INVESTIGATION OF CHURN OF CUSTOMER OF TELECOM INDUSTRY USING MACHINE LEARNING

by

Fathurrahman Yahyasatrio

Thesis submitted to University of Plymouth
in partial fulfilment of the requirements for the degree of

MSc Data Science and Business Analytics

University of Plymouth
Faculty of Science & Engineering

September 2022

Abstract

In the telecommunications industry, the possibility of a customer leaving a product or service, known as customer churn, has been one of the fundamental issues. Its negative impact on the business's overall performance leads to the importance of predicting customer churn accurately. In this work, we provide a predictive framework of customer churn by utilizing machine learning techniques. Following the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology which consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment, we aim to find the best model for classifying and predicting the churner and non-churner. Machine learning classifiers including the Decision Tree, Logistic Regression, Support Vector Machine, Naive Bayes, Neural Networks, AdaBoost, and Random Forest are compared using real-world data of 7,042 telecom users. Using the extracted demographic information, customer account, and subscribed services features of customers, the Random Forest model is found to achieve the best performance and outperform other models with 86.49% of F1-score and 86.29% accuracy performance.

Contents

List of Tables

List of Figures

List of Symbols

k = number of neigbours in the k-Nearest Neighbours algorithm, also refers to the number of folds in k-folds cross validation

X = input data

$P(A)$ = probability of A occurs

Acknowledgements

I have been extraordinarily fortunate to finish this thesis with the input, supervision, and support from numerous people.

First, it is a great pleasure to express my gratitude to Dr. Craig Mc Neile, my supervisor. Without his support, valuable insight, and dedicated supervision, this research would have never been completed.

Second, I also would like to sincerely thank my thesis examiners for their precious suggestions and constructive advice that help to improve the quality of this thesis.

Last but not least, I am forever indebted to my lovely parents for giving me unfailing support and endless energy in finishing this work. I dedicate this thesis to them.

1 Introduction

The telecommunication industry has been changing rapidly. Increase in the number of the telecommunications industry has led to a huge rise in competition where highly competitive environment has driven loss of valuable customers (Ahmed and Maheswari, 2017). The possibility of a customer leaving product or service is known as customer churn and furthermore, which may be attributed to dissatisfaction such as increased prices, poor quality, a lack of features, and privacy issues (Amin et al, 2018).

Customers in the telecommunications industry can select from a number of companies, and in a competitive market, they may simply switch from one service provider to another (Bhargava et al, 2022). Furthermore, they highlighted that in this fiercely competitive market, telecom businesses have an average yearly churn rate of 10-20%, which is significantly higher, and that it costs notably more to acquire a new client than to retain an existing one. According to Ahn et al. (2006), controlling customer turnover and minimizing customer churn are both critical.

According to Kotler (2002), the customer's satisfaction or dissatisfaction with the goods or services after purchase influences subsequent behaviour. A satisfied consumer is more inclined to buy the goods or services again. Dissatisfied customers, on the other hand, may abandon or return the product, refuse to purchase the product, or warn friends. Companies must comprehend the buyer's behaviour at each step and the forces at work. Others' attitudes, unforeseen situational considerations, and perceived risk may all influence purchasing decisions, as can customer levels of post-purchase satisfaction, company post-purchase activities, and consumer post-purchase usage and disposal of the goods.

Ahmed and Maheswari (2017) investigated that a customer does not often churn as a result of a single dissatisfaction experience. They stated that multiple dissatisfaction incidents are normally present before a consumer entirely discontinues doing business with a company, such as speed of download, stability of connection, billing, customer service, navigation and interface or content and services.

Customer churn behaviour has a negative impact on the company's overall performance; a potential reason for reduced sales since new/short-term customers purchase fewer services; helps competitors in obtaining unhappy customers through marketing strategy; leads to drop in revenue; has a detrimental effect on long-term customers; leading to dysfunction, reducing the ratio of potential new customers; acquiring new customers is more difficult (Amin et al. 2016).

The company have recognized the existing customers as the most valuable assets. More clearly, Kotler (2002) indicated that attracting a new customer (Acquisition customer) can cost five times as much as pleasing an existing customer (Customer retention). The increased cost of acquiring new consumers is related to the company's efforts in gathering and analysing data about customers in order to get useful information and insight into their preferences and behaviours. In order to force consumers to move, acquisition customers demand larger incentives. As a result, it makes more financial sense for the firm to focus on maintaining existing customers in order to avoid customer churn (Keramati and Ardabili, 2011).

The benefit of customer churn prediction is that it allows businesses to raise their yearly income, reduce client attrition, and identify and improve areas where their services are weak (Babu and Ananthanarayanan, 2014). Customer churn prediction is performed by analysing data to optimize decisions for the business and identify organizational issues that we will need to consider when making decisions behind gathering, storing and using data (Farnadi et al. 2016).

Singh and Samalia (2014) stated that business intelligence platforms may give business solutions to help organisations minimise customer attrition. In this regard, data mining and machine learning are regarded as important tools in business intelligence because they can use previous customer behaviour to predict their pattern, behaviour, and trends in order to better manage customers.

Machine learning is a subfield of artificial intelligence that aims to extract meaningful knowledge from complicated data structures computationally (Shouval et al, 2020) and become the modern statistical method to provide new opportunities to operationalize previously untapped growing sources of data which help to extract features and identify characteristics from data more accurately and with less human intervention (Sun and Scanlon, 2019). Moreover, they mentioned that big data and machine learning approaches have demonstrated significant potential for data-driven decision support, scientific discovery, and process improvement. The basic characteristic of machine learning is the ability to design systems capable of discovering patterns in data and learning from it without explicit programming. In the case of customer churn prediction, these are behavioural patterns that signal diminishing consumer satisfaction with firm services/products.

Machine learning offers a diverse set of algorithms, including multiclass decision forests, recommendation systems, neural network regression, multiclass neural networks, K-means clustering, and others. Each algorithm is tailored to a certain sort of machine learning task. Accuracy, training duration, linearity, number of parameters, and number of features are all requirements for machine learning situations.

The objectives of this research are to predict the customer churn using machine learning is to compare and provide a predictive framework of customer churn in the telecommunication industry by using various machine learning techniques for accurate prediction.

2 Review

In a highly competitive the telecommunication industry a defensive marketing strategy is becoming more important to face the issue of customer churn (Ahn, Han and Lee, 2006). Customer churn refers to customers who are on the verge of quitting the service (Babu and Ananthanarayanan, 2018).

Mutanen et al. (2006) stated that customer churn is the customer who are at risk of departing and, if feasible, whether those individuals are worth keeping. Migues et al. (2012) explained that customer churn is the customer who will leave the company. Almana et al (2014) has affirmed that customer churn is a popular method for tracking lost consumers which often lose valuable customer. Moreover, it has become critical for industry companies to safeguard their devoted client base and organisational productivity. To conclude, customer churn means the customers who are dissatisfaction and looking out for their best interest and then have an intention to leave existing company.

Oghojafor et al. (2012) investigated the discriminant factors affecting Telecom customer churn that forming several causes are: service plan that is unappealing, inadequate service facilities, excessive call charges, the availability of a better competing service provider, and off-beam advertising. Studies have shown that good service quality leads to reduce customer churn and acquiring of new customers, reduction in cost, enhance brand image, positive words-of-mouth recommendation and enhanced customer loyalty (Kotler, 2012). Lee and Murphy (2008) identified that service quality and value related more to loyalty. Rousan et al. (2010) in their study conclude that loyalty is greatly influenced by service quality.

There are two approach to mitigate customer churn. The first is an untargeted strategy that depends on superior products and mass marketing to strengthen brand alignment and customer retention, while the second, which can be reactive or proactive, is a targeted approach that relies on identifying customers who are likely to churn, providing them with a direct incentive, or customising service plans to keep them (Khan et al., 2010). Furthermore, they stated that if the churn projection is incorrect, the firms will be squandering bonus payment on customers who would have remained anyhow. As a result, the customer churn forecast procedure needs to be as precise as feasible. As a result, we're searching for a way to forecast which consumers are likely to depart.

Omar (2014) found that Multilayer Perceptron (MLP) method effectively identified the most influencing factors in customer churn by using the data collection comprises 11 characteristics of 5,000 randomly picked consumers during a three-month period. Sivansakar and Vijaya (2018) written that the prediction of customer churn is described using Neural Networks and probabilistic possibilistic fuzzy C-means clustering (PPFCM). The first set of experiments confirmed that the PPFCM segmentation can maintain and improve information partitioning results; the second set of experiments assessed the classification outcomes and associated it with other well-known classification methods; the final set of research associates it with other well-known classification methods; and the final set of research associates the proposed hybrid model framework with multiple other recently proposed hybrid organisations.

Saini et al. (2017) mentioned that Decision Tree using exhaustive Chi-square automatic interaction detection (CHAID) technique proven more effective and accurate than others in predicting individuals who are willing to churn in the near future. This dataset is a lengthy customer dataset of around 33,000 customers (active and terminated) that includes demographic as well as service statistics such as their area, call durations utilised at different times of day, prices incurred for services, and lifetime account duration.

Keramati and Ardabili (2011) stated that Binomial Logistic Regression delivering the most important value-added variables that influence customer turnover from 3,150 records and 11 features. Sebastian and Wagh (2017) confirmed that the Logistic Regression method helps to identify the probable customer churn and make the necessary business decisions. The dataset contains 22 variable available. These are related to gender, customer ID, phone service. Moreover, the dataset has over 2,000 customer related information avalaible.

Fei et al. (2017) discovered that combining K-Means with the Nave Bayes method improved accuracy and sensitivity in predicting customer churn in the telecom industry. This dataset contains 5,000 customer caller data points that are randomly divided into 1,667 instances as testing set to evaluate the performance of proposed model, while the remaining 3,333 examples are preserved for training and cross validation. According to Safitri and Muslim (2020), SMOTE Genetic Algorithms and Naive Bayes produce accuracy results of 73% on the implementation of the Naive Bayes method without the preprocessing phase. There are 13 nominal type characteristics and 7 numeric type features in this dataset. Purnamasari et al. (2020) found that cross selling marketing strategies with data mining methods using Naïve Bayes and C4.5 has a good accuracy value of 88.61%. The dataset contains 878 existing data and speedy telephone with 14 attributes with class label attributes (attribute output) which attribute package.

According to Chen et al. (2012), the most efficient model compared to the existing churn prediction methods was found to be Polynomial SVM-based churn prediction model was suggested for identifying whether or not a telecommunication client is a churner using 10-folds cross validation due to its prediction performance, training time, and number of support vector machines. Rodan et al. (2014) mentioned that SVM were optimised using a grid search with a specific evaluation criteria that may be altered depending on the cost of the retention effort approach. The dataset comprises 11 attributes of 5,000 randomly selected consumers with prepaid subscriptions during a three-month period.

3 Methodology/Procedure

In this chapter, we describe the methodology and sequence of work to investigate customer churn using machine learning techniques. We follow the methodology of Cross-Industry Standard Process for Data Mining (CRISP-DM).

CRISP-DM is an open and industry-proven common methodologies for extracting knowledge and actionable insights from data according to the business needs (IBM, 2021). Figure 1 depicts the process's six primary phases which includes Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The stages are not strictly ordered, and it is common to go back and forth between them. The arrows on the process diagram reflect the most essential and common linkages between phases.

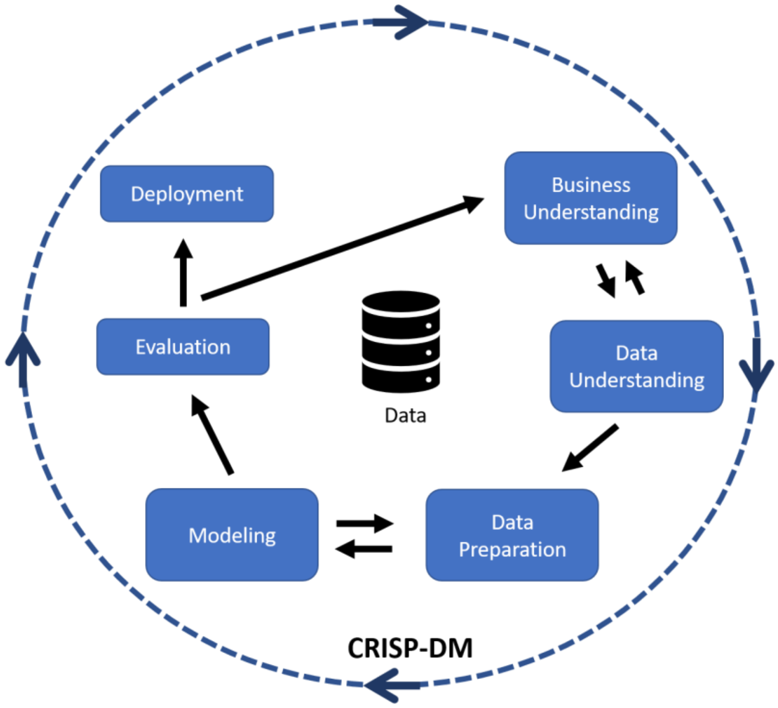


Figure 1. Life cycle of data mining according to the CRISP-DM (IBM, 2021)

The diagram's outermost ring shows the cyclical characteristics of data mining. The data mining process continues after a solution is implemented. Knowledge gained during the process may drive new, more focused commercial enquiries, and future data mining operations will benefit from previous organizations' experiences (IBM, 2021).

3.1 Business Understanding

In the first phase, we begin by focusing on comprehending the project's objectives and needs from a business standpoint. The project leader commonly formalises this business understanding requirement as a problem of knowledge extraction from data (data mining) and creates a preliminary strategy.

3.2 Data Understanding

Starting with preliminary data gathering, in this phase, the organisation moves on to efforts to familiarise itself with the data, detect data quality issues, and reveal early insights into the data. During this stage, the analyst may also identify relevant subgroups in order to build hypotheses regarding hidden knowledge.

The availability of actual and high-quality data is undoubtedly critical in determining the causes of customer churn. We use a real-world data set involves 7,043 customer ID from https://www.kaggle.com/radmirzosimov/telecom-users-dataset as publicly open data sets.

3.3 Data Preparation

The phase of data preparation includes all procedures required to create the ready-to-processed dataset from the raw input data. The most common procedures includes missing value handling, data transformation, data integration, and imbalance data checking and resampling. In missing data imputation, numerous techniques can be used, such as the deletion, averaging, most common values, k-nearest neighbours, etc. Figure 2 shows an example of how to impute the missing values of particular data using k-nearest neighbours if k=2.

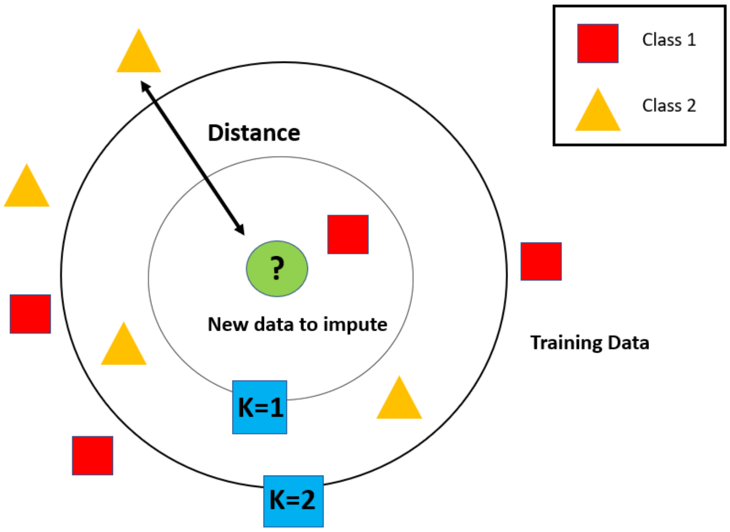


Figure 2. Example of missing value imputation using k-Nearest Neighbours (Navlani, 2018)

The issue of class imbalance is realistic and practical. Class unbalanced issues imply that we must learn from an unequal distribution of data samples between classes. It presents a difficulty to any classifier since it is difficult to learn the minority class samples. There are downsampling/undersampling methods to handle the problem by reducing the number of majority class data, while upsampling methods works by increasing the number of minority class data.

Imbalance ratio quantifies the fraction of data sample size in the majority group to the sample size in the minority group (Noorhalim & Shamsuddin, 2019). Imbalance ratio is widely used to assess the distribution of our data among multiple classes/labels.

3.3.1 Random Undersampling/Downsampling and Random Oversampling/Upsampling

Minority class means the under-represented class/label in the data distribution, while majority class means the class/label that has higher frequency in the data distribution. Oversampling/upsampling entails adding additional copies of a portion of the minority labels to the training data. The simplest technique of oversampling is random oversampling which can be performed more than once (twice, three times, etc.) This is one of the first suggested approaches, and it has also been demonstrated to be resilient. Rather of repeating every item in the minority label, part of them may be picked at random with substitution.

The undersampling/downsampling strategy excludes items from the majority label with or without substitution. The simplest undersampling technique, random undersampling, was one of the first strategies used to correct for class imbalances in a dataset; nonetheless, it may raise the variance of the classifier and is extremely likely to remove relevant or crucial records.

Aside from those simple random downsampling and random upsampling methods, the following more advanced techniques are available:

3.3.2 Synthetic Minority Oversampling Technique (SMOTE)

Synthetic oversampling approaches solve the class imbalanced issue by creating synthetic minority class samples to balance the distribution of majority and minority class samples. In a simple random oversampling approach, the minority class is reproduced from the minority labels of data. While it enhance the amount of data available, it provides no additional information or variety to the machine learning model. In their SMOTE work, Chawla et al. (2002) present a novel approach for creating synthetic data for oversampling purposes for the reasons stated above.

SMOTE generates synthetic data using a k-nearest neighbour method. SMOTE begins with selecting random data from the minority class, and then sets the data's k-nearest neighbours. The random data would then be combined with the randomly chosen k-nearest neighbour to create synthetic data.

3.3.3 Adaptive Sythetic (ADASYN)

Resampling-based techniques were unsatisfactory, prompting researchers to develop SMOTE, which was gradually enhanced by SMOTE variations, ADASYN, and others. ADASYN is a more general framework that finds the impurity of the neighbourhood for each minority observation by considering the ratio of majority observations in the neighbourhood and k (He, et al., 2008).

The core idea behind ADASYN is to employ a weighted distribution for distinct minority class examples based on their level of difficulty in learning, with more synthetic data created for minority class examples that are more difficult to learn than minority class examples that are simpler to learn (He, et al., 2008). As a consequence, the ADASYN technique enhances learning in two ways: (1) lowering the bias induced by the class imbalance, and (2) adaptively pushing the classification decision boundary toward the tougher cases. Simulation analyses on multiple machine learning data sources demonstrate the method's performance across five assessment measures.

3.3.4 Majority-Weighted Minority Oversampling Technique (MWMOTE)

MWMOTE is an improvement to the SMOTE approach that addresses some of the SMOTE technique's issues when there are noisy examples, in which situation SMOTE will commonly create another additional noisy examples (Barua, et al., 2014). In some cases, oversampling approaches may provide incorrect synthetic minority samples, making learning tasks more difficult. To that aim, a novel approach for dealing with unbalanced learning situations dubbed Majority Weighted Minority Oversampling TEchnique (MWMOTE) is introduced. MWMOTE detects the difficult-to-learn informative minority class samples and weights them based on their euclidean distance from the nearest majority class samples. It then uses a clustering technique to produce synthetic samples from the weighted informative minority class samples. This is accomplished such that all of the produced samples fall inside a minority class group.

3.4 Modeling

During this phase, the analyst assesses, chooses, and builds relevant machine learning modelling approaches. Because certain machine learning models, such as neural networks, have unique data format requirements, a loop back to data preparation phase is possible here.

Following is a brief overview of eight well-established and popular methods for churn prediction that take into account the model accuracy, efficiency, and prevalence in the scientific community (Kovac, 2020). The two most popular supervised machine learning tasks are classification and regression/prediction. If our target variable is categorical, it is considered a classification task. Otherwise, regression should be used if we have a numerical target variable.

3.4.1 Decision Tree

The decision tree is a tree-like decision structure that learns from known data labels for classification and prediction using an inductive approach (Han, 2012). The objective is to build a model that predicts the value of a target variable using basic decision rules derived from data attributes. A decision tree is a learning system that uses basic decision-making procedures to divide vast volumes of data into little data groups. Each successful segregation makes the individuals of the outcome group increasingly similar to one another. Decision tree with descriptive and predictive characteristics is simple to interpret, easy to incorporate into databases, and trustworthy.

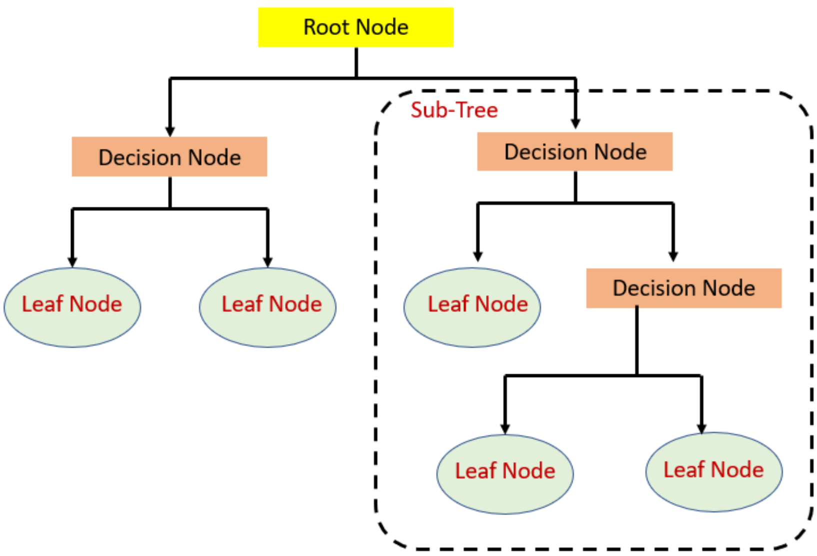


Figure 3. Decision Tree structure (Chouinard, 2022)

3.4.2 Logistic Regression

Logistic Regression is a technique for determining the connection between several independent (explanatory) and dependent (response) variables (Han, 2012). Although it has traditionally been employed in the medical industry. It is a sophisticated regression approach that has recently acquired prominence in social, engineering, science, etc. Because the Least Square Method (LSM) is insufficient in a multivariate model with dependent and independent variable discrimination, Logistic Regression is utilised as an alternative. The probability of the dependent variable with two values is predicted in Logistic Regression analysis. Furthermore, the model's variables are continuous. It is a strategy for categorising observations because of this property.

3.4.3 Support Vector Machine (SVM)

One of the supervised classification algorithms is the support vector machine. SVM is a machine learning technique that predicts and generalises new data by conducting learning on data with uncertain distributions (Han, 2012). SVM's fundamental premise is predicated on the availability of a hyperplane that best differentiates data from two classes. The Support Vector Machine is separated into two types based on the data set's linear and nonlinear separation.

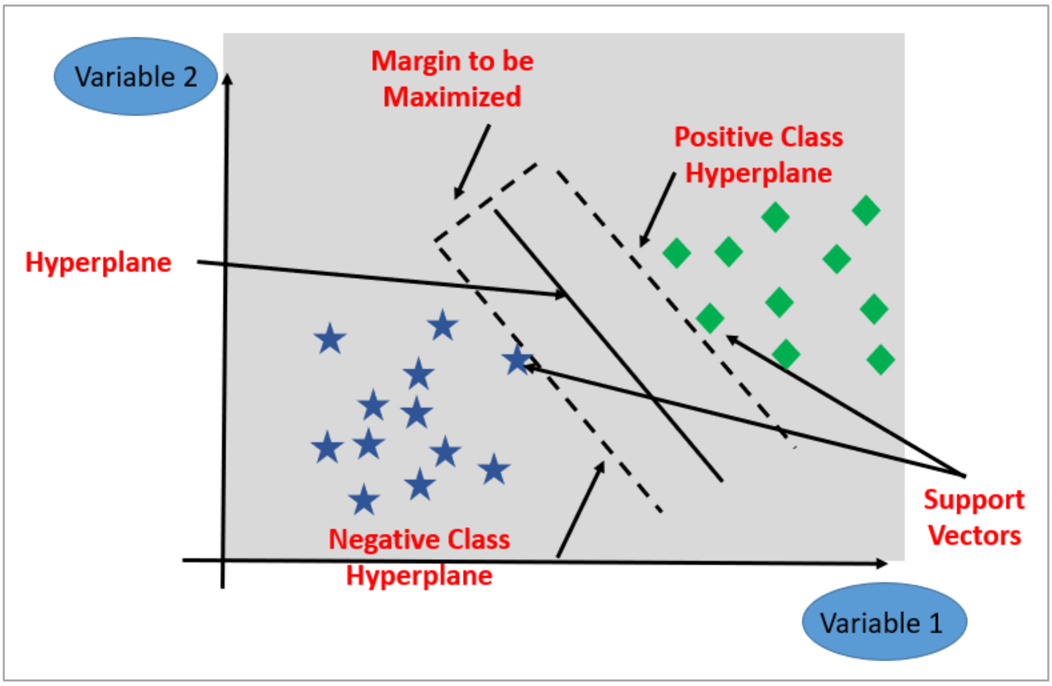


Figure 4. Illustrative example of SVM classifier (Dabakoglu, 2018)

3.4.4 Naive Bayes

The Naive Bayes classifier is a classification technique that labels data using statistical approaches (Han, 2012). It is commonly used in categorization difficulties because to its ease of usage. In principle, it is intended to determine the probability values of each criterion's influence in the Bayesian Classification. In order to estimate the likelihood of a class using data, Naive Bayes estimates the conditional probability of the class to which the data belongs. In this procedure, the Bayes theorem is applied.

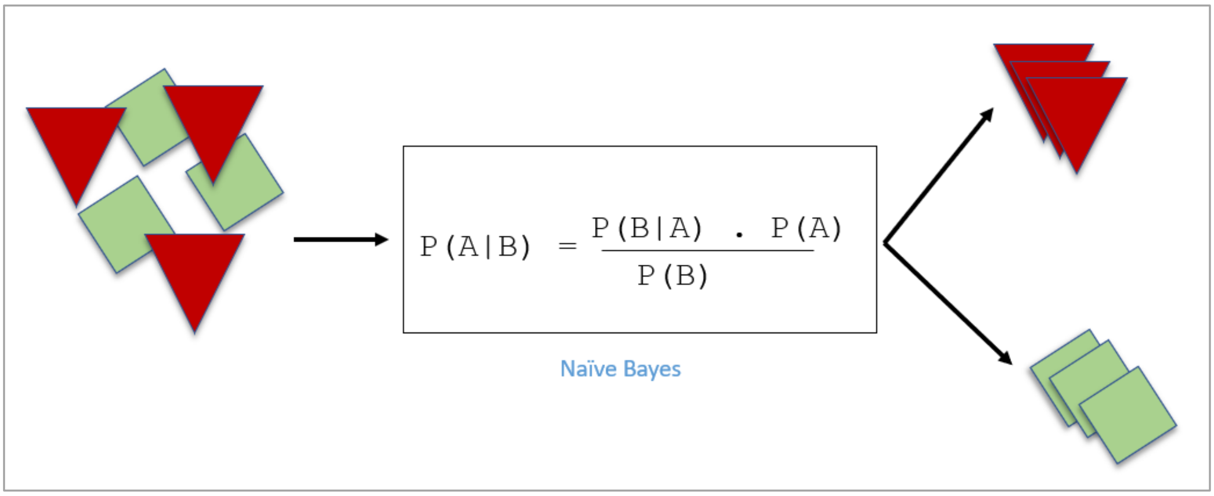


Figure 5. Illustration of Naïve Bayes classifier (Gusain, 2020)

3.4.5 Neural Networks (NN)

Neural networks has been considered as common solution for dealing with difficult problems such as churn prediction. Neural networks can be hardware-based (neurons are represented by real components) or software-based (computer model), and they can employ a wide range of topologies and learning techniques. The Multi-Layer Perception (MLP) model consists of the input layer, a number of hidden layers and an output layer, which is trained using feed-forward and back-propagation mechanism.

In this work, we construct a multilayer perceptron with 3 hidden layers and logistic activation function. Each hidden layer consists of 10 neurons. The k-nearest neighbours technique is used to impute missing values.

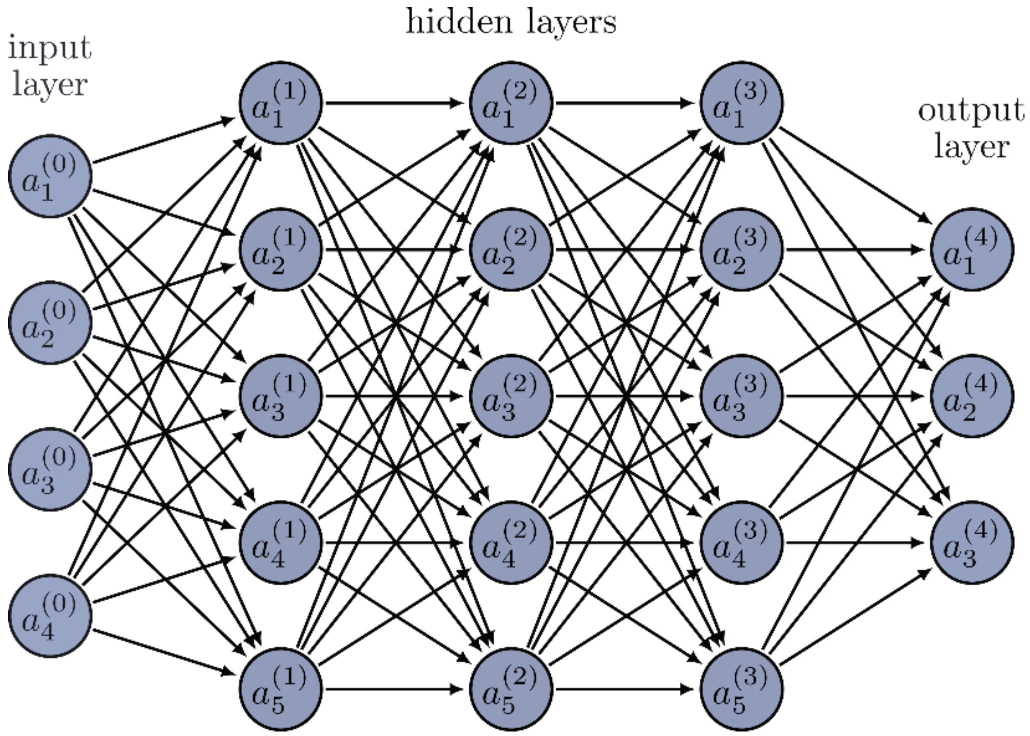


Figure 6. Example of neural network architecture with three hidden layers (Neutelings, 2021)

3.4.6 Adaptive Boosting (AdaBoost)

One of the improvement to the machine learning algorithms is how we can construct a combination of several models and build an ensemble model. An ensemble classifier is commonly more accurate than its individual classifiers. Consider the case of a group that votes by majority. That is, given a data X to classify, it gathers the target class predictions produced by the basis classifiers and outputs the class with the highest probability. Although the basic classifiers may make errors, the ensemble will mislabel X only if more than half of the classification models are incorrect. When the models are diverse, the ensemble produces superior results. That is, there should be low connection among classifiers. The models should also outperform random guessing. Because each base classifier may be assigned to a separate CPU, ensemble techniques can be parallelized (Han, et al. 2012).

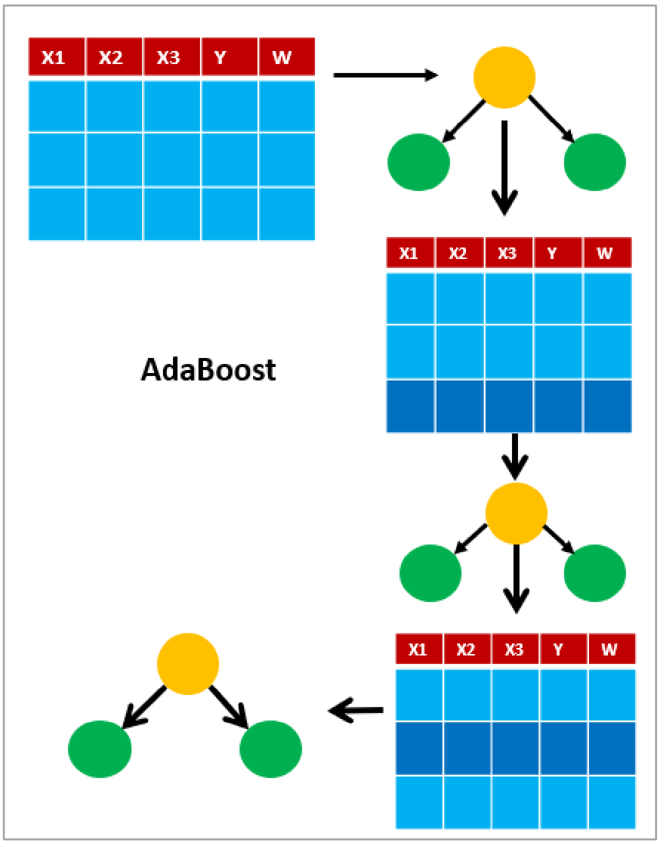


Figure 7. Schematic illustration of AdaBoost method built from decision tree (Prastiwi & Ulubalang, 2020)

AdaBoost, as shown in Figure 7, is an ensemble learning technique that constructs a strong classifier by iteratively selecting and combining a new model, such as a decision tree, which reduce the error of the previously selected model in a sequential manner.

3.4.7 Random Forest (RF)

Random Forest has been considered a versatile and strong supervised machine learning technique that builds and merges many decision trees to form a forest-like structure (Han, 2012). Figure 8 illustrates the schematic example of random forest model built from multiple decision trees. It is applicable to both classification and regression tasks.

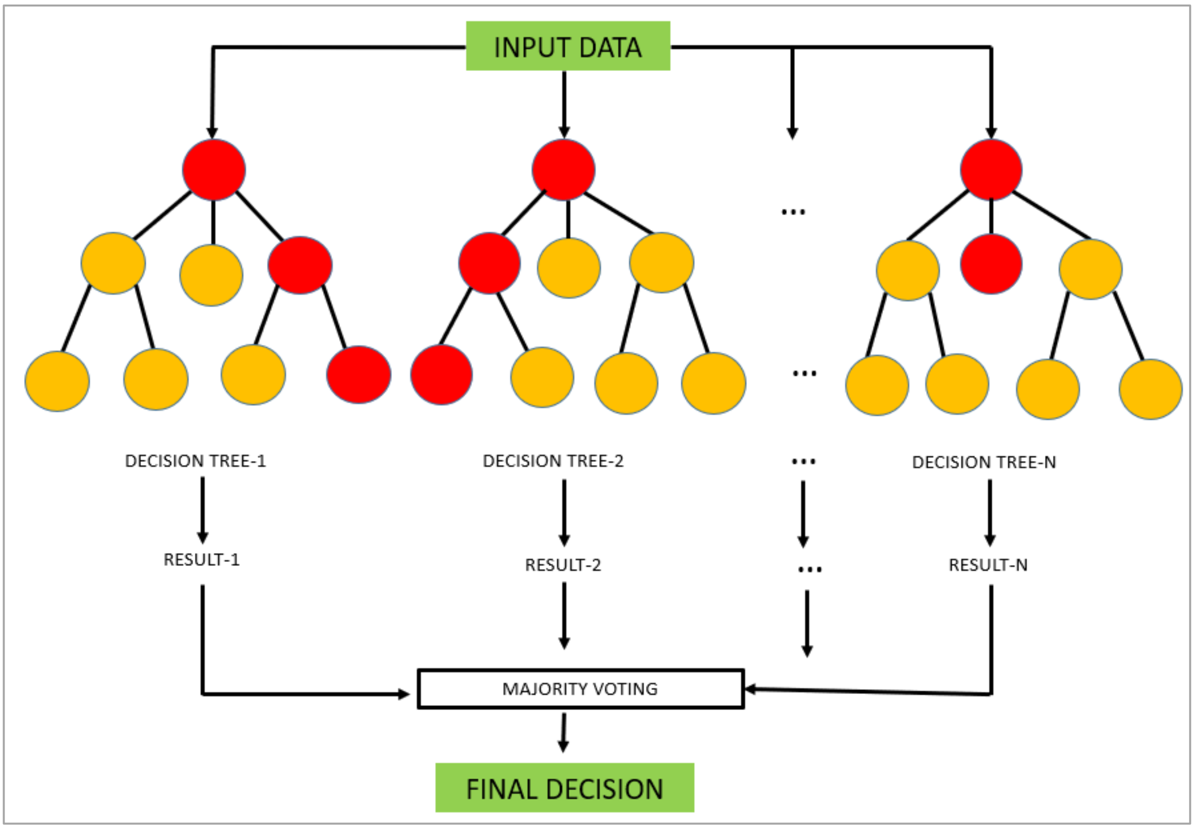


Figure 8. Random Forest method (Sruthi, 2021)

3.5 Evaluation

During this phase, the data scientist assesses, chooses, and builds relevant machine learning modelling approaches. Because certain machine learning models, such as neural networks, have unique data format requirements, a loop back to data preparation phase is possible here.

Evaluating the performance of a churning prediction model is a critical step in ensuring that it can be appropriately generalised (Huang et al., 2012). In fact, evaluation criteria assess a prediction model's ability to accurately rank consumers based on the likelihood of churning. There are well-known metrics for measuring the performance or effectiveness of a predictor model, such as accuracy, precision, recall, and the F-measure, which may successfully demonstrate prediction performance. We utilised these criteria, which are generated based on the confusion matrix provided below, to evaluate the efficacy of machine learning approaches in predicting churn status.

Table 1. Confusion Matrix

		Predicted Class: Churner	Predicted Class: Non-churner
Actual Class	Churner	True Positive	False Positive
	Non-churner	False Negative	True Negative

According to the confusion matrix shown in Table 1, to evaluate machine learning models sufficiently, four terminology should be defined.:

1. TP (True Positive) specifies the number of consumers who should be in the churner group and the prediction model accurately identified them as such.
2. TN (True Negative) shows how many individuals who are in the non-churner group and the prediction model accurately identified them as such.
3. FP (False Positive): the number of customers who are not churners but were wrongly classified as such by the model.
4. FN (False Negative): the number of customers who churn but are wrongly classified as non-churners by the model.

Accuracy is the number of all the correct predictions, and it is calculated using the following equation:

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

Recall is the ratio of real churners which are correctly identified.

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

Precision is the ratio of predicted churners which are correct, and it is calculated using the following formula.

$$Precision = \frac{TP}{(TP + FP)} \qquad (3)$$

F1-score or F-measure is the harmonic average of precision and recall, and it is calculated using equation (4):

$$F1Score = \frac{2\ x\ Precision\ x\ Recall}{(Precision + Recall)} \qquad (4)$$

Sensitivity, also known as the true positive rate (TPR), measures the probability of a churner is an actual churner. This metric is the same as recall.

$$Sensitivity = \frac{TP}{(TP + FN)} \qquad (5)$$

Specificity, also called the true negative rate (TNR), assesses the probability of a non-churner is a truly non-churner.

$$Specificity = \frac{TN}{(TN + FP)} \qquad (6)$$

The positive predictive value (PPV) and negative predictive value (NPV) quantify the ratios of churners and non-churners that are actual churners and non-churners respectively.

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{(TP + FP)} \qquad (7)$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{(TN + FN)} \qquad (8)$$

Balanced accuracy is useful to measure both the sensitivity and specificity criteria simultaneously in the form of an arithmetic average.

$$Balanced\ Accuracy = \frac{(Sensitivity + Specificity)}{2} \qquad (9)$$

Kappa statistic, also known as Cohen's kappa coefficient, assesses the degree of consensus among multiple investigators evaluating the same occurrence, hence consider all cells in the confusion matrix.

$$Kappa\ statistic = \frac{2\ x\ (TP\ x\ TN - FN\ x\ FP)}{(TP + FP)\ x\ (FP + TN) +\ (TP + FN)\ x\ (FN + TN)} \qquad (10)$$

In practice, the data will be split into two parts: training set and testing/validation set. The training set will be used by machine learning model to learn the informative pattern from data. On the other hand, the testing set will be used to evaluate the performance of the trained model. While the random split may not thoroughly evaluate the model on the whole dataset, the k-folds cross validation technique is preferred. Figure 9 illustrates how to split the data and perform model evaluation where each parts of the input dataset has been used either as training and testing/validation set.

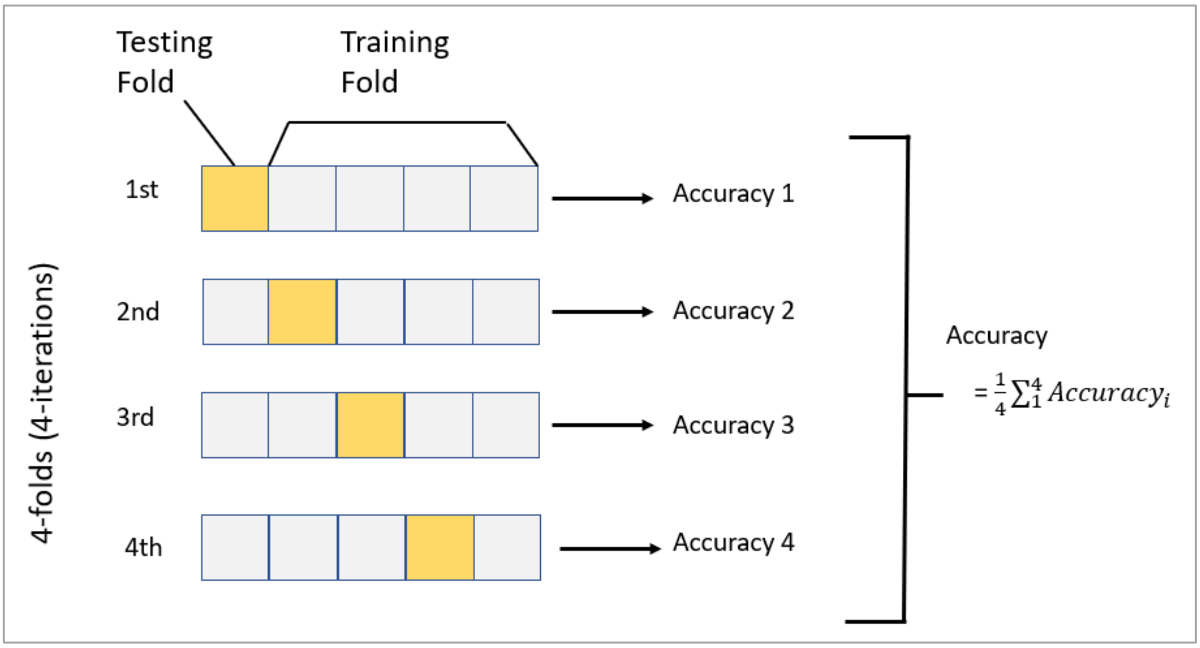


Figure 9. Schematic illustration of k-folds cross validation where k=4 (Ansari, 2021)

3.6 Deployment

In most cases, this phase will include integrating a code version of the model into a computer system. It also includes techniques for scoring or categorising previously unknown data as it emerges. The procedure should apply the new knowledge to the existing business issue. Notably, the code description must incorporate all data preparation procedures prior to modelling. This assures that the model will continue to treat fresh raw data in the same way that it did throughout model construction.

4 Results

4.1 Business Understanding

Numerous customers in the telecoms industry are given the right and ability to choose their preferred provider ffrom a range of telecom companies, and in a competitive market, they may simply switch from one service provider to another. The advantage of customer churn prediction assists businesses in increasing yearly revenue, reducing customer loss, and identifying and improving areas where their offerings are weak.

Our telecommunication customer churn data includes information on an unnamed telecommunications firm that supplied home phone and Internet services to 7,043 customers. It shows which clients have left, stayed, or joined their service. Each client has many vital demographic attributes such as gender, phone service, etc.

4.2 Data Understanding

4.2.1 Data Summary

Data includes the multiple attributes which can be grouped into three main categories of information. The first is demographic information which includes gender, senior citizen, partner, and dependents. The second category of available information is the services subscribed: phone service, multiple line, internet service, online security, online backup, device protection, tech support, streaming TV, and streaming movies. The last information about customer account includes the customer ID, contract, paperless billing, payment method, monthly charges, total charges, and tenure.

Table 2. Summary of Data Description

Variables	Data Types	Number of Missing Data	Mean	Standard Deviation
CustomerID	Categorical	0	-	-
Gender	Categorical	0	-	-
SeniorCitizen	Numerical	0	0.16	0.37
Partner	Categorical	0	-	-
Dependents	Categorical	0	-	-
Tenure	Numerical	0	32.37	24.56
PhoneService	Categorical	0	-	-
MultipleLines	Categorical	0	-	-
InternetService	Categorical	0	-	-
OnlineSecurity	Categorical	0	-	-
OnlineBackup	Categorical	0	-	-
DeviceProtection	Categorical	0	-	-
TechSupport	Categorical	0	-	-
StreamingTV	Categorical	0	-	-
StreamingMovies	Categorical	0	-	-
Contract	Categorical	0	-	-
PaperlessBilling	Categorical	0	-	-
PaymentMethod	Categorical	0	-	-
MonthlyCharges	Numerical	0	64.76	30.09
TotalCharges	Numerical	11	2283.3	2266.77
Churn	Label	0	-	-

Customer ID attributes is considered as not useful for classification as it only store the auto-increment unique identity of customers. Further, the distribution of churner and non-churner labels is investigated from the churn label as shown in Table 3 and Figure 10. The dataset is imbalanced where the majority of customers are non-churners.

Table 3. Number of customer status based on the churn status

Churn Status	Frequency	Percentage
No Churn	5,174	73.46
Churn	1,869	26.54
Total	**7,043**	**100.00**

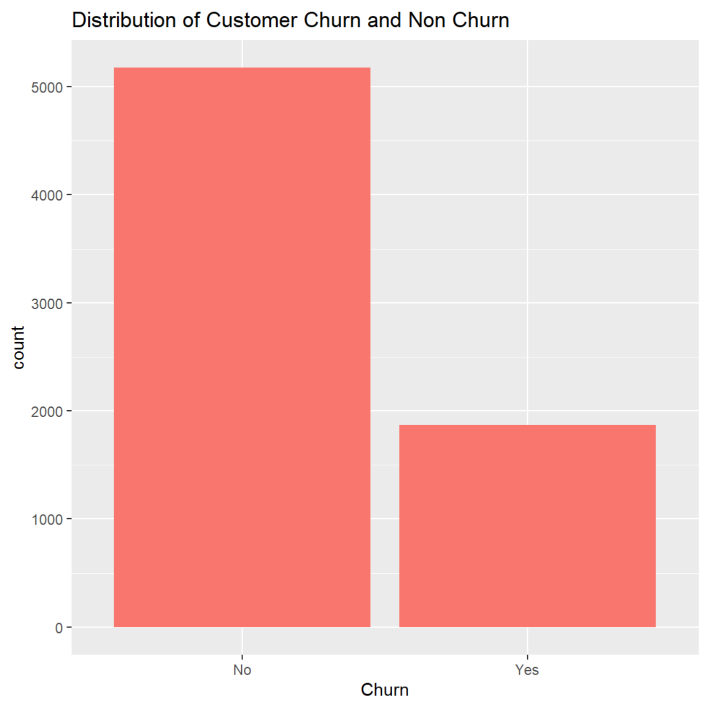


Figure 10. The distribution of churner and non-churner labels

4.2.2 Correlation among Features

In order to gain deeper understanding of the data and to design the most suitable and well-performed classification model, we strive to discover related characteristics from a set of data. To do so, we need to discard irrelevant features that do not contribute much to our target variable, in this case is the churn status.

Fewer features are usually more preferable since they minimise model complexity, and a simpler model is easier to comprehend and interpret. Correlation is a statistical terminology that describes how close two variables are to having a linear relationship with one another. Pearson's correlation coefficient quantifies the strength of a linear relationship between two variables. A Pearson correlation coefficient ranging from -1 to 1 that reflects how closely two variables are connected linearly.

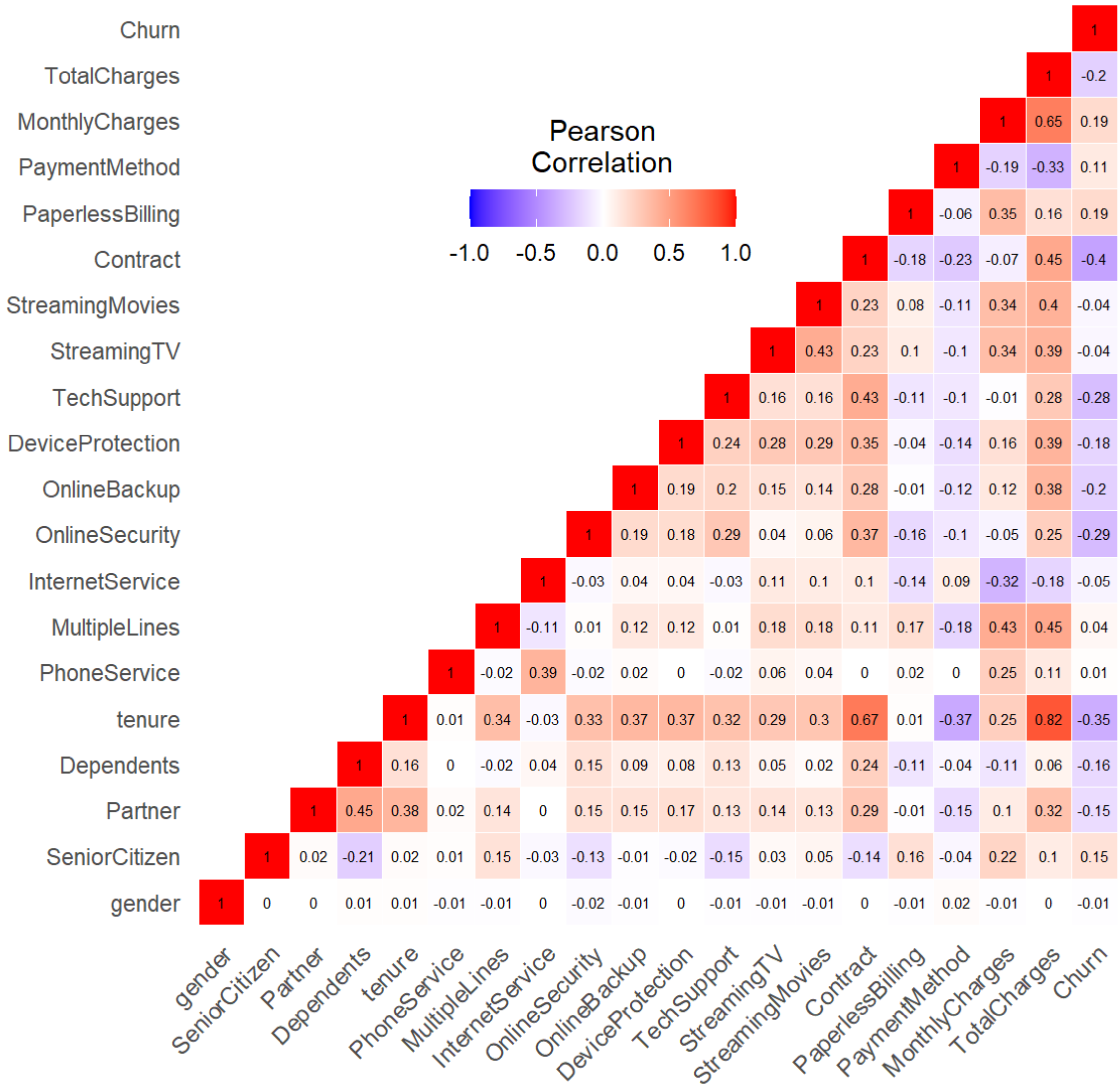


Figure 11. Correlation among Features

Figure 11 shows the Pearson's correlation among features where some notable correlations can be highlighted, such as the tenure status is highly correlated with the total charges and contract.

4.3 Data Preparation

4.3.1 Imputation of Missing Values

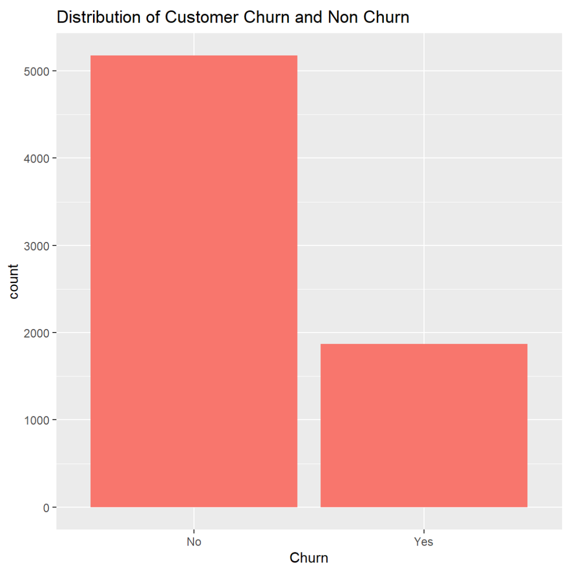
There are missing values in TotalCharges. We perform missing values imputation using k-Nearest Neighbors algorithm, hence all variables are complete.

4.3.2 Resampling Technique for Handling Imbalanced Data

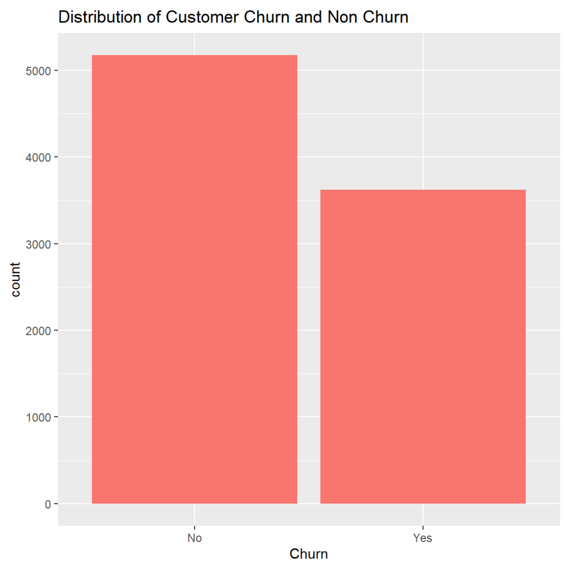
Working with unbalanced datasets presents the difficulty that most machine learning algorithms will overlook, and so perform poorly on, the minority class, despite the fact that performance on the minority class is often the most essential. To deal with the imbalance dataset, we compared multiple techniques include the SMOTE, DATASYN, and MWMOTE.

Imbalance ratio quantifies the fraction of data sample size in the majority group to the sample size in the minority group (Noorhalim & Shamsuddin, 2019). In our case, the majority and minority classes are non-churners and churners respectively. As shown in Figure 12, the original dataset has an imbalance ratio of 0.3612, which is very low. First, we investigate the use of Synthetic Minority Over-sampling Technique (SMOTE) on the data and gained an improved imbalance ratio into 0.7000 as shown in Figure 12. The Adaptive Synthetic (ADASYN) technique gained better results with imbalance ratio of 1.0114. However, the resulted class distribution is unnatural as now the churn label ("Yes") becoming the majority contradict to the original data.

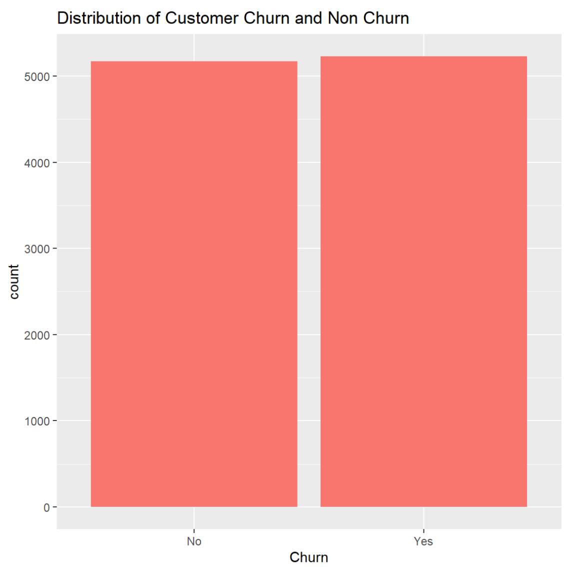
We investigate the method of Majority-Weighted Minority Over-sampling Technique (MWMOTE) to the data and obtained an improved imbalance ratio into 0.9501. With better ratio than other resampling methods and more natural class distribution to the original data, MWMOTE is then selected to modify the imbalanced churn dataset.



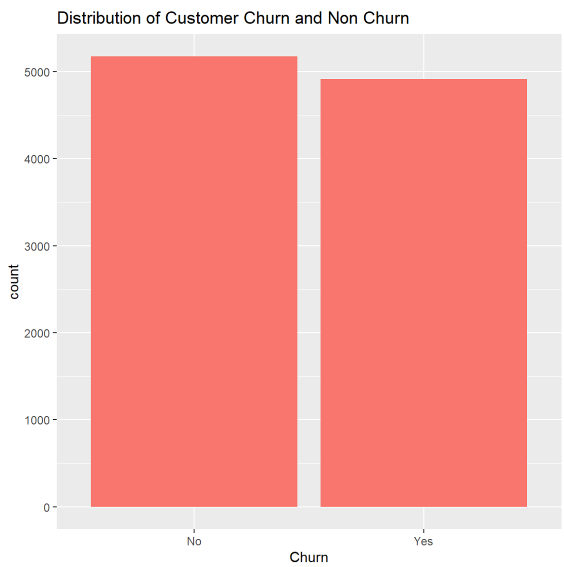
(a) Original,
imbalance ratio of 0.3612



(b) SMOTE,
imbalance ratio of 0.7000



(c) ADASYN,
imbalance ratio of 1.0114



(d) MWMOTE,
imbalance ratio of 0.9501

Figure 12. Comparison of Resampling Techniques for Class Imbalanced

4.4 Modeling

In this phase, we build eight machine learning classification models to distinguish the churners and non-churners. The models include the decision tree, logistic regression, support vector machine, Naïve Bayes, neural networks, AdaBoost, and random forest. To optimize the performance of each model, model parameter tuning is conducted to select the best parameter of each classifier.

4.5 Evaluation

Evaluating the performance of a churning prediction model is an important step in ensuring that it can be generalised effectively. The simulation using 10-folds cross validation under the same environment settings for all models is conducted to ensure a fair evaluation. Table 4 shows the performance comparison of eight machine learning models that we have been built. In terms of accuracy, F1-score, precision, and recall, Random Forest model achieves the best performance.

Table 4. Performance Comparison of Machine Learning Models

Evaluation Metric	Decision Tree	Logistic Regre-ssion	Support Vector Machine	Naive Bayes	Neural Networks	Ada Boost	Random Forest
Accuracy	0.7639	0.8133	0.8178	0.7672	0.8030	0.8560	**0.8629**
F1-Score	0.7621	0.8087	0.8204	0.7646	0.8036	0.7860	**0.8649**
Precision	0.7874	0.8486	0.8314	0.7975	0.8234	0.8050	0.8769
Recall	0.7552	0.7687	0.8221	0.7397	0.7860	0.8110	0.8603
Sensitivity	0.7552	0.7687	0.8221	0.7397	0.7860	0.8000	0.8603

Evaluation Metric	Decision Tree	Logistic Regre-ssion	Support Vector Machine	Naive Bayes	Neural Networks	Ada Boost	Random Forest
Specificity	0.7962	0.8564	0.8285	0.8084	0.8208	0.8430	0.8739
Kappa Statistics	0.5282	0.6272	0.6356	0.5350	0.6061	0.6960	0.7258
Balanced Accuracy	0.7645	0.8143	0.8180	0.7680	0.8034	0.7870	0.8631
Positive Predictive Value	0.7874	0.8486	0.8314	0.7975	0.8234	0.8350	0.8769
Negative Predictive Value	0.7465	0.7793	0.8110	0.7433	0.7853	0.8010	0.8544

From Table 4, we can conclude that Random Forest algorithm obtaines the highest accuracy and F1-score. Compared with the literature, two other existing studies on customer churn prediction of telecom industry are discussed by Ullah, et al. (2019) and Lalwani, et al. (2021).

Lalwani, et al. (2021) compares multiple machine learning model on the same dataset of ours and found that Adaboost and XGboost Classifier gives the highest accuracy of 81.71% and 80.8%. Compared to our results, their study differs in three main processes: handling missing values data, handling class imbalanced data, and k-fold cross validation. First, they remove all records containing missing values while we take benefit of the records and impute the missing values using k-nearest neighbours algorithm. Second, Lalwani, et al. (2021) ignores the class imbalanced data distribution while we perform resampling using the Majority-Weighted Minority

Oversampling Technique (MWMOTE) algorithm. Third, they use 5-folds cross validation in the evaluation, while we take deeper evaluation using 10-folds cross validation. In results, our finding suggests that Random Forest achieves best accuracy of 86.29%.

Ullah, et al. (2019) uses different datasets to classify churn customers data using machine learning algorithms and found that Random Forest algorithm performed better than other machine learning classifiers and obtained 88.63% accuracy. This finding in different datasets is inline with our experiments.

5 Discussion

5.1 Model Interpretation

Machine learning models offer great potential to accurately predict or classify unlabelled data. In predictive modeling, not only the performance of the model that is important, the interpretability is undoubtedly necessary. We used the variable importance to interpret the prediction results of the best model (Random Forest) as shown in Figure 13. The larger value of variable importance the more important the variable is. Contract, monthly charges, and total charges are top 3 important variables in predicting customer churn.

We also used SHAP (Shapley Additive exPlanations) technique to interpret the individual prediction output of any machine learning models. SHAP is based on Shapley values, which determine the significance of a variable by evaluating which label/class would a proposed model resulted in with and without the variable (Shapley, 1953). However, because the order in which a model perceives variables might impact its results, this is accomplished in every feasible sequence to ensure that the characteristics are assessed equally.

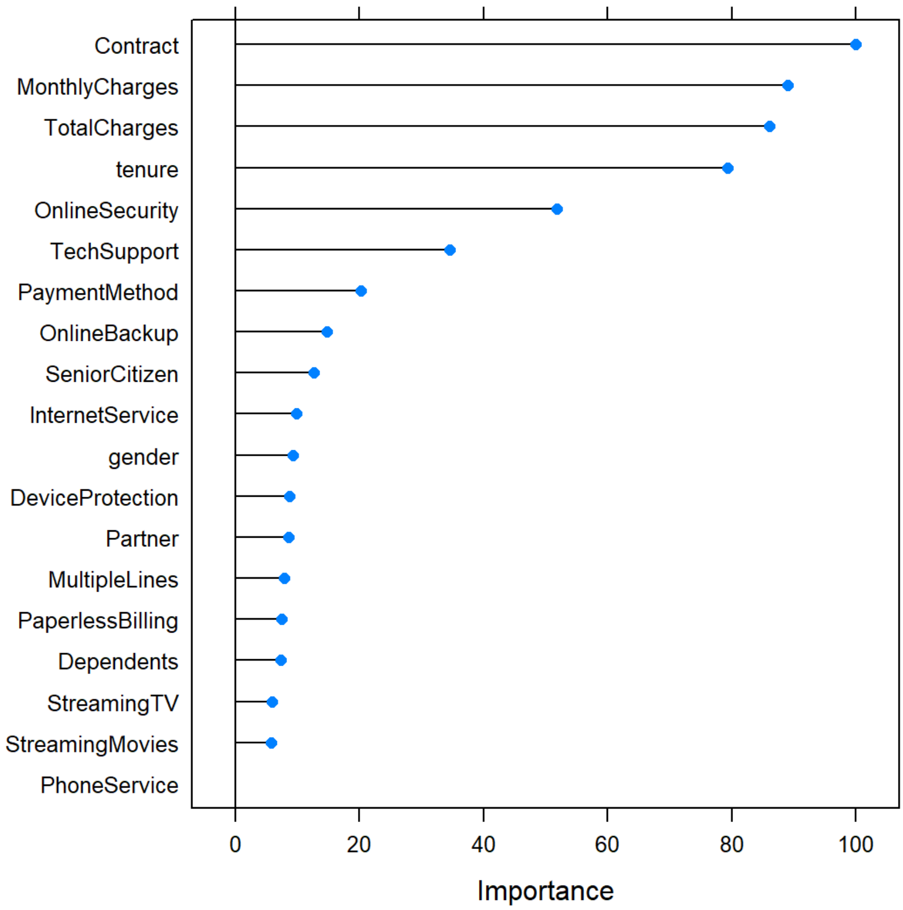


Figure 13. Most important features from Random Forest model

For instance, in linear models, we can utilise their parameters to calculate the overall importance of each variable, however they are scaled with the variable's scale, which can bring inaccuracies and misunderstandings. Furthermore, the coefficient does not account for the feature's local relevance or how it varies with lower or higher levels. The same can be true for variable importances in tree-based models, which is why SHAP is important for model interpretation. However, while SHAP illustrates the influence or importance of each variable on the model's guess, it does not assess the forecast's quality.

In this case, the best model, Random Forest is investigated as shown in Figure 14. For instance, the SHAP results of the 1st customer shows that for this customer, the status of contract, tech support, and phone service subscription leads to its churn status.

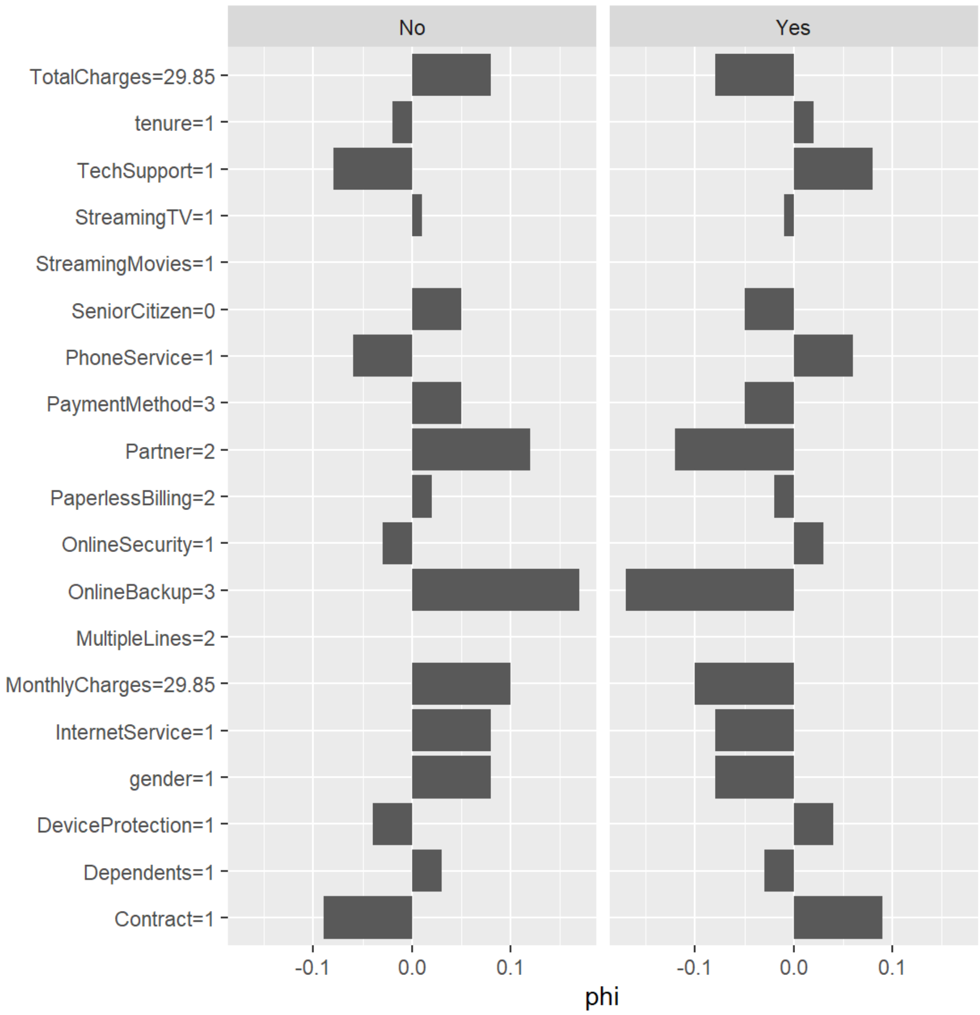


Figure 14. Individual SHAP value for Customer #1

For the 1st customer, we can explain the value of each features/variables that leads to its prediction by Random Forest model. We can interpret the prediction of other customer in a similar way.

6 Conclusion

In this thesis, we aim to find the best model for classifying and predicting the churner and non-churner in the telecommunication business. Machine learning classification models including the Decision Tree, Logistic Regression, Support Vector Machine, Naive Bayes, Neural Networks, AdaBoost, and Random Forest are compared using real-world data of 7,042 telecom users. Using the extracted demographic information, customer account, and subscribed services features of customers, the Random Forest model is found to achieve the best performance and outperform other models with 86.49% of F1-score and 86.29% accuracy performance. The performance of our proposal was better than the classifier designed Lalwani, et al. (2021) on the same data set.

List of References

Adwan, O., Faris H., Jaradat, K., Harfoushi, O., & Ghatasheh, N. (2014). Predicting customer churn in telecom industry using multilayer perceptron neural networks: Modeling and analysis. *Life Science Journal*, *11*(3), 75-81. https://doi.org/10.7537/marslsj110314.11

Ahmed, A. A. Q., & Maheswari, D. (2017). Churn prediction on huge telecom data using hybrid firefly based classification churn prediction on huge telecom data. *Egyptian Infromatics Journal*, *18*(3), 215-220. https://doi.org/10.1016/j.eij.2017.02.002

Ahn, Jae-Hyeon.,Han, Sang-Pil., Lee, Yung-Seop. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korea Mobile telecommunication service industry. *Telecommunication Policy*, *30*(10-11), 552-568. https://doi.org/10.1016/jtelpol.2006.09.006

Almana, A. M., Aksoy, M. S., & Alzahrani, R. (2014). A survey on data mining techniques in customer churn analysis for telecom industry. *Journal of Engineering and Applications*, *4*(5), 165-171.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in telecommunication sector using rough set approach. *Neurocomputing*, *237*(1), 242-254. https://dx.doi.org/10.1016/j.neucom.2016.12.009

Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, Sajid. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, *94*(1), 290-301. https://doi/org/10.1016/j.jbusres.2018.03.003

Amin, Adnan., Shah, Babar., Khattak, Asad, Masood., Moreira, Fernando, Joaqim, Lopes., Ali, Gohar., Rocha, Alvaro., Anwar, Sajid. (2018). Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. *International Journal of Information Management*, *46*(6), 304-319. https://doi.org/10.1016/j.ijnfomgt.2018.08.015

Ansari, F. (2021). Cross Validation Techniques. Analytics Vidhya. https://medium.com/analytics-vidhya/cross-validation-techniques-bacb582097bc

Babu, S., & Ananthanarayanan, NR. (2014). A review on customer churn prediction in telecommunication using data mining techniques. *International Journal Science Engineering Research*, *4*(1), 35-40.

Babu, S., & Ananthanarayanan, NR. (2018). Enhanced prediction model for customer churn in telecommunication using EMOTE. *Advances in Intelligent System and Computing*, *632*(1), 465-475. Springer Nature Singapore. https://doi.org/10.1007/978-981-10-5520-1_43

Barua, S., Islam, M. M., Yao X., & Murase, K. (2014). MWMOTE--Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning. In IEEE

Transactions on Knowledge and Data Engineering, vol. 26, no. 2, pp. 405-425, Feb. 2014, https://doi.org/10.1109/TKDE.2012.232

Bhargava, M., Singh, S., Sharma, J., & Franklin, Vinod, D. (2022). Telecom customer churn prediction. *Lecture Notes on Data Engineering and Communication Technologies – LNDECT*, *92*(1), 325-334. https://doi.org/10.1007/978-981-16-6601-8_30

Chawla, NV., Bowyer, KW., Lawrence, O., Hall, W., Kegelmeyer, P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research (JAIR)*, 16, pp. 321-357.

Chen, Z. Y., Fan, Z. P., & Sun, M. (2012). A hierarchical multiple kernel Support Vector Machine for customer churn prediction using longitudinal behavioural data. *European Journal of Operational Research*, *223*(2), 461-472. https://doi.org/10.1016/jejor.2012.06.040

Chouinard, J. C. (2022). Decision Trees in Machine Learning. JCChouinard. https://www.jcchouinard.com/decision-trees-in-machine-learning/

Crone, Sven, F., Lessmann, Stefan., Stahlbock, Robert. (2006). The Impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*, *173*(3), 781-800. https://doi.org/10.1016/j.ejor.2005.07.023

Dabakoglu, C. (2018). What is Support Vector Machine (SVM)? - with Python. Medium. https://medium.com/@cdabakoglu/what-is-support-vector-machine-svm-fd0e9e39514f

Darrieus, G.J.M. (1931) Turbine having its Rotating Shaft Transverse to the Flow of the Current, United States Patent No. 1,835,018, 8th December 1931.

Devi, P. U., & Madhavi, S. (2012). Prediction of churn behaviour of bank customer using data mining tools. *Business Intelligence Journal*, *5*(1), 96-101.

Farnadi, G., Sitaraman, G., Sushimita, S., Celli, F., Kosinki, M., Stillwell, D., Davalos, S., Moens, M. F., & De Cock, M. (2016). Computational personality recognition in social media. *User Model User-Adap Inter*, *26*(1) 109-142. https://10.1007/s11257-016-9171-0

Fei, T. Y., Shuan, L. H., Yan, L. J., Xiaoning, G., & King, S. W. (2017) Prediction on customer churn in the telecommunications sector using discretization and Naïve Bayes classfier. *International Journal Advance Soft Computer Application*, *9*(3), 23-35.

Gusain, R. (2020). Naive Bayes Classifier from Scratch with Hands On Examples in R. Insight - Data Science Society, IMI, New Delhi. https://insightimi.wordpress.com/2020/04/04/naive-bayes-classifier-from-scratch-with-hands-on-examples-in-r/

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concept and techniques (3rd ed.) Elsevier. Third Edition.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322-1328, https://10.1109/IJCNN.2008.4633969

Hyeon, Ahn, Jae., Pil, Han, Sang., Seop, Lee, Yung. (2006). Customer churn analysis: Churn determinants and mediation effect of partial defection in the Korea Mobile telecommunication service industry. *Telecommunications Policy*, *30*(10), 552-568. https://doi.org/10.1016/jtelpol.2006.09.006

IBM. (2021). IBM SPSS Modeler CRISP-DM Guide.

Keramati, A., & Ardabili, M. S. S. (2011) Churn analysis for an Iranian mobile operator. *Telecommunications Policy*, *35*(4), 344-356. https://doi.org/10.1016/j.telpol.2011.02.009

Khan, A. A., Jamwal, S., & Sepehri, M. M. (2010). Applying data mining to customer churn prediction in an internet service provider. *International Journal of Computer Applications*, *9*(7), 8-14. https://doi.org/10.5120/1400-1889

Khan, Rafi, Ahmad., & Quadri, S. M. K. (2012). Business Intelligence: An integrated approach. *Business Intelligence Journal*, *64*(1), 1-8.

Khodabandehlou, S., & Rahman, M. Z. (2017). Comparison of supervised machine learning techniques for customer churn prediction on analysis of customer behaviour. *Journal of System and Information Technology*, *19*(1), 65-93. https://doi.org/10.1108/JSIT-10-2016-0061

Kotler, Philip. (2002). *Marketing management: Millenium Edition*. (1st ed.). Pearson Custom Publishing.

Kovac, M. (2020). Introduction to the autonomous artificial intelligence systems. *Judgement Proof Robots and Artificial Intelligence*, *1*(1), 47-63. https://doi.org/10.1007/978-3-030-53644-2_4

Lalwani, P., Mishra, M.K., Chadha, J.S., Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing,* 104, 271–29. https://doi.org/10.1007/s00607-021-00908-y

Lee, R., & Murphy, J. (2008). The moderating Influence of enjoyment on customer loyalty. *Australian Marketing Journal*, *16*(2) 11-21. https://doi.org/10.1016/S1441-3528(08)70011-9

Lemmens, A., & Croux, C. (2006). Bagging and boosting classification tress to predict churn. *Journal of Marketing Research*, *43*(2), 277-286. https://doi.org/10.1509/jmkr.43.2.276

Li, Y., Hou, B., Wu. Y., Zhao, D., Xie, A., & Zou, P. (2021). Giant fight: Customer churn prediction in traditional broadcast industry. *Journal of Business Research*, *131*(7), 630-639. https://doi.org/10.1016/j.jbusres.2021.01.022

Laffont, J. J., & Tirole, J. (2001). Competition in Telecommunication. First MIT Press Paperback Edition.

Migueis, V. L., Van den P. L., Camanho, A. S., & Falcao, C. J. F. (2012) Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert System with Applications*, *39*(12), 11250-11256. https://doi.org/10.1016/j.eswa.2012.03.073

Mutanen, T., Nousiainen, S., & Ahola, J. (2010). Customer churn prediction – A case study in retail banking. *Frontiers in Artificial Intelligence and Applications*, 218(1), 77-83. https://doi.org/10.3233/978-1-60750-633-1-77

Navlani, A. (2018). KNN Classification Tutorial using Scikit-learn. Datacamp. https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn

Neutelings, I. 2021. Neural Networks. TikZ. https://tikz.net/neural_networks/

Noorhalim, N., Ali, A., Shamsuddin, S.M. (2019). Handling Imbalanced Ratio for Class Imbalance Problem Using SMOTE. In: Kor, LK., Ahmad, AR., Idrus, Z., Mansor, K. (eds) Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017). Springer, Singapore. https://doi.org/10.1007/978-981-13-7279-7_3

Oghojafor, B., Mesike, G., Bakarea, R., Omoera, C., & Adeleke, I. (2012). Discriminant analysis of factors affecting telecoms customer churn. *International Journal of Business Administration*, 3(2), 59-67. https://dx.doi.org/10.5430/ijba.v3n2p59

Omar, A. (2014). Predicting customer churn in telecom industry using multilayer preception Neural Network: Modelling and analysis. *Life Science Journal*, *11*(03), 72-81.

Prastiwi, A., Ulubalang, D. T. (2020). Boosting Algorithm (AdaBoost and XGBoost). Algotech. https://algotech.netlify.app/blog/xgboost/

Purnamasari, I., Handayanna, F., Arisawati, E., Dewi, L. S., Sihombing, E. G., & Rinawati. (2020). The determination analysis of telecommunications customers potential cross-selling with classification Naïve Bayes and C4.5. *Journal of Physics: Conference Series*, *16*(1), 1-8. https://doi.org/10.1088/1742-6596/1641/1/012010

Rodan, A., Faris, H., Alsakran, J., & Al-Kadi, O. (2014). A Support Vector Machine approach for churn prediction in telecom industry. *Information An International Interdisciplinary Journal*, *17*(8), 3961-3970.

Rosotto, C. M., Sekkat, K., & Varoudakis, A. (2005). Opening up telecommunications to competition and MENA Integration in the World Economy. https://doi.org/10.1002/jld.1185

Rousan, Al., Ramzi, M., & Mohamed, B. (2010). Customer loyalty and the impact of service quality: The case of five star hotels in Jordan. *International Journal of Human and Social Science*, *5*(13), 1-7. https://doi.org/10.5281/zenodo.1331937

Safitri, A. R., & Muslim, M. A. (2020). Improve accuracy of Naïve Bayes classifier for determination of customer churn uses SMOTE and genetic algorithms. *Journal of Computing Exploration*, *1*(1), 70-75. https://doi.org/10.52465/joscexv1i1.5

Saini, N., Monika., & Grag, K. (2017). Churn prediction in telecommunication industry using Decision Tree. *International Journal of Engineering Research and Technology (IJERT)*, *6*(4). 439-443. https://doi.org/10.17577/ijertv6is040379

Sebastian, H. T., & Wagh, R. (2017). Churn analysis in telecommunication using Logistic Regression. *Oriental Journal of Computer Science & Technology*, *10*(1), 207-212. https://dx.doi.org/10.13005/ojcst/10.01.28

Shapley, L. S. (1953). A value for n-person games. Contributions to the Theory of Games, 2.28, pp. 307-317.

Shouval, R., Fein, J. A., Savani, B., Mohty, M., & Nagler, A. (2020). Machine learning and aritificial intelligence in haematology. British Journal Haematology, 192(1), 239-250. https://doi.org/10.1111/bjh.16915

Singh, H., & Samalia, H. V. (2014). A Business intelligence perspective for churn management. *Social and Behavioural Sciences*. *109*(1), 51-56. https://doi.org/10.1016/jsbspro.2013.12.420

Sivasankar, E., & Vijaya, J. (2018) Hybrid PPFCM-ANN Model: An efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and Artificial Neural Network. *Neural Computing and Applications*, *31*(11), 7181-7200. https://doi.org/10.1007/s00521-018-3548-4

Sruthi, E. R. (2021). Understanding Random Forest. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

Sun, A. Y., & Scanlon, B. R. (2019). How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. *Environmental Research Letter*, *14*(7), 1-28. https://doi.org/10.1088/1748-9326/ab1b7d

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, 7, 2169-3536. https://doi.org/10.1109/ACCESS.2019.2914999

Appendices

Table 5. Individual SHAP feature value for Customer #1

	Feature	phi	phi.var	Feature.value	
1	Gender	No	0.1	0.091	gender=1
2	SeniorCitizen	No	0.01	0.010	SeniorCitizen=0
3	Partner	No	0.2	0.162	Partner=2
4	Dependents	No	0.02	0.020	Dependents=1
5	Tenure	No	-0.11	0.119	tenure=1
6	PhoneService	No	-0.09	0.083	PhoneService=1
7	MultipleLines	No	-0.02	0.020	MultipleLines=2
8	InternetService	No	0.11	0.099	InternetService=1
9	OnlineSecurity	No	-0.06	0.057	OnlineSecurity=1
10	OnlineBackup	No	0.22	0.214	OnlineBackup=3
11	DeviceProtection	No	-0.01	0.010	DeviceProtection=1
12	TechSupport	No	-0.07	0.106	TechSupport=1
13	StreamingTV	No	0.02	0.020	StreamingTV=1
14	StreamingMovies	No	0.01	0.010	StreamingMovies=1
15	Contract	No	-0.1	0.091	Contract=1
16	PaperlessBilling	No	-0.05	0.048	PaperlessBilling=2
17	PaymentMethod	No	0.01	0.030	PaymentMethod=3

	Feature	phi	phi.var	Feature.value	
18	MonthlyCharges	No	0.18	0.149	MonthlyCharges=29.85
19	TotalCharges	No	-0.03	0.130	TotalCharges=29.85
20	Gender	Yes	-0.1	0.091	gender=1
21	SeniorCitizen	Yes	-0.01	0.010	SeniorCitizen=0
22	Partner	Yes	-0.2	0.162	Partner=2
23	Dependents	Yes	-0.02	0.020	Dependents=1
24	Tenure	Yes	0.11	0.119	tenure=1
25	PhoneService	Yes	0.09	0.083	PhoneService=1
26	MultipleLines	Yes	0.02	0.020	MultipleLines=2
27	InternetService	Yes	-0.11	0.099	InternetService=1
28	OnlineSecurity	Yes	0.06	0.057	OnlineSecurity=1
29	OnlineBackup	Yes	-0.22	0.214	OnlineBackup=3
30	DeviceProtection	Yes	0.01	0.010	DeviceProtection=1
31	TechSupport	Yes	0.07	0.106	TechSupport=1
32	StreamingTV	Yes	-0.02	0.020	StreamingTV=1
33	StreamingMovies	Yes	-0.01	0.010	StreamingMovies=1
34	Contract	Yes	0.1	0.091	Contract=1
35	PaperlessBilling	Yes	0.05	0.048	PaperlessBilling=2
36	PaymentMethod	Yes	-0.01	0.030	PaymentMethod=3
37	MonthlyCharges	Yes	-0.18	0.149	MonthlyCharges=29.85

	Feature	phi	phi.var	Feature.value	
38	TotalCharges	Yes	0.03	0.130	TotalCharges=29.85