Original Research Article

Customer Churn Prediction using Machine Learning Models

Abstract

Customer churn is a critical concern for the telecommunication industry. Understanding and predicting customer churn can lead to more effective retention strategies and an increase in profitability. Predicting customer churn allows telecommunication companies to identify potentially dissatisfied customers early on and take proactive measures to retain them. Due to a large client base, the telecom industry generates a large volume of data on a daily basis. Decision makers and business analysts stressed that acquiring new customers is more expensive than retaining existing ones. Business analysts and customer relationship management (CRM) analysts must understand the reasons for customer churn as well as behaviour patterns from existing churn data. This paper proposes a churn prediction model that uses classication and clustering techniques to identify churn customers and provides the factors that contribute to customer churning in the telecom sector.

Keywords:machine learning; supervised learning; churn prediction; CRM; telecom; retention

1 Introduction

Business intelligence advancements have become increasingly important in recent years in order to remain effective and competitive in the face of changing business trends. As a result, companies of all sizes (both large and small) have begun to invest in the next level of data analysis and business intelligence. The effective application of business intelligence methods extracts analysis and visualisation of efficiency indicators from significant volume of big data. It lowers the cost and accelerates problem solving with actionable intelligence. However, due to a variety of issues, commercial organisations must make appropriate decisions.With improved business outcomes and established corporate values, the text analytics method makes extensive use of the strength of business intelligence [1]. In businesses, churn management has become vital because it has been proven that, keeping old customers can bring in more profits than getting new consumers [2]. Companies who do not predict consumers who are leaving the business early tend to lose these consumers and obtaining new customers may be time consuming and expensive [3]. Modern companies no longer use traditional-based marketing model to obtain customers, they have also outgrown the use of reports about the past instead the companies now use their acquired data to generate good prediction models with the capacity to supply them with better awareness into the future. With these models, business processes like forecasting of sales are enhanced [4]

With the increase in volume and heterogeneous data, companies now require technologies that will allow them to examine these data [5, 6, 7],. Customer Relationship Management (CRM) is a broad strategy for establishing, managing, and enhancing long-term customer relationships. It is widely recognized and widely applied in a variety of areas, including telecommunications, banking and insurance, retail markets, and so on. Customer retention is one of its key goals. As a result, tools for developing and implementing customer retention models (churn models) are crucial in Business Intelligence (BI) applications. Churning can occur as a result of low customer satisfaction, aggressive competition strategies, new products, regulations, and other factors in today's dynamic market environment. Churn models are designed to detect early churn signals and identify customers who are more likely to quit freely. There has been an increased interest in relevant studies in areas such as the telecommunications industry, banking, insurance companies, gaming and others during the previous decade. As a result, researchers and practitioners believe that developing a customer churn prediction model is critical for maintaining consumers in order to create a model that is both accurate and understandable to discover the clients who wants to churn, as well as their reasons for doing so [8]. The following client types have been identified and they are the customers who are actively looking to switch service providers are known as active churners, passive churners are persons who have had their services cancelled by the companies and finally the silent churners which are clients who may abruptly stop receiving services without warning.

Existing research shows that the primary goal is to identify the valuable churn customer using a large volume of telecom data. However, there are several limitations in existing models that pose significant impediments to solving this problem in the real world. In the telecom industry, a large volume of data is generated, and the data contains missing values, resulting in poor prediction model results. To address the churning prediction problem, machine learning algorithms have been proposed such as Artificial Neural Networks, Decision Trees learning, Regression Analysis, Logistic Regression, Support Vector Machines, Naive Bayes, Sequential Pattern Mining and Market Basket Analysis, Linear Discriminant Analysis and Rough Set Approach. In this paper, a churn prediction model is proposed to evaluate the behaviour of customers. Factors affecting the churning of customers are identified by rules derived from an attribute-based classifier. The results presented shows the efficient of machine learning models in xxxxx. The rest of the paper is organised as follows, a review of related work is presented in section 2. This is followed by the methodology in section 3. The results are presented and analysed in section 4 and the paper is concluded in Section 5.

2 Related Work

Several techniques have been proposed in literature fir churn prediction. These techniques include data mining, machine learning and hybrid strategies. These techniques help businesses identify, predict, and retain churn customers, as well as aid decision-making and CRM. Decision trees are the most commonly used method for predicting problems related to customer churn [9]. Decision trees have the limitation that they are not suitable for complex nonlinear connections between attributes, but perform better for linear data where attributes are interdependent. However, the accuracy of decision trees are improved using pruning [10].

In [11], the authors proposed a forecasting approach that uses a two-phase strategy based on their recency, frequency and monetary value (RFM). The related functions of the RFM are cclassified into four clusters in the first phase. In the second phase, the churn data collected in the first phase are extracted and evaluated using decision trees,Neural Networks and other machine learning algorithms. Experimental results show prediction results are better using hybrid approaches. In [12], the authors proposed a hybrid approach for churn prediction by combining genetic programming with induction algorithm from an existing tree. The proposed algorithm used the behaviour of customers to generate classification rules. The proposed model is used to predict different custom groups based on time of use, location, and underlying social networks, and represents a practical approach to churn models at the human level rather than the mathematical level. The authors in [13] used three models to evaluate the performance of churn models. The models include ANN, classification trees and logistic regression. They selected ten features based on their exploratory data analysis and business experience. Experimental results show that the hybrid models performed better than independent classification models.

AlShourbaji et al.[14] proposed a novel Feature Selection strategy ACO-RSA. In this approach two metaheuristic algorithms algorithms (ant colony optimization (ACO) and reptile search algorithm (RSA)) are integrated for the selection of subset features that are important for churn prediction. The proposed model is evaluated using the state- of-the-art test functions and open-source datasets for churn preductions. This is evaluated alongside Standard ACO, Gray Wolf Optimiser, Multiverse Optimser and the results show that ACO-RSA outperforms the compared approaches. Similarly, the authors in [15] propose a new framework for saturated markets. They use an effective churn prediction model for monitoring customer churn based on Swish Recurrent Neural Network (S-RNN). The proposed model is tuned to classify regular and churn customers based on their network usgae history. In [16], the authors propose a machine learning churning model with six phases. Preprocessing and analysis of features are the first two phases while the third phase is the feature selection phase using gravitational search algorithm. With ratios of 80% and 20%, the data is divided inot tensile and test set respectively. Logistic regression, support vector machines, decision trees, naives bayes and random forest were evaluated and K-fold cross validation was used to optimise the hyperparameters. The results of the experiments show that Adaboost and XGboost outperformed the other approaches compared.

Al-Najjar et al. [17] proposed a churn prediction model for the prediction of credit card cancellations. In their proposed approach feature selection was adopted with five machine learning models. Independent selection of variabes was carried out using k-nearest neighbor selection, two-level clustering and feature selection. Five machine learning models were evaluated incuding c5-tress, Bayesian networks, chi-square trees for automatic interaction detection (CHAID) and neural networks. Experimental results showed that the integration of multiple feature variables improved the accuracy of the performance of the prediction model.Similarly, Zhang et al. [18] proposed a churn prediction model for the telecom industry. Three chinese telecommunication companies were used for data collection. Their predicion model was built using logistic regression and Fisher's discriminant equation. Their results showed that logistic regression outperformed the other model compared with a prediction accuracy of 94%.

3 Proposed Scheme for the Prediction of Customer Churn

In this section, the proposed model is presented in Fig. 1 with detailed description. The first phase of the proposed model is data preprocessing. This is done to filter the data and reduce noise. In this phase, the most important features are selected from retrieval attributes, correlations, filters and ranking. In the third first, a customer classification algorithm is used for the classification of customers to distinguish between churn and non-churn customers. In this paper, Logistic Regression, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision Trees, Random Forest and XGBoost are used for customer classification and prediction.

Figure 1: Prediction Model for Customer Churn

3.1 Data Preprocessing

Machine learning requires preparation of raw data before it can be used for classification in prediction models. This step is called data preprocessing. This is the first and most important step in building a machine learning model. Data used in machine learning projects may not be clean or properly formatted. Real-world data often contains noise and missing values, and is in an unusable form that cannot be used directly in machine learning models. Data preprocessing is a necessary task to clean data to make it suitable for machine learning models and to improve model accuracy and efficiency.

3.2 Dataset Importation

The first requirement for building a machine learning model is a dataset. A dataset is a collection of data in a specific format for a specific problem. To use this feature, data records are stored as CSV files. NumPy, Matplotlib, Pandas, and Sklearn are basic libraries that need to be imported. This library is used to import data sets that are CSV files. The records are converted to NumPy arrays before training. Matplotlib library is used for data visualization. The SyriaTel dataset was used for this project. The dataset consists of 3333 records of customers and 32 columns. The dataset consists of 2850 records of customers that did not churn. It contains 483 records of customers that churn. The dataset is imbalanced because it has more data of customers that did not churn than customers that churned.

3.2.1 Exploratory Data Analysis and Nullity Correlation

Data was analysed using Pandas which generates an HTML format report containing various data statistics. Some of the key features provided by pandas include Type, missing values, and unique values are all essentials. Figure 2 provides an insight on the exploratory data analysis.

Nullity correlation is a technique used to understand patterns of missing values in a dataset. It ranges from -1 to 1. Zero indicates that the variables do not have correlation. -1 indicates a strong negative correlation where +1 indicates a strong positive correlation. Figure 3 shows the visualization of nullity by column.

3.3 Data Cleaning

This first step of data pre-processing (cleaning) is to remove irrelevant features. The features used are ’international plan’, ’voice mail plan’, ’number vmail messages’, ’total day minutes’, ’total day calls’, ’total day charge’, ’total eve minutes’, ’total eve calls’, ’total eve charge’, ’total night minutes’, ’total night calls’, ’total night charge’, ’total intl minutes’, ’total intl calls’, ’total intl charge’, ’customer service calls’.

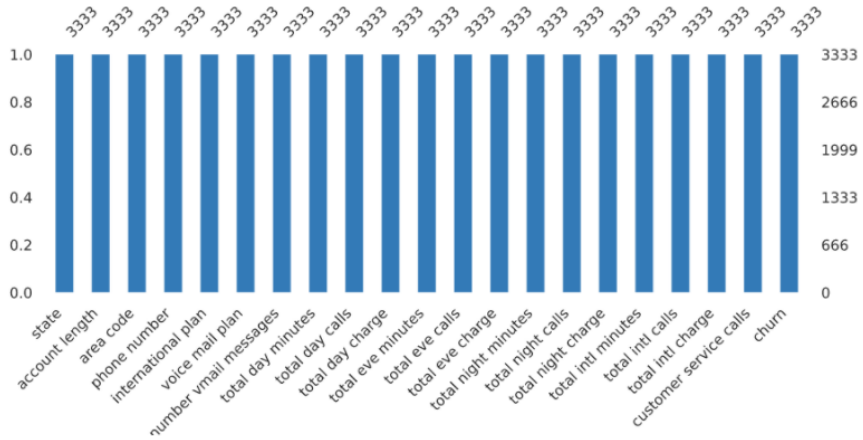

Figure 2: Visualisation of Nullity

3.4 Features Extraction

The next step is to separate the features and the target variable. The "churn" column is the target variable for prediction. Feature extraction is a broad phrase that refers to strategies for creating combinations of variables to get around these issues while still accurately representing the data as shown in Figure 3. Many practitioners of machine learning feel that well optimized feature extraction is the key to building good models. The features of the dataset are: 'Age', 'Sex', 'Chest pain type', 'BP', 'Cholesterol', 'FBS over 120', 'EKG results', 'Max HR', 'Exercise angina', 'ST depression', 'Slope of ST', 'Number of vessels fluro', 'Thallium'. Standardization was applied to the dataset to keep all the features on the same scale. It also helps to speed up the training process.

3.5 Label Encoding

Label encoding in Figure 3 is the process of translating labels into numeric format so that they may be read by machines. Machine learning algorithms can then better decide how those labels should be used. In supervised learning, it is a crucial pre-processing step for the structured dataset. Label encoding was performed to convert the categorical variables in the target column to numbers. The number 0 represents the false class (customers that did not churn). While the number 1 represents the true class (customers that churn).

3.6 Feature Scaling

The last step in the data cleaning process is feature scaling. The method of feature scaling used is standardization. Standardization is applied to the training set and test set to keep all the features on the same scale. It also helps to speed up the training process. The formula for standardization is shown in Equation 1.

$$x = \frac{[x - min(x)]}{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}} \tag{3.1}$$

3.7 Training Model

The dataset was used to train six machine learning algorithms. The algorithms include Logistic Regression, K Nearest Neighbours (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest and XGBoost.

3.7.1 Logistic Regression

For Logistic Regression, the dataset was split into the training and validation set. 75 $\%$ of the data was used for training and 25 $\%$ was used for validation. There are 2499 records in the training set and 834 records in the validation set. It predicts the output based on probability. It uses the sigmoid function. The output lies between 0 and 1. If the threshold is 0.5 then values from 0 to 0.49 are false, while values between 0.5 to 1.0 are true. It is a linear classifier. The Logistic Regression class was imported from the linear model module of the Sci-kit Learn library. The maximum number of training iterations was set to 300 while penalty was set to 'none'.

3.7.2 Training with K-Nearest Neighbour (KNN) Algorithm

The operating premise of the KNN algorithm is based on assigning a weight to each data point, which is referred to as a neighbour. KNN Classifier class was imported from the neighbors module of the Sci-kit Learn library. The parameter $n_neighbors$, which represents the number of neighbors, was set to 5. The metric parameter represents the metric to be used to calculate the distance from the k neighbors. The metric was set to 'minkowski', which is the default value. The p parameter is the power parameter. It is used only when the Minkowski metric is used. Power parameter for the Minkowski metric. When p = 1, this is equivalent to using Manhattan distance (l1), and euclidean distance (l2) for p = 2. Euclidean distance was used to calculate the distance between two points by setting p to a value of 2.

3.7.3 Support Vector Machine Algorithm

Support Vector algorithm is one of the most popular supervised learning algorithms. The goal of a support vector machine is to create the best decision boundary that can separate n-dimensional space into classes. The linear kernel of the support vector machine will be used to train the model. The SVC (Support Vector Classifier) class is imported from the SVM module of the Sci-kit learn library. The sigmoid SVM kernel is used. The other SVM kernels available in the Sci-kit Learn library are 'poly', 'rbf', 'linear' and 'precomputed.'

3.7.4 Decision Tree Algorithm

The Decision Tree is a tree-structured classifier. The internal nodes represent the features of the dataset, the branches represent the decision rules, and each leaf node represents the outcome. I import the Decision Tree Classifier class from the tree module of the Sci-kit Learn library. The criterion parameter is a function for determining the quality of a split. The criterion is "entropy." The min samples split represents the minimum amount of samples needed to separate an internal node in the decision tree. This parameter helps us avoid overfitting. The min samples split was 25 which means once 25 samples remain, they should not be split again into various classes.

3.7.5 Random Forest Algorithm

Random Forest is based on ensemble learning, which is a process of combining multiple classifiers to solve a complex problem. This algorithm uses multiple decision trees to make predictions. It takes less training time when compared to other algorithms. Random Forest Classifier class was imported from the ensemble module of the Sci-kit Learn library. The n estimators parameter represents the number of decision trees we use to build our random forest classifier. The n estimators parameter is 100. It means we are using 100 decision trees in our random forest classifier. The min samples split is 25, and the criterion parameter is ‘entropy.

3.7.6 XGBoost Algorithm

XGBoost which denotes “Extreme Gradient Boosting”. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements Machine Learning algorithms under the Gradient Boosting framework. It provides parallel tree boosting to solve many data science problems in a fast and accurate way. The XGBClassifier class was imported from the XGBoost library.

4 Results and Discussion

In this section, the results for different supervised machine learning models and the model’s performance using the same dataset by splitting the data into 75

4.1 Performance Metrics

To evaluate the performance of machine learning model, it is very important to utilize the correct metric. When an incorrect metric is used, it may cause the machine learning model to perform poorly when used in real life. Some standard evaluation metrics are:

4.1.1 Accuracy

This is used to deduce which machine model has the best capacity to recognize relationships and patterns between variables in the training dataset. It can be calculated by dividing the total number of correct predictions by the total number of predictions and then multiply by 100 (Juba and Le, 2019).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

where True Positive (TP) denotes that the actual value is true, and model predicts true. False Positive (FP) is also called as Type I Error; the actual value is false, and the model predicts true. True Negative (TN): The actual value is false, and model predicts false. False Negative (FN): This is also called as Type II Error, the actual value is true, and model predicts false.

4.1.2 Precision

Precision is the ratio of true positives and total positives predicted (Juba and Le, 2019). The formula for Precision is given as:

$$P = \frac{TP}{TP + FP} \tag{4.2}$$

4.1.3 Recall/Sensitivity/Hit-Rate

The recall metric focuses on type-II errors (FN). Although it cannot measure the existence of type-I error which is false positives (Juba and Le, 2019).

$$R = \frac{TP}{TP + FN} \tag{4.3}$$

4.1.4 F1-score

This is the combination of precision and recall. The F1- score is the harmonic mean of Precision and Recall (Juba and Le, 2019). The formula of the two essentially is:

$$R = \frac{Precision}{Precision + Recall} \tag{4.4}$$

4.2 Performance of Machine Learning Models

The results obtained after training machine learning models to customers that will churn in a telecommunication industry are given in the following subsections.

4.2.1 Logistic Regression

The class of interest is the ‘Churned’ class. The logistic regression model correctly classified 25 customers that churned. It misclassified 18 customers as churned, whereas the customers did not leave. It correctly classified 695 customers as retained. It misclassified 96 customers as retained, whereas they actually churned. The confusion matrix which shows the number of true positives, false positives, true negatives, and false negatives is shown in Figure 3. As shown in the summary of results in Figure 9 Logistic Regression had accuracy of 86 $\%$, precision of 84 $\%$, recall of 86 $\%$ and also F1-score of 83 $\%$.

4.2.2 K-Nearest Neighbour (KNN) Algorithm

The KNN model correctly classified 49 customers that churned. It misclassified 20 customers as churned, whereas the customers did not leave. It correctly classified 693 customers as retained. It misclassified 72 customers as retained, whereas they actually churned. The confusion matrix which shows the number of true positives, false positives, true negatives, and false negatives is shown in Figure 4 . A summary of performance analysis shown in Figure 9 shows that KNN had accuracy of 89 $\%$, precision of 88 $\%$, recall of 89 $\%$ and also F1-score of 88 $\%$.

4.2.3 Support Vector Machine

The SVM model correctly classified 10 customers that churned. It misclassified 54 customers as churned, whereas the customers did not leave. It correctly classified 659 customers as retained. It misclassified 111 customers as retained, whereas they actually churned. The confusion matrix which shows the number of true positives, false positives, true negatives, and false negatives is shown in Figure 5. A summary of performance analysis shown in Figure 9 shows that SVM had accuracy of 80 $\%$, precision of 75 $\%$, recall of 80 $\%$ and also F1-score of 77 $\%$

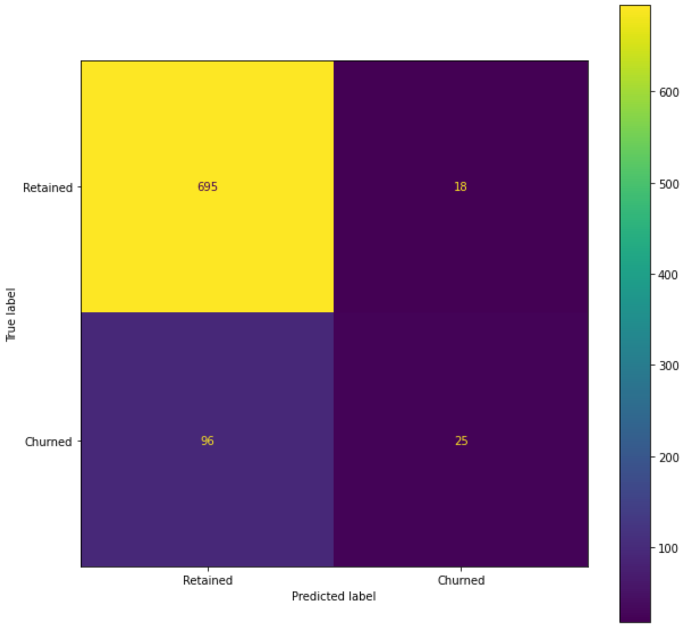

Figure 3: Confusion Matrix for Logistic Regression

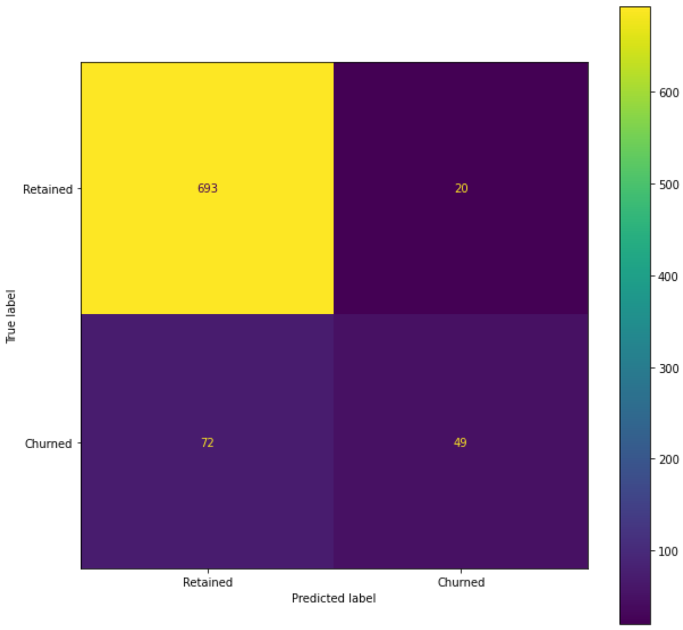

Figure 4: Confusion Matrix for K-Nearest Neighbour (KNN)

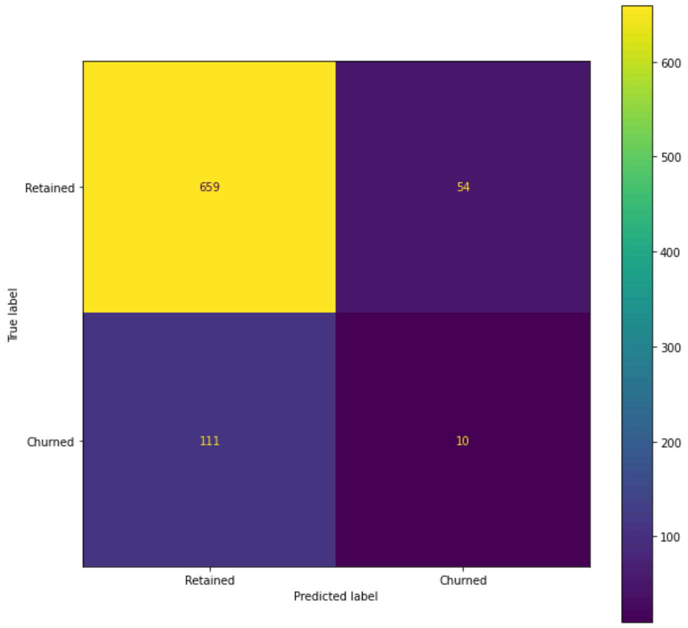

Figure 5: Confusion Matrix for Support Vector Machine

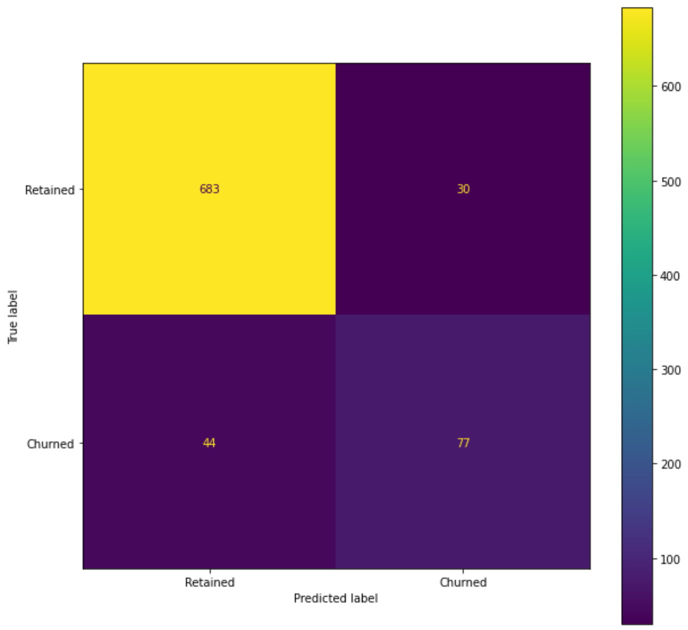

Figure 6: Confusion Matrix for Decision Tree Algorithm

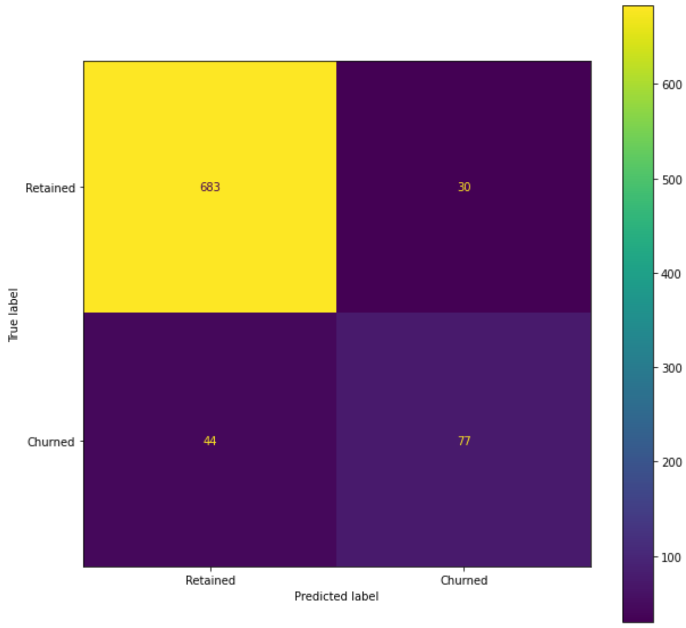

Figure 7: Confusion Matrix for Random Forest Algorithm

4.2.4 Decision Tree Algorithm

The Decision Tree model correctly classified 77 customers that churned. It misclassified 30 customers as churned, whereas the customers did not leave. It correctly classified 683 customers as retained. It misclassified 44 customers as retained, whereas they actually churned. The confusion matrix which shows the number of true positives, false positives, true negatives, and false negatives is shown in Figure 6. A summary of performance analysis shown in Figure 9 shows that Decision tree had accuracy of 91 %, precision of 90 %, recall of 91 % and also F1-score of 90 %

4.2.5 Random Forest Algorithm

The Random Forest model correctly classified 99 customers that churned. It misclassified 9 customers as churned, whereas the customers did not leave. It correctly classified 704 customers as retained. It misclassified 22 customers as retained, whereas they actually churned. The confusion matrix of Random Forest is shown in Figure 7. A summary of performance analysis shown in Figure 9 shows Random Forest had accuracy of 96%, precision of 96%, recall of 96 % and also F1-score of 96 %.

4.2.6 XGBoost Algorithm

The XGBoost model gave us the lowest number of false positives, and highest number of true negatives. While the random forest model gave us the lowest number of false negatives, and highest number of true positives. The confusion matrix of XGBoost model is shown in Figure 8. A summary of performance analysis shown in Figure 9 shows that XGBoost is the best model for predicting customers that have churn with accuracy rate of 96%.

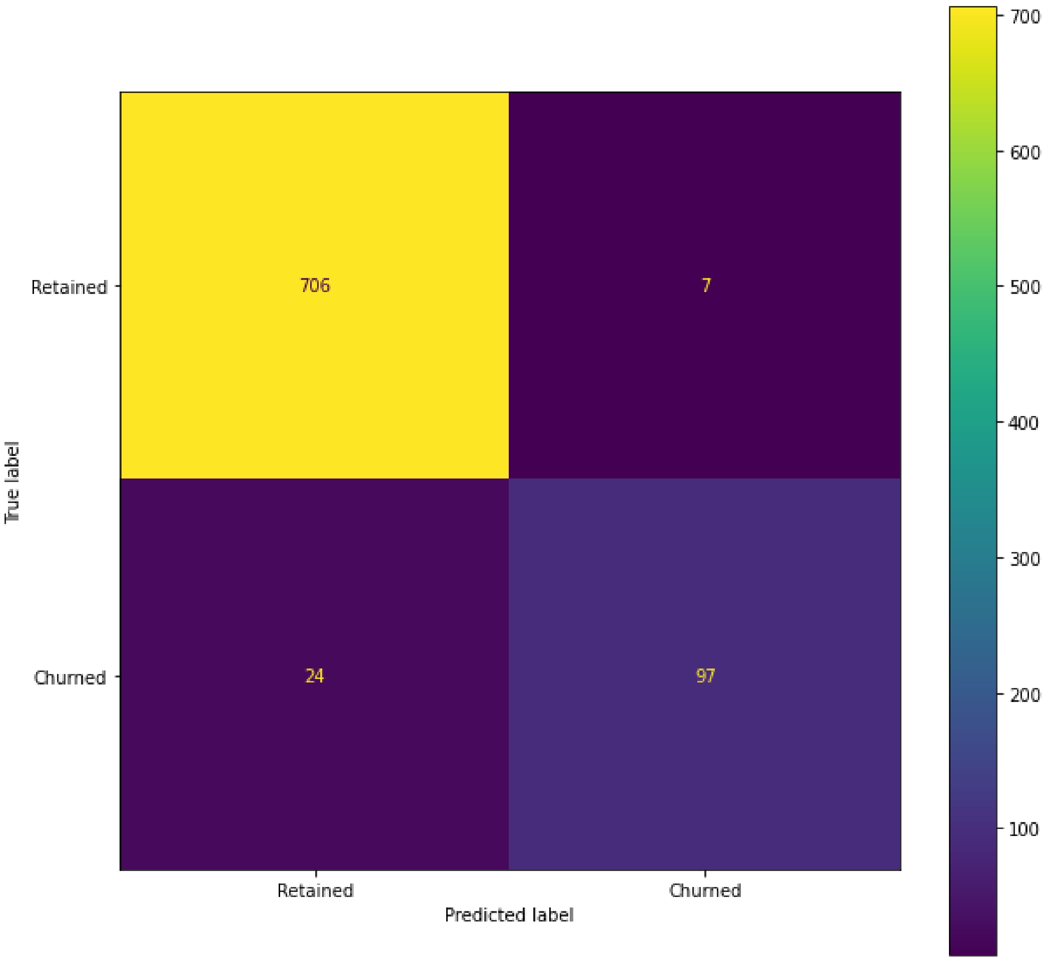

Figure 8: Confusion Matrix for XGBoost

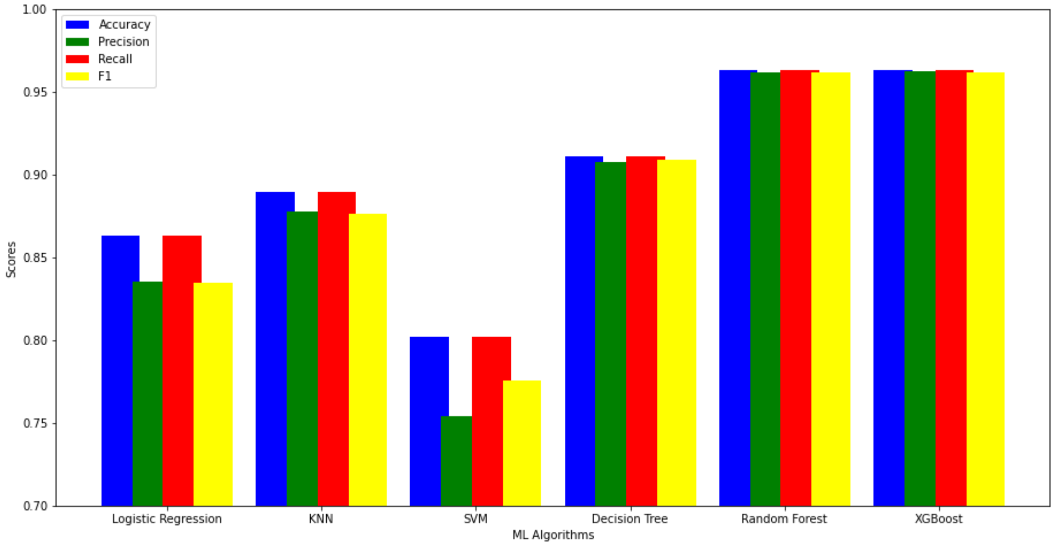

Figure 9: Performance Comparison of Machine Learning Models

4.2.7 Summary of Results

In machine learning, there are certain metrics used to determine if the machine models made progress or not. All machine models need a metric to evaluate their performances. In this research, certain performance metrics like precision, recall, f1-score and accuracy were used to determine the performance of each model. Precision which can also be referred to as positive predictive value, is the fraction of relevant instances among the retrieved instances. From the result obtained in this research, XGBoost outperformed other models with mean precision of 96 % which is the ability to identify only customers that are about to churn, recall (also known as sensitivity) of 96 % which is the fraction of relevant instances that were retreived by XGBoost and F-scores also known as F-measure of 96 % is a measure is a measure of a model's accuracy on a datatset. However, when compared to the works of the authors in [14], the results of their investigation reveal that logistic regression outperforms artificial neural networks and random forests. The logistic regression displays highest accuracy of 100 %, followed by 98.44 % accuracy for random forest classifier, and 85.55 % accuracy for ANN classifier. When comparing the result of this research to that of Mishra and Reddy [**?**], Ensemble based classifiers namely Bagging, Boosting and Random Forest for Churn Prediction in telecom industry. The Random Forest outperformed other approaches with 91.66 % of accuracy, error rate of 8.34 %, specificity rate of 53.54 %, and sensitivity rate of 98.89 %. In Raja and Pandian (2020), XGBoost classifier was compared to KNN and Random Forest Classifiers and it provided a superior accuracy score and F-1 score. From the result of Mohammad et al. (2019) and Mishra and Reddy (2017), it can be seen that different models were chosen based on their accuracy score except for Raja and Pandian (2020) in which XGBoost outperformed other model. This is due to the fact that for each model, there are different hyper-parameters that can be used. Each model needs to be fine-tuned to give the best accuracy. In this research, the XGBoost outperformed other models due to the hyper-parameter that was used to determine the accuracy of predicting customers that will churn.

4.2.8 Logistic Regression

5 CONCLUSIONS

Predicting customer churn in the telecommunication industry involves a complex interplay of techniques, data, and ethical considerations. While significant strides have been made in this field, continued research and innovation are necessary to address ongoing challenges and adapt to the rapidly changing landscape of the telecommunications industry. In this paper, a customer churn prediction model is proposed for data analytics. Six Machine learning algorithms including Logistic Regression, K-Nearest Neighbour, Support Vector Machine, Random Forest, Decision Tress and XGBoost were used in predicting customer churn in the telecommunication sector. Experimental results show that XGBoost outperformed the other algorithms followed by Random Forest and Decision Tree algorithms.

References

[1] Irina V. Pustokhina, Denis A. Pustokhin, Aswathy RH, T. Jayasankar, C. Jeyalakshmi, Vicente García Díaz, and K. Shankar. Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms. *Information Processing Management*, 58(6):102706, 2021.

[2] Michel Ballings, Dirk Van den Poel, and Emmanuel Verhagen. Improving customer churn prediction by data augmentation using pictorial stimulus-choice data. In Jorge Casillas, Francisco J. Martínez-López, and Juan Manuel Corchado Rodríguez, editors, *Management Intelligent Systems*, pages 217–226, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[3] Hemlata Jain, Ajay Khunteta, and Sumit Srivastava. Churn prediction in telecommunication using logistic regression and logit boost. *Procedia Computer Science*, 167:101–112, 2020. International Conference on Computational Intelligence and Data Science.

[4] Tonya Boone, John E. Boylan, Robert Fildes, Ram Ganeshan, and Nada Sanders. Perspectives on supply chain forecasting. *International Journal of Forecasting*, 35(1):121–127, 2019. Special Section: Supply Chain Forecasting.

[5] Elisabetta Raguseo. Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management*, 38(1):187–195, 2018.

[6] Anietie N. Okon, Steve E. Adewole, and Emmanuel M. Uguma. Artificial neural network model for reservoir petrophysical properties: porosity, permeability and water saturation prediction. *Modeling Earth Systems and Environment*, 7(4):2373–2390, Nov 2021.

[7] Chinedu Pascal Ezenkwu, Uduak Idio Akpan, and Bliss Utibe-Abasi Stephen. A class-specific metaheuristic technique for explainable relevant feature selection. *Machine Learning with Applications*, 6:100142, 2021.

[8] T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, and K.Ch. Chatzisavvas. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55:1–9, 2015.

[9] Sanaz Moayer and Scott Gardner. Integration of data mining within a strategic knowledge management framework. *International Journal of Advanced Computer Science and Applications*, 3(8), 2012.

[10] Ramakrishna Vadakattu, Bibek Panda, Swarnim Narayan, and Harshal Godhia. Enterprise subscription churn prediction. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 1317–1321, 2015.

[11] Ali Tamaddoni Jahromi, Mehrad Moeini, Issar Akbari, and Aram Akbarzadeh. A dual-step multi-algorithm approach for churn prediction in pre-paid telecommunications service providers, 2010.

[12] V. Yeshwanth, V. Vimal Raj, and M. Saravanan. Evolutionary churn prediction in mobile networks using hybrid learning. In *The Florida AI Research Society*, 2011.

[13] Kristof Coussement and Dirk Van den Poel. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1):313–327, 2008.

[14] Ibrahim Al-Shourbaji, Na Helian, Yi Sun, Samah Ibrahim Alshathri, and Mohamed Abd Abd Elaziz. Boosting ant colony optimization with reptile search algorithm for churn prediction. *Mathematics*, 2022.

[15] R. Sudharsan and E. N. Ganesh. A swish rnn based customer churn prediction for the telecom industry with a novel feature selection strategy. *Connection Science*, 34:1855 – 1876, 2022.

[16] Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh Chadha, and Pratyush Sethi. Customer churn prediction system: a machine learning approach. *Computing*, 104:271 – 294, 2021.

[17] Dana Al-Najjar, Nadia Al-Rousan, and Hazem M. Al-Najjar. Machine learning to develop credit card customer churn prediction. *J. Theor. Appl. Electron. Commer. Res.*, 17:1529–1542, 2022.

[18] Tianyuan Zhang, Sérgio Moro, and Ricardo F. Ramos. A data-driven approach to improve customer churn prediction based on telecom customer segmentation. *Future Internet*, 14:94, 2022.