PREDICTING CUSTOMER CHURN USING MACHINE LEARNING AND DATA MINING TECHNIQUES

***Abdul Ahmed Abdul[1], Bashir Shehu[2], Abubakar Muhammad Abbas[3]**

[1]Infomation Technology Department,
Shamrock Innovations.
[2]Computer Science Department,
Ahmadu Bello University.
[3]Computer Science Department,
Bayero University Kano.
*Corresponding author email: ahmadabdul592@gmail.com, Tel: +2348032642267

Abstract
Customer churn is a significant and primary concern for large companies, especially in the telecommunications sector, due to its direct impact on revenue. As customers are the main source of profit, retaining them is crucial. Churners are individuals who leave a company's service for various reasons. Accurately predicting customer behavior is essential for reducing churn rates. Research indicates that acquiring a new customer is more costly than retaining an existing one. Therefore, to retain customers, service providers need to understand the reasons for churn, which can be uncovered through data analysis. This paper develops a churn prediction model to help companies identify customers most likely to churn. The model utilizes machine learning techniques, to determine the primary factors driving customer churn and to identify the most effective prediction algorithm. The dataset comprises demographic details, data usage, and service types of 3,333 customers, spanning 11 attributes, sourced from Global AI. This investigation explores the use of these algorithms for churn prediction and highlights the growing importance of data services for subscribers. As such, these insights are valuable for both service providers and researchers in understanding service quality from the perspective of mobile subscribers which can lead to a revaluation of existing benchmarks and an enhancement of customer satisfaction. Random Forest algorithm was found to yield the best results in predicting customer churn.

Keywords: Customer churn, telecommunications, machine learning, churn prediction model, data analysis, customer retention, prediction algorithms.

Introduction

Emphasizing the higher costs associated with attracting new customers compared with retaining existing customers, and the fact that long-term customers tend to produce more profits, (Verbeke et al., 2011) assert that customer retention increases with profitability. Many competitive organizations have realized that a key strategy for survival within the industry is to retain existing customers. Tsai and Chen., (2010) argued that “this leads to the importance of churn management.” Customer churn represents a basic problem within the competitive atmosphere of the telecommunication industry. The telecommunications sector has become one of the main industries in developed and developing countries, and the technical progress and the increasing number of operators raised the level of competition (Gerpott et al., 2001). Companies are working hard to survive in this competitive market depending on multiple strategies. Tree main strategies have been proposed to generate more revenues (Wei & Chiu 2002):

- acquire new customers,
- upsell the existing customers, and
- increase the retention period of customers.

However, comparing these strategies taking the value of return on investment (RoI) of each into account has shown that the third strategy is the most profitable (Wei & Chiu, 2002), proving that retaining an existing customer costs much less than acquiring a new one (Qureshii et al., 2013), in addition to being considered much easier than the upselling strategy (Ascarza, 2016). To apply the third strategy, companies have to decrease the potential of customer churn, known as "the customer movement from one provider to another" (Bott. 2014). The development and digitalization of the world have led to new ways of doing business and companies all over the globe have been forced to adapt (Shkurti and Muca, 2014). Subscription-based services are one of the outcomes of the explosive digitalization that has taken the world by storm with this, comes both possibilities and challenges that require modern-day solutions (Blank and Hermansson, 2018). The digitalization era has transformed the business landscape, not only in terms of operational processes but also by providing consumers with a vast array of subscription-based service options. This abundance of choices poses a significant challenge for companies, as retaining customers may become increasingly difficult. However, digitalization within organizations can lead to reduced labor costs, enhanced efficiency, and a more comprehensive understanding of internal operations. These factors are crucial for maintaining a competitive edge and gaining an advantage over rival companies in the market (Shkurti and Muca, 2014).

More so, as information technology is growing, the available amount of data and information has increased significantly during the past years. This rapid growth has enabled storage and processing of great amounts of data while increasing the necessity of automatically finding valuable information and creating knowledge (Buo and Kjellander, 2014) with meaningful information extracted from the stored data, firms can make appropriate decisions to grow the business. With the rapid growth in available data and information, the application of data mining techniques and machine learning algorithms has seen a surge, due to their ability to handle and analyze great amounts of data (Mishra and Rani, 2017). Digitalization has also brought forward an ongoing trend to improve current data processing activities as a part of customer relationship management (CRM) strategies. The idea of knowledge management and customer relationship management has lately obtained more attention in the subscription-based business model and the concepts focus on the distribution of resources to customer-centric activities, to be able to increase competitive advantages (Gebert, 2003 and Sergue, 2020). Customer knowledge is the knowledge businesses can obtain through customer interaction (Khodakarami and Chan, 2014). Customer relationship management systems support interaction between businesses and customers to collect, store, and analyze data to get an overview of their customers (Khodakarami and Chan, 2014). Such systems have evolved through the past years and by using technology and different data analysis tools, enterprises can find patterns in customer's behaviour, which would be almost impossible to discover manually (Buo and Kjellander, 2014). Such patterns could vary from a customer's purchasing behaviour or patterns related to customer churning. In a subscription-based business model, a fundamental part of success is to minimize the rate of customers ending their subscriptions, in other words, to minimize churn (Buo and Kjellander, 2014).

Customer churning refers to the action of when a customer chooses to abandon their service provider (Garcia et al., 2017). The term is relatively new and has gained more relevance with the emergence of online services. Firms across the globe recognize customer churning as a great loss since they have already invested in attracting these customers. This is one of the major reasons that customer retention is beneficial for a firm. Customers can churn for many reasons and it is hard to pinpoint a general reason for churning. The availability of information has given consumers a bargaining power, and nowadays customers can easily find the service provider, that provides the same product with a more satisfying deal (Gordini and Veglio, 2017). To manage this, firms invest in customer churn prediction, which means that companies try to predict which of their customers will churn to apply preventative measures. These preventive measures could differ depending on the reason a customer might churn, and could be for example, offering a lower price or including an extra service. As mentioned earlier, analysing customer behavior is the basis for predicting customers who might churn, which is important for many reasons. One reason is that for companies who rely on subscription-based income, it can make a big difference on whether they can keep a steady income level or need to change their services to keep customers. Another reason is that, compared to retaining customers, attracting new ones is costlier and firms can save money by retaining their existing customer base (Ullah, 2019).

More so, with new companies continuously emerging in the telecommunications industry, and new customers to those companies sparse, markets have matured to a point where they have saturated, so possible new customer acquisition is scarce. Because of this, companies are beginning to recognize that their most valuable assets are their existing customers. Holding onto existing customers is not an undemanding procedure, with further complications arising from governing bodies such as NCC. These authorities have emerged to act in the customers' best interests, and the promotion of healthy competition amongst service providers to ensure that the customer has a choice, and services are not monopolized. A further challenge that is commonly being recognized regarding the acquisition of new customers is lack of data, making targeted sales campaigns difficult to deploy.

In today's business world companies are recognizing that customer value and increased revenue is more likely to come from their existing customer base than from new customer acquisition. The reasoning behind this is that companies know their existing customers, already have a relationship with them, and amplitude of data on them. In common recognition of this problem, industry has seen an emergence of customer relationship management (CRM) products. Software companies have realized that CRM has become a receptive topic within industry and therefore an area of potential wealth and opportunity for the development and promotion of products that promise to not only boost customer retention, but also for increasing selling opportunities (Hadden et al., 2007).

Also, in the past years, companies have been able to store and process huge amounts ofdata while realizing that being customer-centric was becoming a main requirement to stand out from the competition. Indeed, due to saturated markets, focusing on Customer Relationship Management (CRM) to retain an existing customer base is not optional anymore, but an absolute necessity for competitive survival. In its research for Bain and Company, Frederick Reichheld stated that the cost of acquiring a new customer could be higher than that of retaining a customer by as much as 70%, and that increasing customer retention rates by a mere 5% could increase profits by 25% to 95%. More generally, data-driven decision-making is way for businesses to make sure their next move will benefit both them and their customers. Almost every company, especially in the Tech ecosystem has now put into place a tracking process, to gather data related to their customers' behavior. The data to track varies along with the specific business model of each company and the problem service they aim to address. By analyzing how, when, and why customers behave a certain way, it is possible to predict their next steps and have time to work on fixing issues beforehand. This prediction and quantification of the risk of losing customers can be done globally or individually and is mainly used in areas where the product or service is marketed on a subscription basis. Churn prediction is generally made by studying consumer behavior or observing individual behavior that indicates a risk of attrition. It involves modeling and machine learning techniques that can sometimes use a considerable amount of data.

Related Works

Prabadevi et al., (2023) focus on customer churning analysis using machine learning algorithms to predict client churn early on, allowing businesses to take proactive measures to retain customers. The study tests various algorithms like stochastic gradient booster, random forest, logistics regression, and k-nearest neighbors, achieving accuracies ranging from 78.1% to 83.9%. This demonstrates the effectiveness of machine learning in predicting customer churn accurately. Lalwani et al., (2021), conduct a study related to churn prediction using a machine learning approach. They use Logistic Regression, Naïve Bayes, Support Vector Machine, Decision Tree, Random Forest, and various boosting algorithms such as XGBoost and AdaBoost; to conduct research of this magnitude, they use comprehensive data from a telecom provider that consists of 7000 instances. Also, the dataset includes 21 features, categorized into numerical and categorical values which consist of usage information such as number of calls, but also specific customer data, such as gender. The accuracy of their study ranges from 74 to 82 percent, recall from 74 to 82 percent, precision from 57 to 81 percent, and F-measure from 63 to 81 percent. They conclude that the boosted classifiers, XGBoost and AdaBoost, performed best and achieved the best accuracy.

Vafeiadis et al., (2015) compare different machine learning classifiers to predict customer churning. The data used in their study is provided by a telecom provider and consists of 5000 samples and multiple variables. The data mostly contains usage information such as call duration and number of texts sent. In addition, specific subscription information is also used, such as subscription period. The different classifiers compared are Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes, and Logistic Regression. They use boosting to compare the single learners with their boosted versions; however, Naïve Bayes and Logistic Regression are not boosted due to imitations in their parameters. To evaluate the performance of the different classifiers, they use precision, recall, accuracy, and F-measure, which are calculated from the confusion matrix. Based on these measures they conclude that boosting significantly improves the performance in terms of both accuracy and F-measure for all of the considered classifiers. On the other hand, without the use of boosting, they suggest that the Support Vector Machine is a good tool for customer churn prediction. Idris (2012) proposed an approach based on genetic programming with AdaBoost to model the churn problem in telecommunications. The model was tested on two standard data sets. One by Orange Telecom and the other by cell2cell, with 89% accuracy for the cell2cell dataset and 63% for the other one.

Masarifoglu and Buyuklu (2019), study customer churn prediction using survival analysis, which is a group of statistical methods used for studying events based on certain circumstances, such as customer churn prediction, emphasizing that the question they answer is the time until a certain event occurs. Their model estimation is based on the survival function and the hazard function. They employed the Kaplan-Meier method and the Cox model in R. Using this approach, the preliminary risks of customer churning are explained and quantified. Survival curves and hazard ratios are also presented, which they argue, allow service providers to take preventative measures against churners. They use historical telecom data acquired from a mobile operator for their study, which contains information regarding 10365 randomly selected customers. The data consists of categorical variables represented as dummy variables. These variables represent information regarding subscription type, length of subscription but also customer specific information such as gender. The model estimation is based on the survival function and the hazard function which they agreed that the purpose of this estimation is to identify potential churn characteristics and estimate customer churn by calculating probabilities for survival. Huang et al., (2015) studied the problem of customer churn in the big data platform.

The goal of the researchers was to prove that big data greatly enhances the process of predicting the churn depending on the volume, variety, and velocity of the data. Dealing with data from the Operation Support department and Business Support department at China's largest telecommunications company needed a big data platform to engineer the fractures. Random Forest algorithm use and evaluated using AUC.

Makhtar et al., (2017) proposed a model for churn prediction using rough set theory in telecom. As mentioned in his paper Rough Set classification algorithm outperformed other algorithms like Linear Regression, Decision Tree, and Voted Perception Neural Network. Various researchers studied the problem of unbalanced data sets where the churned customer classes are smaller than the active customer classes, as it is a major issue in churn prediction problems. Amin et al., (2016) compared six different sampling techniques for oversampling regarding telecom churn prediction problem. The results showed that the algorithms (mega-trend diffusion function (MTDF) and rules-generation based on genetic algorithms) outperformed the other compared oversampling algorithms.

Guo-en and Wei-dong (2008) focused on building a customer churn prediction model using SVM in the telecommunication industry. They compared this method with other techniques such as decision tree (DT), artificial neural networks, naïve Bayesian (NB), and logistic regression. The results proved SVM to be a simple classification method of high capability yet good precision. Kumar and Ravi (2008) used data mining to predict credit card customer churn. They used multilayer perceptron (MLP), logistic regression, Decision tree (DT), random forest, radial basis function, and SVM techniques.

Nie et al., (2011) built a customer churn prediction model by using logistic regression and Decision tree (DT) based techniques within the context of the banking industry. In their study, Lin et al. (2011) used rough set theory and rule-based decision-making techniques to extract rules related to customer churn in credit card accounts using a flow network graph (a path-dependent approach to deriving decision rules and variables). They further showed how rules and different kinds of churn are related.

Sharma and Panigrahi (2011) applied neural networks to predict customer churn from cellular network services. The results indicated that neural networks could predict customer churn with an accuracy of higher than 92%. Yu et al., (2011) applied neural network, SVM, DT, and extended SVM (ESVM) techniques to forecast customer churn;

of the methods studied, ESVM performed best. Dahiya and Talwar. (2015) presented new features-based logistic regression (LR), linear classifier (LC), NB, DT, MLP neural networks, and SVM. In their experiments, each technique produced a different output. Data mining by evolutionary learning (DMEL) could show the reason or probability of a churning phenomenon; DT, however, could only show the reason. LR, NB, and MLP could provide probabilities of different customer behaviors. LC and SVM could distinguish between a churner and a non-churner.

Burez and Van den Poel (2009) studied the problem of unbalanced datasets in churn prediction models and compared the performance of Random Sampling, Advanced Under Sampling, Gradient Boosting Model, and Weighted Random Forests. They used (Area under curve (AUC), Lift) metrics to evaluate the model. The result showed that under sampling technique outperformed the other tested techniques.

Brandusoiu, et al. (2016) presented an advanced methodology of data mining to predict churn for prepaid customers using a dataset for call details of 3333 customers with 21 features, and a dependent churn parameter with two values: Yes/No. Some features include information about the number of incoming and outgoing messages and voicemail for each customer. They applied the principal component analysis algorithm (PCA) to reduce data dimension. Tree machine learning algorithms: Neural Networks, Support Vector Machine, and Bayes Networks were used to predict churn factor. They used Area Under Curve (AUC) to measure the performance of the algorithms. The UAC values were 99.10%, 99.55% and 99.70% for Bayes Networks, Neural Networks and support vector machine, respectively.

Methodology

Figure 1 below shows the technique employed in the sttudy and the methodology is adapted from Marsland (2014) as depicted. The proposed machine learning process includes data collection and preparation, feature selection, algorithm choice, parameter and model selection, training, and evaluation. We do not consider the proposed machine learning process as a linear process as such there is a lot of back and forth between the steps to create the most optimal model, in our case a churn prediction model.

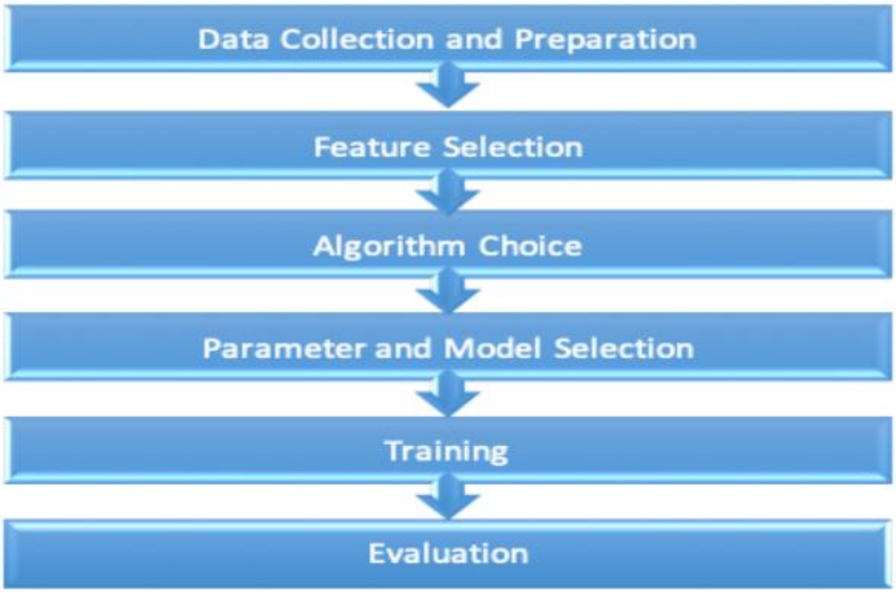

Figure 1: Machine learning process proposed by Marsland (2014)

Data Collection and Preparation: Once the insights to look for are identified, the data sources necessary for further predictive modeling can be decided. The dataset used for this research contains demographic details of customers, their total charges, and the type of service they receive from the company. It comprises churn data of over 3333 customers spread over 11 attributes obtained from Global AI (Figure 2). It can be used to analyze all relevant customer data and develop focused customer retention programs, implying that each row represents a customer, and each column contains the customer's attributes described in the column Metadata.

```
# Read csv
data = pd.read_csv("churn.csv")
data.head()
```

	Churn	AccountWeeks	ContractRenewal	DataPlan	DataUsage	CustServCalls	DayMins	DayCalls	MonthlyCharge	OverageFee	RoamMins
0	0	128	1	1	2.7	1	265.1	110	89.0	9.87	10.0
1	0	107	1	1	3.7	1	161.6	123	82.0	9.78	13.7
2	0	137	1	0	0.0	0	243.4	114	52.0	6.06	12.2
3	0	84	0	0	0.0	2	299.4	71	57.0	3.10	6.6
4	0	75	0	0	0.0	3	166.7	113	41.0	7.42	10.1

Figure 2 A Snapshot of the dataset used

Feature Selection: Consists of determining the most impactful features for a problem. Feature selection is used to identify the most relevant features and is often used due to its performance-enhancing properties (Tang et al., 2014). This requires in-depth knowledge of the data being used to determine which features should be included (Marsland, 2014). According to Tang et al., (2014), many features can result in overfitting, meaning that the model learns from what can be considered noise to the point that the result is negatively affected. Hence, learning from noise complicates generalization and the ability to correctly classify unseen data instances. We decided to identify non-important features using a less complicated and straightforward approach, where the importance of the features is scored based on their prediction ability. We ranked the features based on their importance score and removed all non-important features with a score below 0.05. There exist more advanced methods for feature selection, such as recursive feature elimination (RFE) (Chen and Jeong, 2007). We decided not to use such methods since they are more time-consuming and exceed the purpose of our study. The results we obtained were sufficiently good using feature importance. Since Naïve Bayes assumes that each feature is independent, feature selection using feature importance was not performed.

Algorithm Choice: Based on the scope of the problem, the optimal algorithm for the considered problem can vary. There are numerous algorithms to choose from when faced with a problem, for instance prediction tasks. However, not every algorithm is suitable and fits the criteria for the task at hand, which is why several algorithms are rejected. Since the problem we are trying to solve is a classification problem, all algorithms used for regression problems are rejected. Common classifiers, such as Naïve Bayes, Random Forest, XGBoost, Support Vector Machine, Logistic Regression, and Hidden Markov's Model are used in prediction problems, however, previous studies show that ensemble learners are preferred for classification problems and they are rapidly becoming the standard choice among algorithms due to their performance enhancing characteristics (Kumar and Jain, 2020; Van Wezel and Potharst, 2007). Hence, our choice of algorithms is based on previous studies. We chose to use Decision Tree, which applies classification and regression tree algorithm (CART); Random Forest, which applies bagging and Logistic Regression which applies stacking, which performs well for several real-world problems including churn prediction.

Parameter and Model Selection: Many machine learning algorithms need to be fed with parameters such as the number of estimators. Parameter turning is the process of optimizing the parameters of a machine learning algorithm and choosing the ones that obtain the best result for the desired problem. In our study, we investigate how machine learning can be used to predict customer churning within the telecommunication industry; hence we considered parameter tuning to be beyond the scope of this study. In addition, we considered it to be too big of a task to accomplish. We decided to use default parameters and values for our algorithms and models.

Training: The process of training refers to using labelled data to build a model, which preferably performs well on unseen instances. The dataset should be split in a training set used for training and fitting of a model, and a test set used for evaluation. Supervised learning refers to learning by examples, meaning that each example contains inputs (features) and a corresponding output. To train the model, correlations are found between inputs and outputs. We decided to use supervised learning since our dataset is labelled and includes inputs with the corresponding output for each customer. We decided to randomly split the dataset into a training set and a test set with the proportions of 70:30 percent. Unfortunately, there is no distinct answer to how the split between training and testing data should be done and the choice is made from the researchers. Previous studies related to machine learning have used different splits such as 50/50, 60/40, 70/30, 80/20, and 90/10, but there is no research that suggests that one proportion preferred over another (Raschka, 2018). However, researchers argue that the bigger the dataset, the less test data is needed. For example, if there is data that contains 200 000 data instances, 10% can be more than enough to see if the model performs adequately but if the sample size contained 3000 data instances, 10% might be too little to validate the performance of the model and we might need to choose a larger test sample. On the other hand, there is no distinct answer to what a big or small dataset is. After consideration and a discussion with a domain expert we came to the conclusion that our dataset, including approximately 3333 data instances, is relatively big and we chose to split our dataset with the proportions of 70:30 percent.

Evaluation: Evaluation is an important part of the machine learning process, where the performance of the classifier should be evaluated. In order to select the optimal machine learning algorithm different indicators are used for different types of problems (Hongquig et al., 2018). In previous literature regarding churn prediction, the majority of researchers evaluate their model's using accuracy, precision, recall, and F-measure, which are all calculated from the confusion matrix. We chose to evaluate the performance and efficiency of our algorithms using the above-mentioned metrics. In the case of imbalanced data, accuracy is not the most optimal metric used for evaluation since it does not always fully reflect the performance of the algorithm (Tyagi and Mittal, 2020). To handle this problem, we evaluate the classifiers based on precision, recall, and F1 score on the target class, that is Churn. E valuation is done on unseen instances from the test set, in other words data, which the algorithm has not been trained on. This is why the data should be split into a training and a test set to be able to determine how well the model performs on examples not used during training.

Evaluation Metrics: Machine learning models are not always perfect for the given data and are in need of evaluation to see how well the model performs. The most common ways to evaluate binary classifiers are using certain metrics, which are accuracy, recall, precision and F1-score (Sokolova and Lapalme, 2009). These metrics can be calculated through a confusion matrix, shown in table below.

	Actual	
Predicted	True Positive (TP)	False Negative (FN)
	False Positive (FP)	True Negative (TN)

Table 1 Confusion Matrix for Binary Classification

A confusion matrix is a machine learning concept, which includes information about the actual and predicted classifications, used to describe the performance of the classifier (Deng et al., 2016). True Positives (TP) and True Negatives (TN) represent the correctly classified test instances while False Negatives (FN) and False Positives (FP) represent the incorrectly classified test instances (Witten et al., 2016).

Accuracy is a measure that shows the overall effectiveness of the classifier (Sokolova and Lapalme, 2009). It is a metric showing the rate of total correctly classified instances. According to Deng et al. (2016), accuracy is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad \text{(Eq.1)}$$

Precision is a measure that shows the proportion of correctly predicted positive instances. The metric shows how often the model is correct when predicting the target class, in our application, churners. According to Deng et al., (2016) precision shows the accuracy of predicting a specific class and is calculated as follows:

$$Precision = \frac{TP}{TP+FP} \quad \text{(Eq.2)}$$

Recall is a measure, which shows the effectiveness of the classifier to determine examples labelled as positive (Sokolova and Lapalme, 2009). It shows the ability of the binary classifier to identify instances of a specific class (Deng et al., 2016). Recall is calculated as follows:

$$Recall = \frac{TP}{TP+FN} \quad \text{(Eq.3)}$$

The F-score is often used for evaluating the performance of a classifier. The F-score is a measure that takes both precision and recall into consideration and is typically defined as the harmonic mean of precision and recall. A better combined recall and precision is achieved as F-score is closer to 1. (Vafeiadis et al., 2015)

$$F - Score = \frac{2*Precision*Recall}{Precision+Recall} \quad \text{(Eq.4)}$$

Experiment and Result Analysis

The results were obtained using Python 3.10.6 by utilizing the Jupyter Libraries from Anaconda. The various libraries used include scikit learn, numpy, pandas, matplotlib and seaborn. The results obtained in comparing the performance of the various algorithms are narrated below. A thorough comparison of algorithms based on the metrics mentioned above gives a comprehensive insight into the performance and efficiency of each of them. From the table below, we observe that by comparing the three models, random forest seems to yield the highest level of accuracy. it also has a better f1-score for predicting customer churn, we concluded that Random Forest is the best model of the three models used.

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.8770351328191945	0.5618279569892473	0.6276276276276276	0.5929078014184397
Random Forest	0.921165381319623	0.8117154811715481	0.5825825825825826	0.6783216783216783
Logistic Regression	0.8641816623821765	0.58	0.17417417417417416	0.2678983833718245

Table 2 comparison of model using evaluation metrics

More so, accuracy could be a misleading metric in the case of an imbalanced dataset, since it does not tell much on how the classifiers perform on specific classes. However, in our research, it is more important to predict actual churners with high accuracy and minimizing False Negatives rather than minimizing the number of False Positives, that is, non-churners classified as churners. In Table 2 above, we present the performance of the classifiers for the target class, that is churners. In our study, the focal point is to predict churners with a good accuracy. By analyzing the results, Logistic Regression does not perform well in predicting churners, in particular with the default imbalanced dataset, the results were far from adequate. The Random Forest classifier and the Decision Tree classifier perform remarkably better on churn prediction than the Logistic Regression classifier does. The Decision Tree classifier and the Random Forest Classifier perform similarly.

Furthermore, our results show that for the considered classifiers, Logistic Regression, Decision Tree, and Random Forest, the overall accuracy ranges between 86.4% and 92.1%, indicating an overall good performance as shown in the figure below.

Figure 2: Model Performance

On the other hand, one can question what a suitable score of accuracy is. Accuracy measures the classifiers' ability of correctly classifying all the test instances, in our case including both non-churners and churners. But what is actually a good score? This is a common question when evaluating machine learning algorithms, and unfortunately there is no distinct answer to this question. We argue that a good score of accuracy depends on the considered application and what prediction problem is investigated. 90% accuracy in one application may not be as good as 90% in another application. For example, for email spam detection, 90% accuracy is not bad, meanwhile that only 10% of the emails are wrongly classified. In medical predictions, regarding potential patients in a certain risk group, 10% wrong classifications are far from acceptable. It should be emphasized that accuracy alone is not a sufficient metric used for evaluation (Sergue, 2020), which is also shown by the above examples. In particular, when the considered dataset is imbalanced to a high degree, accuracy can be a misleading metric. In our case, the used dataset consists of 98% non-churners and 2% churners, which in fact could give us a 98% accuracy by predicting all test instances as non-churners. If this would be the case, the result is far from acceptable and a 98% accuracy would not be good. Based on this reasoning, accuracy gives in our case a skewed view of the actual performance of the classifiers. Based on the above reasoning, we can therefore not assume that the performance of the classifier is good by only looking at accuracy. Alternative evaluation metrics, which are commonly used in the literature are precision, recall, and F1-score, all indicating the performance of the classifier on a specific target class. In our case the target class is churn, and our results based on the above-mentioned alternative metrics show that, except for Decision Tree and Logistic Regression, Random Forest has the best obtained precision of 81.1%, recall of 58.3%, and F1-score of 67.8% is actually good at predicting churners in our application.

Reference

Amin A, Anwar S, Adnan A, Nawaz M, Howard N, Qadir J, Hawalah A, and Hussain A. (2016) Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. IEEE Access. 4:7940–57

Anil Kumar D. and Ravi V, (2008) Predicting credit card customer churn in banks using data mining. Int J Data Anal Tech Strateg 1(1):4–28

Ascarza E, Iyengar R, and Schleicher (2016) M. The perils of proactive churn prevention using plan recommendations: evidence from a feld experiment. J Market Res; 53(1):46–60

Blank C and Hermansson T, (2018). A Machine Learning approach to churn prediction in a
subscription-based service, KTH, Stockholm.

Bott. (2014) Predicting customer churn in telecom industry using multilayer perceptron neural networks: modelling and analysis. Igarss;11(1):1–5

Brandusoiu I, Toderean G, and Ha B. (2016) Methods for churn prediction in the prepaid mobile telecommunications industry. International conference on communications. p. 97–100

Buo D and Kjellander M, (2014)"Predicting Customer Churn at a Swedish CRM-system Company," Linkoping’s Universitet, Linkoping.

Burez D and Van den Poel. (2009) Handling class imbalance in customer churn prediction. Expert Syst Appl. 36(3):4626–36

Chen XW and Jeong JC, (2007) "Enhanced recursive feature elimination," in Sixth International Conference on Machine Learning and Applications (ICMLA 2007).

Dahiya, K. and Talwar, K. (2015) ‘Customer churn prediction in telecommunication industries using data mining techniques – a review’, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, No. 4, pp.417–433.

Deng X, Liu Q, Deng Y, and Mahadevan S (2016) An improved method to construct basic probability assignment based on the confusion matrix for classification problem. Inf Sci 340–341:250–161

Farquad M, Ravi V, and Raju S (2014) Churn prediction using comprehensible support vector machine: An analytical CRM application. Applied Soft Computing 19:31–40

Garcia DL, Nebot A and Vellido A, (2017)"Intelligent data analysis approaches to churn as
a business problem: a survey," Knowledge and Information Systems, vol. 51, no. 3, pp. 1-56.

Gebert H, Geib M, Kolbe L and Brenner W, (2003) "Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts," Journal of Knowledge Management, vol. 7, no. 5, pp. 107-123.

Gerpott TJ, Rams W, Schindler A. (2001) Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommunication Policy. 25:249–69

Gordini N and Veglio V, (2017) "Customer's churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry," Industrial Marketing Management, vol. 62, pp. 100-107.

Guo-en X and Wei-dong J (2008) Model of customer churn prediction on support vector machine. Syst Eng Theory Pract 28(1):71–77

Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer-assisted customer churn management: State-of-the-art and future trends. Computers and Operations Research 34(10):2902–2917

Hongqing G, Peiyong S, Wenzhong G and Kun G, (2018) "Component-based Assembling Tool and Runtime Engine for the Machine Learning Process," in International Conference on Cloud Computing, Big Data and Blockchain.

Huang F, Zhu M, Yuan K, Deng EO. (2015) Telco churn prediction with big data. In: ACM SIGMOD international conference on management of data. p .607–18

Idris A, Khan A, Lee YS. (2012) Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). p. 1328–32

Khodakarami F and Chan Y, (2014) "Exploring the role of customer relationship management (CRM) systems in customer knowledge creation," Information & Management, vol. 51, pp. 27-42.

Kumar A. and Jain M, (2020) Ensemble Learning for AI Developers: Learn Bagging, Stacking, and Boosting Methods with Use Cases, Apress.

Lalwani P, Kumar MM, Singh J, Chadha and Sethi P, (2021) "Customer churn prediction system: a machine learning approach," Computing, pp. 1-24.

Lin CS, Tzeng GH, and Chin YC (2011) Combined rough set theory and flow network graph to predict customer churn in credit card accounts. Expert Syst Appl 38:8–15

Makhtar M, Nafs S, Mohamed M, Awang M, Rahman M, Deris M. (2017) Churn classification model for local telecommunication company based on rough set theory. J Fundam Appl Sci.9(6):854–68.

Marsland S, (2014) Machine Learning: An Algorithmic Perspective, 2nd edition ed., Chapman and Hall/CRC.

Masarifoglu M and Buyuklu AH, (2019)"Applying Survival Analysis to Telecom Churn Data," American Journal of Theoretical and Applied Statistics, vol. 8, no. 6, pp. 261-275.

Mishra K and Rani R, (2017)"An inclusive survey on machine learning for CRM: a paradigm shift," in 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India.

Nie G, Rowe W, Zhang L, Tian Y, Shi Y (2011) Credit card churn forecasting by logistic regression and decision tree. Expert Systems with Applications 38(12):15,273–15,285

Prabadevi. B, Shalini. R. S, Kavitha. B. (2023). Customer churning analysis using machine learning algorithms. International journal of intelligent networks, doi: 10.1016/j.ijin.2023.05.005

Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. (2013) Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. p.131–6

Raschka S, (2018) "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," University of Wisconsin–Madison.

Sergue M, (2020)"Customer Churn Analysis and Prediction using Machine Learning for a B2B SaaS company," KTH, Stockholm.

Sharma A and Panigrahi PK (2011) A neural network-based approach for predicting customer churn in cellular network services. Int J Comput Appl 27(11):26–31

Shkurti R and Muca A (2014) "An Analysis of Cloud Computing and Its Role in Accounting Industry in Albania," Romanian-American University, vol. 8, no. 2, pp. 219-229.

Sokolova M and Lapalme G, (2009) "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427-437.

Tang J, Alelyani S and Li H, (2014) "Feature Selection for Classification: A Review," p. 37.

Tsai CF and Chen MY (2010) Variable selection by association rules for customer churn prediction of multimedia on demand. Expert Syst Appl 37:2006–2015.

Tyagi S and Mittal S, (2020) "Sampling Approaches for Imbalanced Data Classification Problem in Machine Learning," in Proceedings of ICRIC 2019.

Ullah I, Raza B, Malik AK, Imran M, Islam SU and Kim SW, (2019) "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," IEEE Access, pp. 60134-60149.

V. Verdhan, (2020) Supervised Learning with Python: Concepts and Practical Implementation Using Python, publisher logo Apress.

Vafeiadis T, Diamantaras K, Chatzisavvas K (2015) A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory 55:1–9.

Van Wezel M and Potharst R, (2007) "Improved customer choice predictions using ensemble
methods," European Journal of Operational Research, vol. 181, no. 1, pp. 436-452.

Verbeke W, Dejaeger K, Martens D, Hur J, and Baesens B (2011) New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. European Journal of Operational Research 218:211–219

Wei CP and Chiu LT (2002) Turning telecommunications call details to churn prediction: a data mining approach. Expert Systems with Applications 23(2):103–12

Witten IH, Frank E, Hall MA and Pal CJ, (2016) Data Mining: Practical Machine Learning Tools and Techniques, San Francisco: Elsevier Science & Technology.

Yu X, Guo S, Guo J, Huang X (2011) An extended support vector machine forecasting framework for customer churn in e-commerce. Expert Syst Appl 38:1425–1430

Zineldin, M. (2000) TRM: Total Relationship Management