate the process of pattern recognition. ML algorithms are able to work with large amounts of information and discover minor patterns of churn that could otherwise be overlooked by human analysts. The RF and SVM have high accuracy and F1 scores, which implies that the two models are good in representing non-linear relationships between consumer behavior and the likelihood of churn, and hence can be used to make solid predictions.

Also, ML offers flexibility when selecting the model, and algorithms are available for different types of data and business requirements. To give an example, LR has the advantage of interpretability, which is essential to the decision-maker in regulated industries, and RF, as well as other ensemble methods, is the best option to work with nonlinear and complicated data patterns. The use of an ensemble such as RF also improves generalization of the model and minimizes overfitting as indicated in the balanced performance across all metrics.

5.2. Practical use cases for business decision-makers

To business decision-makers, ML-driven CP means concrete advantages and an action plan that can entail a substantial decrease in customer loss.

The churn models can assist the companies to identify customer dissatisfaction in the early stages, and therefore, they may provide personalized retention approaches to the customers, for example, discounts, improved plans, or individual customer care. In the field of banking and finance, ML models can identify customers at risk of leaving their bank or credit card company, which would allow one to provide them with a timely incentive or consultations to prevent this. In the same way, in e-commerce, the companies can use CP models to reduce cart abandonment and enhance customer retention by detecting the behavioral patterns that predict churn and starting personalized marketing campaigns.

The accuracy of models such as DT and RF implies the minimization of false positives. Consequently, there will be no wastage of resources by businesses trying to avoid the targeting of customers unlikely to churn. ML helps companies prioritize

retention efforts according to the risk of churn and results in more effective resource utilization, better retention rates, and CLV.

5.3. Addressing challenges in churn prediction

5.3.1. Imbalanced datasets

The imbalance of the classes, with non-churners being more numerous than churners, represents one of the greatest difficulties in the CP. This imbalance may lead to a situation where models can show good results in terms of accuracy but will not identify churners properly. This is clear by the fact that the sensitivity scores have been lower in all the models compared to their specificity. As an illustration, although DT has a high specificity (92.68%), it has rather low sensitivity (41.88%), which means that it is effective at detecting non-churners but not at detecting churners.

In order to counter this, methods like SMOTE and majority undersampling are necessary. Also, other measures of model evaluation, in addition to accuracy, sensitivity, F1 score, and precision, are important in the case of imbalanced data. These measures make sure that the model will not only be able to predict the majority group but also classify the minority group (churners) with enough accuracy.

5.3.2. Model complexity vs. interpretability

The other issue that is common with CP is the balance between achieving high model complexity and interpretability. More advanced models, like RF and SVM, tend to provide superior results (e.g., the accuracy of RF of 80.74% and F1 score of 61.34%) but are not transparent, which can be an issue in industries where it is essential. On the contrary, LR is simpler to comprehend but can be weak in case the connections between input features and churn are nonlinear.

It is especially relevant in the regulated industry, such as healthcare or finance, where a company has to explain its decisions. In these applications, models that can be interpreted or combined with explainability methods, such as LR, SHAP, or LIME, may be used to obtain the required transparency without compromising performance.

5.3.3. Cross-dataset generalization

While the models achieved consistent performance during cross-validation, the evaluation was limited to a single telecom dataset. Churn drivers can vary widely across sectors such as media subscriptions, cloud platforms, and financial services. Consequently, model robustness cannot be fully confirmed without testing on multiple heterogeneous datasets that capture distinct behavioral characteristics.

Since the dataset used in this study originates from a telecommunications environment, the resulting behavioral patterns, feature importance, and churn triggers inherently reflect subscription-driven service models. Real-world churn, however, differs significantly across domains. For example, e-commerce churn may be influenced by seasonal purchasing trends and basket abandonment, whereas banking churn may depend on credit utilization history, account risk scores, and incentive structures. Therefore, evaluating the proposed models on multiple cross-industry datasets is essential to mitigate domain-specific bias and improve external validity. Future work should incorporate diverse datasets such as e-commerce clickstream logs, financial service transaction histories, and media subscription cycles to compare changes in accuracy, recall, and F1 score under distinct churn drivers. Additionally, conducting experiments under

segmentation scenarios (e.g., tenure cohorts, CLV tiers, and contract-type clusters) would enable assessment of model adaptability to behavioral sub-populations. This scenario-driven validation approach would help determine whether ensemble-based models like RF and kernel-based algorithms such as SVM maintain predictive superiority beyond telecommunications. Ultimately, this broader evaluation design would enhance the generalization potential of churn analytics across real-world business environments and prevent conclusions from being constrained to a single industry dataset.

5.4. Role of evaluation metrics in model selection

The choice of the appropriate CP model is a crucial process that should be made based on several metrics. Accuracy is not enough, especially in the case of imbalanced data. Majority class predictive models can be very accurate but fail to predict churners, which is the most important objective of CP.

Such measures as sensitivity (recall) are important in order to measure the churners, the customers that a business should target to retain. Accuracy aids in the assurance that the customers who are marked as those to be retained are really in danger of exiting, which optimizes the resources of marketing and customer support. A more complete picture of the performance of models can be presented using the F1 score that balances precision and recall. RF (F1 score of 61.34) and SVM (F1 score of 60.94) in this study have a good balance of these metrics and hence qualify as the best options when precision and recall are important. The business decision-makers ought to prioritize the metrics with the strategic objectives of the organization. Sensitivity should be given the first priority in case the main objective is to reduce churn and include all the potential churners. Precision ought to be focused on in case it is required not to spend resources on consumers who are not likely to churn. To have a balanced approach, the F1 score is more of a complete picture, and therefore, it should be taken as the main evaluation metric.

5.5. Insights gained from comparative analysis

The comparative evaluation of the five ML models gives important insights about their advantages, disadvantages, and applicability in the field of CP:

- 1) RF can be spotted as the most all-around optimal, scoring the most significant accuracy, F1 score, and close-to-top precision and specificity. It is strong and flexible and can effectively model nonlinear and complex interactions and thus is suitable in the high-volume industries where customer data is huge.
- 2) SVM is distinguished by the balance of the performance, especially sensitivity (57.89%) and specificity (88.12%). It has a high potential in identifying both churners and non-churners; thus, it is very good to use in an industry where both classes of individuals have to be identified accurately.
- 3) LR provides a good trade-off between interpretability and performance. It can also be applied quite well in situations where transparency of the model is critical, although its relatively poor performance in more complicated data settings means it would not be a good choice in highly nonlinear CP problems.
- 4) Although DT is very specific and accurate, its recall and F1 score are low, which means that it is overfitting and does not generalize well. It is possible that this model can work out in the simpler cases, but it must be applied with caution in real life.

- 5) KNN is the poorest of them all, particularly in F1 score (56.87%). It has lower performance in large, noisy datasets in CP tasks due to its sensitivity to data characteristics such as scale and noise.

All in all, the analysis indicates the necessity of applying such ensemble techniques as RF and such kernel-based approaches as SVM to make successful predictions of churn. These models not only generalize well on various types of data but also provide high predictive power, which makes them useful tools in data-driven decision-making in customer-centric industries.

5.5.1. Dataset source and external validity

While the Kaggle dataset provides a well-structured and widely used benchmark for comparative evaluation, it may not fully reflect the heterogeneity, noise, or behavioral dynamics present in proprietary enterprise data. Public datasets often lack historical interaction logs, marketing campaign exposure, and sentiment data, which can influence churn behavior in real-world deployments. Accordingly, model performance reported in this study should be interpreted as indicative rather than definitive for all business contexts.

5.5.2. Real-world deployment considerations

In industrial environments, CP models are typically integrated into CRM platforms, where predictive outputs trigger automated retention campaigns. Deployment pipelines must include data ingestion layers, batch or real-time inference modules, monitoring dashboards, and feedback loops for model retraining. In practice, inference latency, scalability, and regulatory compliance are critical operational constraints that require careful engineering beyond model performance alone.

6. Cost Implications and Resource Considerations

6.1. Computational cost of training and deploying models

Among the most important factors of applying ML models to CP is the fact that both training and deployment are computationally expensive. Such complicated models as RF and SVM are computationally heavy and need a lot of memory, processing power, and time to train, particularly on large data. RF is an ensemble algorithm that requires extensive training time and space since a number of DTs are constructed. Moreover, SVM, especially when nonlinear kernels are used, is based on complex computations on the optimization process, particularly in the case of high-dimensional data. It implies that the larger the size of the data, the more time and resources it consumes to train the model exponentially.

LR is, on the other hand, a lighter model in terms of computation. It learns a lot quicker, needs less resources, and is highly applicable in real time or in situations where the hardware is not able to provide much. LR is very efficient, and hence, it is suitable to deploy it in the first stage, test it, or apply it in systems with limited computing capabilities and memory like mobile systems or embedded systems.

The latency of model predictions is a relevant consideration when it comes to production. RF can be more accurate but can take longer to make predictions, given that the model has already been trained (e.g., it is possible that it takes longer to make an inference than in simpler models such as LR or DT). LR is

simple, and as such, it is faster to make predictions; hence, it is better suited in an environment where low-latency predictions are important. Companies, therefore, need to make a trade-off between the performance of the model and the efficiency of the operations, especially on systems that need quick decision-making or those that need to be updated in real time.

6.2. Trade-offs and real-world deployment challenges

6.2.1. Simplicity vs. power

A typical trade-off exists when deciding on whether to use simple models such as LR or powerful models such as RF. LR is quick, simple to apply, and very interpretable. It is useful for simple issues in which the associations between features are linear and in which interpretability is important. It is however not well-suited for handling intricate nonlinear dependencies in the data and hence has limited capability to predict more complex CP use cases.

RF, on the other hand, is more accurate and generalizes better especially in the modeling of the complex, intricate patterns in datasets. This increases its strength as an instrument of CP, as it is proven in the study in terms of accuracy, precision, and F1 score. Nonetheless, it is at the cost of more computational resources, training time, and interpretability. RF may act as a black box, which is why it may be difficult to comprehend how the model reaches its predictions by the stakeholders.

6.2.2. Interpretability vs. performance

The other important trade-off is interpretability and performance. In regulated industries (finance, healthcare, and insurance), stakeholders of the business are usually interested in simple models that can be easily described and explained. LR offers an insight into the need for attributes in the form of coefficients, thus being transparent and interpretable. On the contrary, more sophisticated models, such as RF although guaranteeing better predictive accuracy, are more likely to act as black boxes; that is, their decision-making process is more difficult to understand.

In the case of industries in which explainability must be performed to meet regulatory or operational needs, a potential solution is to use strong models (RF or SVM, for instance) and combine them with XAI frameworks (SHAP or LIME, respectively). These are tools that give some visibility of how a model makes its predictions; thus, they add more complexity to the deployment process, which could impact the overall efficiency and transparency of the whole system.

6.2.3. Real-world deployment challenges

The implementation of ML models in the real-life business environment is associated with a number of issues that businesses have to deal with:

Data availability and quality: Among the main issues is the fact that companies should have quality and relevant data. The quality of historical customer data is extremely important to ML models, particularly in the context of churn forecasting. When the data is incomplete, biased, or noisy, it can severely restrict the effectiveness of the model, resulting in poor performance of the prediction.

System integration: The ML models must be integrated smoothly with the current CRM systems and other decision-support systems that are deployed by businesses. It is crucial to make sure that the models will be capable to work in real time,

and the predictions can be passed to the customer engagement or retention process.

Monitoring and maintenance: After deployment, there should be constant monitoring to check that the models are up to date as the behavior of customers changes with time. There is a possibility of concept drift, that is, variations in data patterns over time, which can lead to a reduction in the performance of the model with time. This challenge can be mitigated by regularly updating the models with new data and retraining them to make sure that the predictions remain relevant.

In addition to that, the costs of running the model should also be taken into consideration by businesses, including the resources required to perform inference in real time and the human resources necessary to update and monitor the models.

6.2.4. Data availability constraints

Access to comprehensive, proprietary customer datasets is often restricted due to privacy regulations and organizational confidentiality, which limits opportunities to validate models across diverse markets. Consequently, publicly available datasets serve as an accessible foundation for benchmarking but should ideally be complemented with domain-specific data for production-level deployment.

7. Conclusion and Future Work

In this study, the performance across several ML algorithms, namely, LR, KNN, SVM, DT, and RF, was tested on several metrics in terms of CCP. The findings indicated that RF was more effective than other models in terms of accuracy (80.74%), F1 score (61.34%), and precision and proved to be robust in terms of dealing with nonlinear relationships and obtaining a desirable combination of sensitivity and specificity, thus suitable for actual CP. The study has added value to the literature in that it compares the models in various dimensions such as computational efficiency, interpretability, and prediction performance, which makes the ensemble and kernel-based models such as RF and SVM important, particularly in the high turnover industries. Nevertheless, there are limitations to the analysis, which include the bias of the dataset, which can interfere with the applicability of the results to the rest of the population, and overfitting of such models as DT. These limitations indicate that future studies must be based on strong validation and a variety of data. The potential tasks that can be done later include the deployment of the methods of drift detection in order to track the variations in customer behavior and keep the model updated continuously via incremental learning, and the creation of a system capable of real-time churn forecasting integrated with CRM platforms to enable timely and proactive retention measures. Furthermore, the analysis was conducted on a single public dataset, which may restrict the generalizability of the findings across different industries, regions, and customer portfolios. Future research should incorporate multiple real-world datasets sourced from enterprise CRM systems to evaluate model robustness under varying demographic distributions, churn triggers, and market conditions. Cross-dataset validation and domain adaptation techniques may further enhance external validity and promote scalability in live customer retention environments. Future extensions should incorporate multiple heterogeneous real-world datasets sourced from financial services, media subscriptions, cloud platforms, and retail loyalty programs. Cross-domain validation, transfer learning, and domain adaptation strategies can further assess model robustness and reduce

dependence on saturated public benchmarks. Evaluating models across varying temporal cycles would also reveal their resilience to shifting behavioral patterns. Although RF became the most efficient model, improving adaptability, minimizing bias, and developing solutions that are scalable to be used in real time and on a large scale in the industry should become the center of future research.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at <https://www.kaggle.com/datasets/blastchar/teleco-customer-churn/data>.

Author Contribution Statement

Ramshankar Nagarajan: Methodology, Software, Validation, Resources, Visualization. **Siva Subramanian Raju:** Conceptualization, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Athiraja Atheeswaran:** Methodology, Software, Validation, Visualization.

References

- [1] Wu, Z., Jing, L., Wu, B., & Jin, L. (2025). A PCA-AdaBoost model for E-commerce customer churn prediction. *Annals of Operations Research*, 350(2), 537–554. <https://doi.org/10.1007/s10479-022-04526-5>
- [2] Saha, S., Saha, C., Haque, M. M., Alam, M. G. R., & Talukder, A. (2024). ChurnNet: Deep learning enhanced customer churn prediction in telecommunication industry. *IEEE Access*, 12, 4471–4484. <https://doi.org/10.1109/ACCESS.2024.3349950>
- [3] Imani, M., Joudaki, M., Beikmohammadi, A., & Arabnia, H. (2025). Customer churn prediction: A systematic review of recent advances, trends, and challenges in machine learning and deep learning. *Machine Learning and Knowledge Extraction*, 7(3), 105. <https://doi.org/10.3390/make7030105>
- [4] Soleiman-garmabaki, O., & Rezvani, M. H. (2023). Ensemble classification using balanced data to predict customer churn: A case study on the telecom industry. *Multimedia Tools and Applications*, 83(15), 44799–44831. <https://doi.org/10.1007/s11042-023-17267-9>
- [5] Harish, A. S., & Malathy, C. (2023). Evaluative study of cluster based customer churn prediction against conventional RFM based churn model. In *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies*, 1–6. <https://doi.org/10.1109/ICEEICT56924.2023.10156962>
- [6] C, J. K., Anumala, A., E.P, S., & E.S, B. (2024). Bank customer churn and extra benefits prediction using machine learning model. In *2024 Second International Conference on Advances in Information Technology*, 1–6. <https://doi.org/10.1109/ICAIT61638.2024.10690731>
- [7] Balaji, G., Gowtham, N. A., Tarun, S., Prajapati, G., & Manikandan, N. (2024). Customer churn prediction using machine learning algorithms. In *2024 International Conference on Emerging Research in Computational Science*, 1–6. <https://doi.org/10.1109/ICERCS63125.2024.10895079>
- [8] Rafi, M., Ahmad, M. F., Varshitha, K., Siri Varsha, T., Lahari, K., Haque, M. A., . . . , & Dutta, S. R. (2024). Customer churn prediction employing ensemble learning. In *2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Applications*, 207–211. <https://doi.org/10.1109/ICCCMLA63077.2024.10871687>
- [9] Srinivasan, R., Rajeswari, D., & Elangovan, G. (2023). Customer churn prediction using machine learning approaches. In *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering*, 1–6. <https://doi.org/10.1109/ICECONF57129.2023.10083813>
- [10] Vanitha, M., & Faritha Banu, J. (2025). Enhancing customer retention: A federated machine learning framework for banking churn prediction. In *2025 International Conference on Automation and Computation*, 143–148. <https://doi.org/10.1109/AUTOCOM64127.2025.10957571>
- [11] Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: A machine learning approach. *Computing*, 104(2), 271–294. <https://doi.org/10.1007/s00607-021-00908-y>
- [12] De, S., & Prabu, P. (2022). Predicting customer churn: A systematic literature review. *Journal of Discrete Mathematical Sciences and Cryptography*, 25(7), 1965–1985. <https://doi.org/10.1080/09720529.2022.2133238>
- [13] Fujo, S. W., Subramanian, S., & Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, 11(1), 185–198. <https://doi.org/10.18576/isl/110120>
- [14] Xiahou, X., & Harada, Y. (2022). B2C e-commerce customer churn prediction based on k-means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(2), 458–475. <https://doi.org/10.3390/jtaer17020024>
- [15] AL-Najjar, D., Al-Rousan, N., & AL-Najjar, H. (2022). Machine learning to develop credit card customer churn prediction. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(4), 1529–1542. <https://doi.org/10.3390/jtaer17040077>
- [16] Lukita, C., Bakti, L. D., Rusilowati, U., Sutarman, A., & Rahardja, U. (2023). Predictive and analytics using data mining and machine learning for customer churn prediction. *Journal of Applied Data Sciences*, 4(4), 454–465. <https://doi.org/10.47738/jads.v4i4.131>
- [17] Faritha Banu, J., Neelakandan, S., Geetha, B. T., Selvalakshmi, V., Umadevi, A., & Martinson, E. O. (2022). Artificial intelligence based customer churn prediction model for business markets. *Computational Intelligence and Neuroscience*, 2022(1), 1703696. <https://doi.org/10.1155/2022/1703696>
- [18] Vo, N. N. Y., Liu, S., Li, X., & Xu, G. (2021). Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Systems*, 212, 106586. <https://doi.org/10.1016/j.knsys.2020.106586>
- [19] Agarwal, V., Taware, S., Yadav, S. A., Gangodkar, D., Rao, A., & Srivastav, V. K. (2022). Customer—churn prediction using machine learning. In *2022 2nd International Conference on*

- on *Technological Advancements in Computational Sciences*, 893–899. <https://doi.org/10.1109/ICTACS56270.2022.9988187>
- [20] Usman-Hamza, F. E., Balogun, A. O., Capretz, L. F., Mojeed, H. A., Mahamad, S., Salihu, S. A., . . . , & Salahdeen, N. K. (2022). Intelligent decision forest models for customer churn prediction. *Applied Sciences*, 12(16), 8270. <https://doi.org/10.3390/app12168270>
- [21] Poudel, S. S., Pokharel, S., & Timilsina, M. (2024). Explaining customer churn prediction in telecom industry using tabular machine learning models. *Machine Learning with Applications*, 17, 100567. <https://doi.org/10.1016/j.mlwa.2024.100567>
- [22] Sana, J. K., Abedin, M. Z., Rahman, M. S., & Rahman, M. S. (2022). A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection. *PLOS One*, 17(12), e0278095. <https://doi.org/10.1371/journal.pone.0278095>
- [23] Li, A., Yang, T., Zhan, X., Shi, Y., & Li, H. (2024). Utilizing data science and AI for customer churn prediction in marketing. *Journal of Theory and Practice of Engineering Science*, 4(05), 72–79. [https://doi.org/10.53469/jtpes.2024.04\(05\).10](https://doi.org/10.53469/jtpes.2024.04(05).10)
- [24] Asif, D., Arif, M. S., & Mukheimer, A. (2025). A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry. *Results in Engineering*, 26, 104629. <https://doi.org/10.1016/j.rineng.2025.104629>
- [25] Usman-Hamza, F. E., Balogun, A. O., Nasiru, S. K., Capretz, L. F., Mojeed, H. A., Salihu, S. A., . . . , & Awotunde, J. B. (2024). Empirical analysis of tree-based classification models for customer churn prediction. *Scientific African*, 23, e02054. <https://doi.org/10.1016/j.sciaf.2023.e02054>
- [26] Rajarajan, K., Vishal Reddy, D., & Priya, S. (2024). Anticipating customer churn in telecommunication using machine learning algorithms for customer retention. In *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics*, 1–7. <https://doi.org/10.1109/ICDCECE60827.2024.10548668>
- [27] J, S., Gangadhar, C., Arora, R. K., Renjith, P. N., Bamini, J., & Chincholkar, Y. D. (2023). E-commerce customer churn prevention using machine learning-based business intelligence strategy. *Measurement: Sensors*, 27, 100728. <https://doi.org/10.1016/j.measen.2023.100728>
- [28] Sudharsan, R., & Ganesh, E. N. (2022). A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. *Connection Science*, 34(1), 1855–1876. <https://doi.org/10.1080/09540091.2022.2083584>
- [29] Prabadevi, B., Shalini, R., & Kavitha, B. R. (2023). Customer churning analysis using machine learning algorithms. *International Journal of Intelligent Networks*, 4, 145–154. <https://doi.org/10.1016/j.ijin.2023.05.005>
- [30] De, S., P, P., & Paulose, J. (2021). Effective ML techniques to predict customer churn. In *2021 Third International Conference on Inventive Research in Computing Applications*, 895–902. <https://doi.org/10.1109/ICIRCA51532.2021.9544785>

How to Cite: Nagarajan, R., Raju, S. S., & Atheeswaran, A. (2026). Data-Driven Churn Prediction: Evaluating ML Models for Business Retention Strategies. *Journal of Computational and Cognitive Engineering*. <https://doi.org/10.47852/bonviewJCCE62026861>