Ahliya Journal of Business Technology and MEAN Economies
[Published by Palestine Ahliya University]
Vol. 01 Issue 01 (2024)
ISSN: 3007-9691

Predictive Customer Analytics: Machine Learning for Churn Prediction and Retention

Tasneem Qaraeen[*1], Nora Qaqour [**1], Sameh Taqatqa [***1]
[1] Faculty of Engineering and Information Technology, Palestine Ahliya University (Palestine)
[*] tasneemqar7@gmail.com
[**] nuraqaqour@gmail.com
[***] Sameh@paluniv.edu.ps

Received:17/06/2024 Accepted:03/07/2024 Published:15/07/2024

Abstract:
Customer churn presents a big challenge in the industry. Businesses have to deal with the problem of customers stopping using their products and services due to dissatisfaction, competitive offers, more affordable alternatives, or changing needs. Churn can be damaging to businesses since it causes revenue loss and higher costs. To address this issue, our research aimed to develop a prediction model that helps predict customer churn. We started with getting the data set about telecommunication. Our analysis and model development were based on this dataset. Then we did data visualization to gain a better understanding of the data through multiple charts. After that, we performed data preparation. First, we did data transformation, data cleansing to address missing values and outliers; feature selection was done, and finally, in this step, the data set was split into testing and training sets. Multiple machine learning algorithms were used for modeling, such as decision trees, random forests, logistic regression, support vector machines, Naïve Bayes, and neural networks. Following model development, we evaluated the model performance with each algorithm using tune model hyperparameters. The decision tree algorithm performed the best with %96.7 accuracy, %96.9 precision, %99.3 recall, and %98.1 F1-score. These findings showed how effective decision tree algorithms are in predicting customer churn. This predictive model will enable telecommunication businesses to predict potential churn, make retention strategies, reduce customer churn and increase customer retention rates.

Keywords: *Customer churn; Machine learning; Tune Model Hypermeters; Accuracy; Precision; Recall; F1-score.*

التحليلات التنبؤية للعملاء: التعلم الآلي للتنبؤ بالتوقف والاحتفاظ به

تسنيم قراعين*[1]، نورة قعقور**[2]، سامح طقاطقة***[3]

[1] كلية الهندسة وتكنولوجيا المعلومات، جامعة فلسطين الأهلية (فلسطين)

* tasneemqar7@gmail.com

** nuraqaqour@gmail.com

*** Sameh@paluniv.edu.ps

تاريخ الاستلام:2024/06/17 تاريخ القبول:2024/07/03 تاريخ النشر:2024/07/15

ملخص:

يُمثل تسرّب العملاء تحديًا كبيرًا في قطاع الصناعة. إذ يتعين على الشركات التعامل مع مشكلة توقف العملاء عن استخدام منتجاتها وخدماتها بسبب عدم الرضا أو العروض المنافسة أو البدائل الأكثر تكلفة أو تغير احتياجاتهم. وبهذا يمكن أن يكون للتسرب تأثير ضار على الشركات؛ كونه يتسبب في خسارة الإيرادات وزيادة التكاليف. ولمعالجة هذه القضية، هدفت الدراسة إلى تطوير نموذج يساعد في التنبؤ بتسرب العملاء. إذ تم الحصول على مجموعة بيانات حول الاتصالات. وتم بناء التحليل وتطوير النموذج استنادًا إلى هذه المجموعة. ثم تم إجراء تصوّر للبيانات لفهم أفضل للبيانات من خلال الكثير من الرسوم البيانية. بعد ذلك، قام الباحثون بتحضير البيانات. وتحويل البيانات وتنظيفها لمعالجة القيم المفقودة والقيم الشاذة؛ وتم اختيار الميزات، وأخيرًا، في هذه الخطوة، تم تقسيم مجموعة البيانات إلى مجموعات اختبار وتدريب. وتم استخدام مجموعة من خوارزميات التعلم الآلي للنمذجة، مثل الأشجار القرار، والغابات العشوائية، والانحدار اللوجستي، وآلات المتجهات الداعمة، وخوارزمية نايف بايز، والشبكات العصبية. وبعد تطوير النموذج، تم تقييم أداء النموذج مع كل خوارزمية باستخدام ضبط معلمات النموذج. إذ حققت خوارزمية شجرة القرار أفضل أداء بدقة إجمالية بلغت (96.7%)، ودقة للتصنيف (96.9%)، واسترجاع (99.3%)، ودرجة F1 بلغت 98.1%. وأظهرت هذه النتائج مدى فعالية خوارزميات شجرة القرار في التنبؤ بتسرّب العملاء. وبهذا سيمكن هذا النموذج التنبؤي شركات الاتصالات من التنبؤ بالتسرب المحتمل، ووضع استراتيجيات للاحتفاظ بالعملاء، وتقليل تسرّب العملاء وزيادة معدلات الاحتفاظ بالعملاء.

الكلمات المفتاحية: *تسرّب العملاء؛ التعلم الآلي؛ ضبط معلمات النموذج؛ الصحة (الدقة الإجمالية)؛ الضبط (الدقة)؛ الاسترجاع؛ درجة F1.*

1. Introduction

Nowadays businesses are facing several challenges due to competition and market trends among these challenges, one of the major issues for businesses is customer churn or could be known as customer attrition, Churn is the process by which clients leave a business and stop using its services or products for a variety of reasons, such as dissatisfaction with the service or products, better offers from competitors, switch to a more affordable alternative or changing in customer needs, The number of customers who come in and go out over a given time period is measured as the churn rate, Customer churn is acknowledged as one of the contributing factors that lead to a business's or organization's decrease in profitability as well as other financial and social losses (Alwis., 2018). To address this challenge businesses have turned to customer relationship management (CRM) a comprehensive approach for establishing, maintaining and enhancing relationships with customers, it is widely used in a variety of industries, including retail, banking, insurance, and telecommunications. Retaining customers is one of the CRM system's primary goals and the most challenging one and, it is considered the core of CRM because it is now far more expensive to acquire new clients than it is to retain current ones, customer retention is a significant benefit for all businesses and a key tactic for ensuring sustained profitability and organizational success, also Effectively managing customer churn through CRM strategies is an essential requirement to sustain relationship with customers and ensuring customer retention (Vafeiadis, 2015; Seyed et al., 2019).

However, the true transformation arises when machine learning methods are seamlessly integrated using historical data from CRM systems with machine learning algorithms to develop a customer churn prediction model, the development of an efficient customer churn prediction model has been gaining attention among academics, businesses, and marketing researchers who are collectively focused on identifying the reasons behind customers churn (Umayaparvathi & Iyakutti, 2016).

Significantly, this shift comes with considerable cost advantages, attracting new customers is five to six times more expensive than keeping existing ones, also researchers indicate that the implementation of a churn predicting model can raise business profit from 25% to 85% when the churn rate is reduced by 5%, in conclusion, Churn prediction models when effectively implemented can increase a company's profits, reduce costs while ensuring customer retention and its considered a key tactic for ensuring sustained profitability and organizational success (Ullah et al., 2019).
Recognizing the importance of this topic, Customer Churn Prediction System aims to use a data-driven system that predicts or forecasts which customers are likely to stop using a company's products or services in the near future using historical customer data and machine learning algorithms, allowing enterprises to take proactive measures to preserve these key connections.

2. Literature Review

Churn prediction modeling is important for businesses as it allows them to proactively retain customers, minimize revenue loss, and enhance profitability. This section will review previous research and studies that have explored this topic:

The researchers in Brandusoiu and Toderean (2013) proposes an innovative methodology that uses data mining techniques on a dataset of call detail records to estimate customer churn in mobile telecommunication industry. This paper compares and evaluates three data mining algorithms: Quick Unbiased Efficient Statistical Tree, Chi-squared Automatic interaction Detection Tree (DT), and Classification and Regression Tree. By analyzing the findings, they discovered that all three models perform almost equally well at predicting both churners and non-churners. The three model have a good predicting performance (around 80%) for customers churn, in addition , Researchers in Wagh et al. (2024) focus on addressing churn issues. They employ classification techniques to identify subscribers likely to leave and collect reasons behind their churn in the telecom industry. The primary objective is to examine different ML algorithms to know reasons for churn by using model prediction. The system collects customer data using classification algorithms like Random Forest (RF), machine learning techniques such as K-Nearest Neighbors (KNN), and DT classifiers. It provides an effective business model that evaluates churn data and accurately predicts churn, enabling proactive actions to prevent revenue loss. The system achieves 99% accuracy using the RF classifier for churn prediction, with a precision and recall of 99%. This research not only enhances churn prediction but also has implications for other business sectors, providing predictive models to improve customer service and reduce churn effectively.

In study Kisioglu and Topcu (2011) a system is created to identify the most significant factors influencing on of clients who are likely to leave in Turkey's telecom companies. A Bayesian Belief Network (BBN) model is built to identify the actions of consumers who are likely to leave. The CHAID (Chi-squared Automatic Interaction Detector) algorithm is used to discretize continuous data since BBN require discrete variables. The average number of minutes spent on calls, average billing amount, and the frequency of calls to people from different providers are the factors that were found to be the most significant in explaining customer churn.

Researchers in Gürsoy (2010) aimed to identify the factors that contribute to the telecom company's customer attrition also to identify customers who want to churn and target them with targeted campaigns. Among the classification techniques used are logistic regression (LR) and decision trees (DT) analysis to identify the causes of customer attrition. The analysis shows subscribers with the following attribute have very high tendency to churn: Subscribers who lack discounted packages are more likely to churn, so creating diverse and appealing packages to match various calling behaviors is essential. Additionally, the volume of incoming local and long-distance calls is a crucial factor in churn analysis. Developing packages tailored to inner-city and long-distance call patterns can help address subscriber concerns. Furthermore, a decrease in the proportion of calls received from the same operator has been found to heighten the likelihood of churn.
Based on the analysis and the provided recommendations, fostering stronger relationships with existing subscribers, boosting demand for the company's various services, minimizing churn rates, and increasing business profit margins can all be achieved.

Researchers in Bilişik and Sarp (2023) The objective was to identify the optimal model for predicting customer churn in the telecommunications industry using machine learning techniques. The study utilized corporate customer data from one of Turkey's top three telecom companies, featuring a dataset with 593 columns specifically designed for corporate customer analytics. The initial phase of the analysis involved data preparation, including descriptive data analysis. Four different algorithms—RF, LR, DT, and gradient boosting—were employed to build churn prediction models. Among these, the regression algorithm was identified as the most effective for predicting customer churn

Researchers in Kavitha et al. (2020) is focusing on customer churn detection to retain customers and reduce churn rates. This research uses machine learning techniques to predict customers likely to cancel subscriptions, offering better services and reducing churn rates. the researchers used various algorithms like RF, XGBoost & Logistic Regression to find accurate values to predict churn of customers, they implemented the model on a dataset that is trained and tested to have the most accurate values to predict, they found out after performing several experiments on the proposed churn model using machine learning algorithms on the dataset that RF is the most accurate and efficient compared with Extreme Gradient Boosting (XGBoost) & LR. Finally, they visualized the data using the convolutional neural network (CNN) model. These experiments demonstrated the effectiveness of RF in predicting customer churn.

In Ahmad (2019) the research centers on creating a churn prediction model for telecom operators to forecast customers at risk of leaving. Employing machine learning methods on a big data platform, the model utilizes the Area under Curve (AUC) as its standard measure. The model achieved a performance rating with an AUC value of 93.3%. It was validated on a substantial dataset from SyriaTel Telecom Company, testing its efficacy across four algorithms: DT, RF, Gradient Boosted Machine Tree (GBM), and XGBOOST, the study analyzed the performance of classification algorithms using different sizes of training data and unbalanced datasets, The system was trained, tested, and evaluated using the dataset, which included all customer data collected over a nine-month period. Four algorithms were tested by the model: XGBOOST, RF, GBM, and DT. However, using the XGBOOST algorithm produced the best outcomes.

In Ebrah and Elnasir (2019) the researchers aimed to identify the most accurate churn prediction model in the telecom sector and determine the key feauturs influencing customer churn using machine learning algorithms like Naive Bayes (NB), Support Vector Machine (SVM), and Decision Trees. They conducted their analysis on IBM Watson and cell2cell datasets, evaluating model performance based on the Area Under Curve (AUC). The models achieved AUC scores of 0.82, 0.87, and 0.77 for NB, SVM, and decision trees respectively on IBM Watson, and 0.98, 0.99, and 0.98 on cell2cell. SVM yielded the highest accuracy with AUC scores of 0.87 on IBM Watson and 0.99 on cell2cell, surpassing previous studies on the same datasets. These findings are crucial for businesses aiming to devise effective customer retention strategies.

In Umayaparvathi and Iyakutti (2016) A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics : Are reviewed in this research The many customer data categories that are accessible in publicly accessible datasets, as well as the predictive models and performance measures employed in the literature for churn prediction in the telecom sector. There are 50,000 records in the dataset overall. The results were as follows: 3672 customers are churners and the remaining 46328 are non-churners. Based on the information gathered about its clients, such as the client's credit score, bill and payment information, use patterns, and value-added services;3) consumer demographics and personal information; and customer care service details. The present study started by presenting the problem of churn prediction and underscored the significance of employing predictive modeling techniques to tackle the issue of customer attrition in the telecommunication sector. They provide a summary of the distinct features of churn prediction datasets that are available to the general public. They also examined and classified the different consumer characteristics used in churn prediction.

3. Research Methodology

The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology is utilized to develop a predictive model for customer churn. CRISP-DM serves as a framework for data mining projects, outlining the project phases, tasks, and outputs. It begins with the data understanding phase, involving initial data collection to familiarize oneself with the data, identify data quality issues, and gain preliminary insights. The next phase, data preparation, involves creating a final dataset through tasks like attribute selection, data cleansing, and transformation for modeling purposes. Following this is the modeling phase, where the appropriate machine learning model is selected. The evaluation phase assesses the model's accuracy in predicting churn rates to ensure it meets business objectives.

Finally, in the deployment phase, the model is implemented based on project requirements, ranging from generating reports to establishing a repeatable data mining process, ensuring clarity on necessary actions for utilizing the developed models (Wirth & Hipp, 2000; Martinez-Plumed et al., 2021).

CRISP-DM methodology was selected for this research due to its Structured Approach making sure you complete every necessary phase of the research life cycle which can assist in staying organized and avoiding missing any important steps.

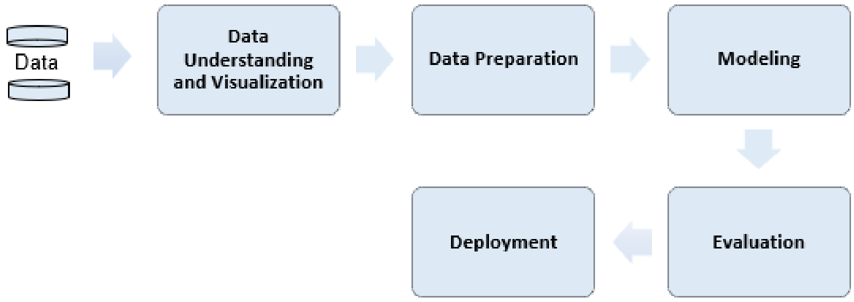

Figure 1. Research Methodology

3.1 Data Source

The dataset in this research is from Kaggle for a telecommunication company (www.kaggle.com/datasets/mnassrib/telecom-churn-datasets) the consumption data of customers who use this company services. The dataset contains customer's information and historical behavior while using the service that the telecommunication company provide. This public dataset contains 21 attributes mostly about customer usage patterns.

3.2 Data understanding and visualization

Table 1. provides a detailed overview of the features within the dataset, outlining their types and meanings.

Table 1. Dataset Description

	Features	Description	Type
1	State	Customer state (churn or not churn)	Character
2	account length	How long the account has been active, measured in months.	Numeric
3	area code	The digit code at the beginning of the customer's phone number	Numeric
4	phone number	The customer's contact number.	Numeric
5	international plan	customer has a plan that includes international calling or not.	Character
6	voice mail plan	customer has a plan that includes voicemail service or not.	Character
7	number vmail messages	The number of voicemail messages the customer has.	Numeric
8	total day minutes	The total duration of daytime calls.	Numeric
9	total day calls	The total count of daytime calls.	Numeric
10	total day charge	The total cost incurred from daytime calls.	Numeric
11	total eve minutes	The total duration of evening calls.	Numeric
12	total eve calls	The total count of calls made during the evening.	Numeric
13	total eve charge	The total cost incurred from evening calls	Numeric
14	total night minutes	The total duration of calls made during the night	Numeric
15	total night calls	The total count of calls made during the night	Numeric
16	total night charge	Total charge for the night calls.	Numeric
17	total intl minutes	Total minutes of international calls.	Numeric
18	total intl calls	Total number of international calls.	Numeric
19	total intl charge	Total charge for the international calls.	Numeric
20	customer service calls	Number of times the customer called customer service.	Numeric
21	Churn	Whether the customer churned or not (True/False).	Character

This dataset contains a total of (3333) records. Among these, 483 customers are churners, representing (14.5%) of the total. The remaining (2850) customers are non-churners, accounting for (85.5%) of the dataset, as shown in figure 2.

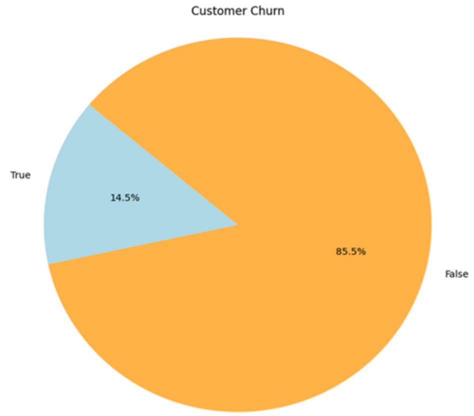

Figure 1. Customer churn and non-churn

3.3 Data preparation

Data preparation is a crucial preprocessing stage for data analysis, serving as a component of both data warehousing and data mining. The quality of the input data has a significant impact on the quality of the analysis results. During the data preparation phase, activities involve selecting tables, records, and attributes from the original raw data that will be fed into modeling tools. Tasks include cleaning the data to address inconsistencies and errors, as well as transforming it to ensure compatibility with modeling tools (Wirth & Hipp, 2000).

3.4 Data Transformation

Data transformation is a process in data preparation that involves converting data from one format or structure into another in order to prepare the data in a way that is compatible with the algorithms and analytical techniques to be used, this process is essential. By implementing these changes, the data is made more receptive to the analysis tools of choice, allowing for a more insightful and meaningful investigation of the patterns, relationships, and insights contained within the dataset (Sattler & Box, 2001). One crucial aspect of data transformation is encoding. Categorical values cannot be processed by machine learning algorithms and nearly all deep learning architectures. This implies that in order for the algorithms to solve problems with regression or classification, their input must be numerical. As a result, encoding techniques must be used to convert these categorical variables into numerical values (Dahouda & Joe, 2021).

In our research we used encoding to transform categorical variables to numerical variables as presented in the table 2 and figure 3.

Table 2: Encoding of Categorical Variables

Categorical variables	Numerical value
international plan	Yes=0, no=1
voice mail plan	Yes=0, no=1
Churn	True=0, false=1
State	0=IN, 1=KS, 2=OH, 3=NJ, 4=OK,5=AL,6=MA, 7=MO, 8=LA, 9=WV, 10=RI, 11=IA, 12=MT, NY=13, 14=ID, 15=VT, 16=VA, 17=TX, 18=FL, 19=CO, 20=AZ, 21=SC,22=NE,23=WY, 24=HI, 25=IL, 26=NH,27=GA, 28= AK,29= MD,30= AR,31= WI, 32= OR, 33=MI, 34= DE, 35= UT, 36=CA ,37=MN ,38=ME,39=SD ,40=NC ,41=WA, 42= NM, 43= NV, 44= DC, 45=KY, 46=MS, 47=TN, 48=PA, 49=CT, 50=ND.

```
      international plan  voice mail plan  churn  state
0                      1                0      1      1
1                      1                0      1      2
2                      1                1      1      3
3                      0                1      1      2
4                      0                1      1      4
...                  ...              ...    ...    ...
3328                   1                0      1     20
3329                   1                1      1      9
3330                   1                1      1     10
3331                   0                1      1     49
3332                   1                0      1     47

[3333 rows x 4 columns]
```

Figure 3. Data Transformation

3.5 Data Cleansing

Data cleansing is the procedure of identifying and rectifying errors, inconsistencies, and inaccuracies found within a dataset. The objective is to enhance the overall quality of the data by addressing issues that could adversely affect the results of data analysis or machine learning models. One major task in data cleansing is dealing with missing values in a dataset, missing values are the absence of data for a specific variable or observation. There are a number of reasons why these records could be missing, such as mistakes made during data collection, non-responders to surveys, or just the absence of pertinent or unavailable information. Since missing values can affect the accuracy and reliability of statistical analysis and machine learning models, handling them is an essential part of data preprocessing (Abdallah & Webb, 2017) Several methods can be employed to handle missing values effectively Common approaches include deletion methods like listwise or pairwise deletion, imputation methods such as mean or regression imputation, predictive modeling techniques like K-Nearest Neighbors, and more advanced methods like multiple imputation and matrix factorization (Krause et al., 2020). In our research, we found no missing values, as shown in Figure 4.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3333 entries, 0 to 3332
Data columns (total 21 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   state                   3333 non-null   object
 1   account length          3333 non-null   int64
 2   area code               3333 non-null   int64
 3   phone number            3333 non-null   object
 4   international plan      3333 non-null   int64
 5   voice mail plan         3333 non-null   int64
 6   number vmail messages   3333 non-null   int64
 7   total day minutes       3333 non-null   float64
 8   total day calls         3333 non-null   int64
 9   total day charge        3333 non-null   float64
 10  total eve minutes       3333 non-null   float64
 11  total eve calls         3333 non-null   int64
 12  total eve charge        3333 non-null   float64
 13  total night minutes     3333 non-null   float64
 14  total night calls       3333 non-null   int64
 15  total night charge      3333 non-null   float64
 16  total intl minutes      3333 non-null   float64
 17  total intl calls        3333 non-null   int64
 18  total intl charge       3333 non-null   float64
 19  customer service calls  3333 non-null   int64
 20  churn                   3333 non-null   int64
dtypes: float64(8), int64(11), object(2)
memory usage: 546.9+ KB
None
```

Figure 4. Dealing with missing data

3.6 Feature Selection

The process of finding and selecting the most important variables from a dataset enhance the accuracy and efficiency of a model is known as feature selection with the aim to improve performance such as estimated accuracy, visualization and comprehensibility of learned knowledge. techniques for devising best feature selection approaches are gain ratio, information gain, Relief and Correlation based feature selection (Khalid & Nasreen., 2014).

Table 3: Feature selection techniques

Feature Selection Techniques	Features
Correlation	International plan, Customer service calls, Total day minutes, Total day charge, Voice mail plan, Total eve minutes, Total eve charge, Number voice mail messages, Total international charge, Total international minutes, Total international calls, Total night charge, Total night minutes, State, Total day calls, Account length, Phone number, Total eve calls, Area code, Total night calls.
Gain Ratio	Customer service calls, international plan, Phone number, Total day minutes, Total day charge, Total international charge, Total international minutes, Voice mail plan, Number voice mail messages, Total international calls, Total eve minutes, Total eve charge, State, Account length, Area code, Total night calls, Total night minutes, Total day calls, Total eve calls, Total night charge.
Information Gain	Phone number, Total day charge, Total day minutes, Customer service calls, international plan, State, Voice mail plan, Number voice mail messages, Total International calls, Total international charge, Total international minutes, Total eve charge, Total eve minutes, Total night minutes, Account length, Area code, Total day calls, Total night calls, Total eve calls, Total night charge.
Relief	State, Customer service calls, Total day minutes, Total day charge, Total eve minutes, Total eve charge, Area code, Voice mail plan, international plan, Number voice mail messages, Total international charge, Total international minutes, Total night minutes, Total night charge, Total day calls, Account length, Total eve calls, Total night calls, Total international calls, Phone number.

In our research we choose features used the previous feature selection techniques and the frequency with which each feature as shown in table 3. appears at the top of the rankings provided by each feature selection technique is used to identify the features that are most significant overall. Based on this, the result showed that the most important features are international plan, Customer service calls, Total day minutes, Total day charge, Voice mail plan, Total eve minutes, total eve charge and number vmail messages. And the least important features are Total night calls, Total eve calls, Area code, Total night charge, Phone number and Total day calls.

3.7 Splitting Data

Data splitting, which involves dividing a dataset into two separate sets for testing and training, is a commonly used technique for model validation. In order to evaluate and compare predictive performance while avoiding overfitting concerns on the training set, models are trained on the training set and validated using the testing set. The most common technique for splitting data is random subsampling, which selects rows at random for testing without replacement and keeps the remaining rows for training to guarantee a representative validation process (Joseph, 2022). In our study, we split the data into a 70/30 ratio because our dataset is relatively small, and this split gave us the best results. This allocation means that 70% of the data is used for training our models, while the remaining 30% is set aside for testing purposes

3.8 Generating Machine Learning Models

Machine learning (ML) is the scientific study of statistical models and algorithms used by computer systems to carry out certain tasks without being explicitly programmed. These models, which are essentially computer programs, are designed to recognize patterns in data or make predictions. Machine learning can be supervised or unsupervised. Supervised learning is used when data is limited and clearly labeled for training purposes. These models are created from machine learning algorithms. Data scientists employ various machine learning algorithms as the foundation for different models, depending on their specific purposes, such as prediction modeling or categorization. When a given algorithm is applied to data, the data is altered to better handle a certain task and develops into a machine learning model (Mahesh, 2020; Sullivan, 2022).

Supervised Machine Learning (SML) entails looking for algorithms that can infer general hypotheses from examples supplied by outside sources and use those predictions to forecast results for upcoming occurrences. Supervised learning encompasses two primary types: classification and regression. Both involve mapping an input x to an output y. In classification, the goal is to assign inputs to predefined classes based on training data. Regression, on the other hand, predicts continuous values and aims to establish a best-fit line or curve that represents the relationship within the data. In summary, regression deals with quantitative outputs, while classification deals with qualitative outputs (Mohamed, 2017; Hu et al., 2021; Konieczny & Idczak, 2016; Soofi & Awan, 2017).

There are many algorithms for classification tasks in machine learning. Some of the most commonly used ones include:

A. Decision Tree (DT)

DT algorithm, a popular data mining method for classification systems and prediction algorithms. It highlights its ease of use, interpretability, and robustness with missing values. DT can handle both discrete and continuous variables and have applications in various fields. Key applications include variable selection, assessing the importance of variables, handling missing values, prediction, and data manipulation. The process of building DT involves selecting input variables, splitting nodes, determining stopping criteria, and pruning to improve model accuracy. By segmenting data based on attributes highly associated with churn, it provides valuable insights into customer behavior and supports effective retention strategies (Song & Lu, 2015).

In these algorithms, the following hyperparameters were used: The maximum number of leaves per tree was 20, the minimum number of samples per leaf node was 10, the learning rate was 0.2, and 100 trees were constructed.

B. Random Forest (RF)

RF it is an ensemble learning technique applied to regression and classification. It creates a large number of DT from a random subset of data using the bagging technique. The final DT are created by combining the output of each DT in the RF. The RF Algorithm consists of two stages: the creation of RF and the prediction using the RF classifier that was established in the first stage (Alzubi et al., 2018).

In these algorithms, the following hyperparameters were used: The number of decision trees was 8, with a maximum depth of 32 for each tree. Each node was split using 128 random splits, and there was a minimum of 1 sample required per leaf node.

C. Logistic Regression (LR)

LR models are statistical models that depict the connection between one or more independent variables and a categorical dependent variable “which can only have certain discrete values”. They are frequently used to examine how predictor variables on categorical outcomes, which are frequently binary results like the presence or absence of something. A logistic function is used in LR to evaluate the odds of various events (Model, n.d.).

In these algorithms, the following hyperparameters were used: The L1 regularization weight was set to 1, the L2 regularization weight was also set to 1, and the memory size for L-BFGS was 20.

D. Support Vector Machines (SVM)

SVM are efficient algorithms used in data mining and machine learning for tasks such as classification and regression. By maximizing data margins, SVM effectively separate different categories of data (Bäck et al., 2012).

In these algorithms, the following hyperparameters were used: The number of iterations was set to 1, and the lambda value was 0.001.

E. Neural Network (NN)

NN algorithm is a machine learning technique. NN were initially developed with the aim of mimicking the functioning of the human nervous system to tackle various machine learning tasks. They achieve this by modeling computational units in a way that resembles human neurons. The overarching goal behind neural networks is to advance artificial intelligence by constructing machines whose architecture replicates the computational processes observed in the human nervous system (Kizilkan et al., 2022).

In these algorithms, the following hyperparameters were used: The number of hidden nodes was 100, the learning rate was 0.1, there were 100 learning iterations, the initial learning weights diameter was 0.1, and the momentum was 0.

F. Naïve Bayes (NB)

The NB algorithm is a machine learning technique used for classification tasks, particularly in natural language processing and text classification. It relies on Bayes' theorem to calculate the conditional probability of different classes given independent features. Despite its simplicity, NB can perform well in practical applications when the independence assumption is approximately true or when combined with other techniques (Chen et al., 2020).

4. Model Performance

In the model evaluation phase of our research, we assess the performance of models. Since our research employs supervised learning with a classification approach, we used the confusion matrix as the primary method for estimating accuracy.

A. Confusion matrix

Confusion matrix serves as the fundamental tool for evaluating accuracy. It serves as a succinct summary table detailing both correct and incorrect predictions generated by a classifier or classification model, particularly for binary classification tasks. This N x N matrix is instrumental in evaluating the performance of a classification model, where N represents the number of target classes. Visualizing the confusion matrix enables individuals to gauge the accuracy of the model, as diagonal values indicate the number of accurate classifications. Structured as a square matrix, with actual values represented in columns and predicted values in rows, or vice versa, the confusion matrix provides a comprehensive assessment of a model's predictive prowess, aiding in the iterative refinement of machine learning algorithms for real-world applications (Visa et al., 2011; Ting, 2017).

TP "True Positive": The model correctly predicted a positive value when the actual value was positive.

FP "False Positive": The model predicted a positive value incorrectly when the actual value was negative.

FN "False Negative": The model predicted a negative value incorrectly when the actual value was positive.

TN "True Negative": The model correctly predicted a negative value when the actual value was negative.

Table 4: Confusion Matrix for which algorithm

		Actual	
		YES	**NO**
Predicted	**YES**	**TP** 1978	**FP** 64
	NO	**FN** 13	**TN** 278

B. Accuracy

Accuracy is a metric used to measure the model's capability to predict the output correctly (class label) correctly by calculating the ratio between the numbers of correct predictions divided by the total number of predictions. A high accuracy indicates that the model makes correct predictions overall. Accuracy (Hossin & Sulaiman, 2015; Lozano et al., 2017).

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

C. Recall

Recall is a metric used to measure how many were correctly predicted positive by the model out of all the positive instances in the dataset. It's calculated by dividing the true positive to the sum of true positive and false negative. A high recall indicates that the model is most likely will not miss positive outcomes and is more sensitive to identifying them (Green, 2018; Dridi et al., 2021).

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

D. Precision

Precision is a metric used to measure how many true positive predictions made by the model are true in the dataset it represents the accuracy of the positive prediction made by the model. It is calculated by dividing the true positive by all the positive predictions made by the model. A higher precision means that the model makes accurate predictions (Green, 2018; Dridi et al., 2021)

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

E. F1-score

The F1-score is a metric that combines Precision and Recall into a single measure to assess the effectiveness of a prediction model. Since it combines both precision and recall into one value this provides a balanced measure of the model accuracy also f1 score shows the robustness of the prediction model (Green, 2018; Dridi et al., 2021).

$$\mathrm{F1} = 2 \times \frac{1}{\frac{1}{\mathrm{Precision}}+\frac{1}{\mathrm{Recall}}} \quad (4)$$

4.1 Tune Model Hypermeters

Tuning is considered an essential step in the machine learning workflow to improve model performance. This entails determining model specific parameters, such as the number of layers in a neural network or the number of trees in a Random Forest, as well as important hyperparameters like learning rate, batch size, and number of epochs. Model's accuracy, efficiency, and capability for generalization can all be greatly increased through proper tuning (Joy et al., 2016; Bardenet et al., 2013).

4.2 Evaluation

Our predictive model was developed for each algorithm, and then we evaluated it using a confusion matrix and other metrics. AS shown in table 5.

Before using tune model hyperparameters**:**

Table 5: Model evaluation using 70:30 as splitting

Algorithm	TP	FN	FP	TN	Accuracy	Precision	Recall	F- Score
DT	811	48	59	82	0.893	0.932	0.944	0.938
RF	817	42	62	79	0.896	0.929	0.951	0.940
LR	841	18	125	16	0.857	0.871	0.979	0.922
SVM	831	28	123	18	0.849	0.871	0.969	0.917
NN	**838**	**21**	**81**	**60**	**0.898**	**0.912**	**0.976**	**0.943**
NB	836	23	120	21	0.857	0.874	0.973	0.921

In table 5, the performance of the algorithms is ranked according to their performance from best to worst. the NN algorithm was the highest overall performance with an Accuracy of 0.898, Precision of 0.912, Recall of 0.976, and F-Score of 0.943. Following closely is the RF algorithm with an Accuracy of 0.896, Precision of 0.929, Recall of 0.951, and F-Score of 0.940. DT, achieving an Accuracy of 0.893, Precision of 0.932, Recall of 0.944, and F-Score of 0.938. While LR, SVM and NB algorithms display decent performance, they fall slightly short in accuracy compared to others algorithm. On another hand the performance of algorithms after using tune model hyperparameters shown in table 6.

Table 6: Model evaluation using tune model hyperparameters

Algorithm	TP	FN	FP	TN	Accuracy	Precision	Recall	F- Score
DT	**1978**	**13**	**64**	**278**	**0.967**	**0.969**	**0.993**	**0.981**
RF	1979	15	99	243	0.951	0.952	0.992	0.972
LR	1953	38	289	53	0.860	0.871	0.981	0.923
SVM	1949	42	287	55	0.859	0.872	0.979	0.922
NN	1956	35	224	118	0.889	0.897	0.982	0.938
NB	1942	49	287	55	0.856	0.871	0.975	0.920

After using hyperparameter tuning as shown in Table 6, the performance improved for all algorithms. The best algorithm was the decision tree, and its performance improved after tuning with %96.7 accuracy, %96.9 precision, %99.3 recall, and %98.1 F-score. Random forest came second with %95.1 accuracy, %95.2 precision, %99.2 recall and %97.2 F-score. Also Logistic Regression and SVM showed improvement in performance, but for neural network and naïve bayes performance slightly decreased in terms of accuracy, precision, recall, and F-Score.

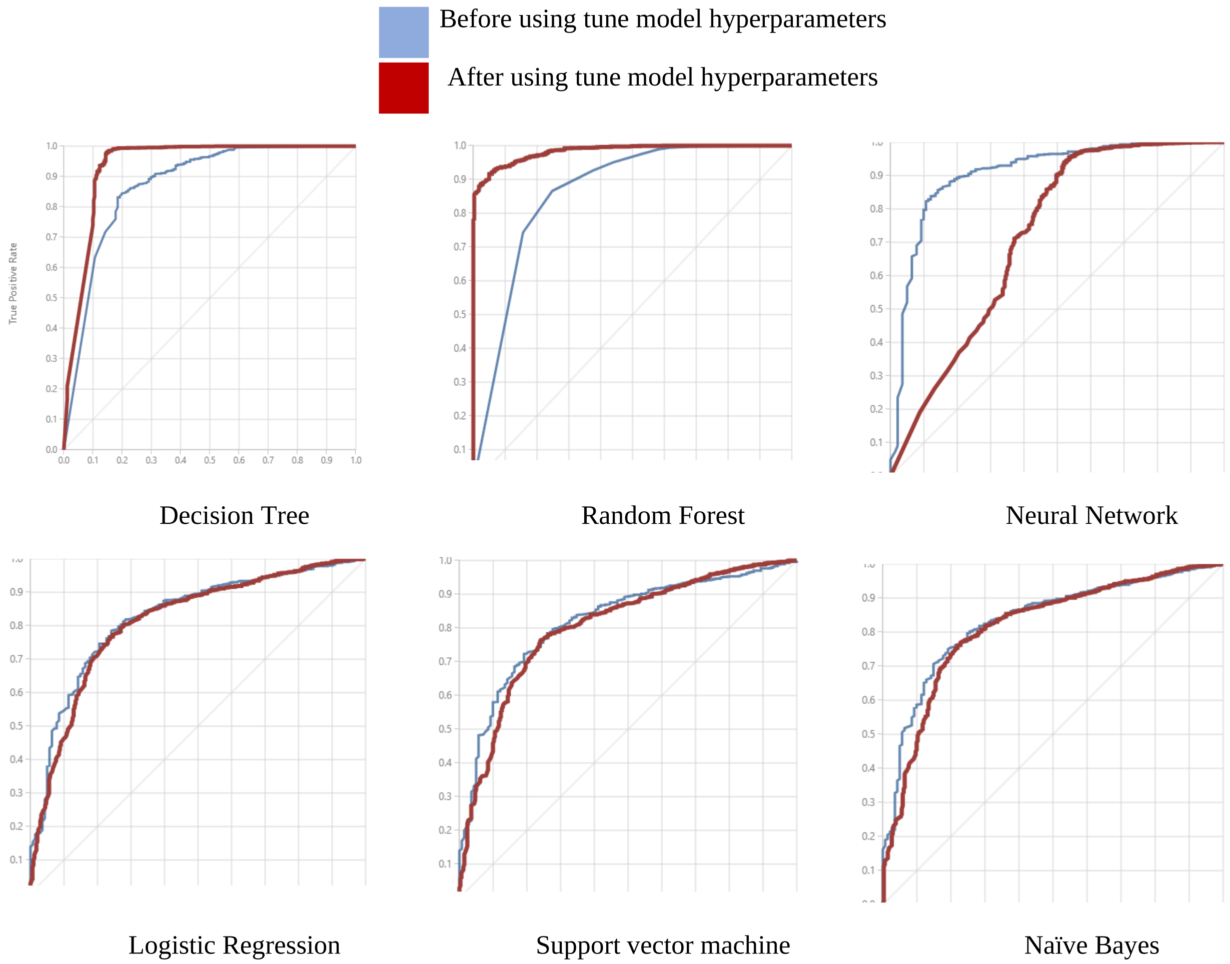

Figure 5. Model evaluation before and after tuning model hyperparameters for all algorithms.

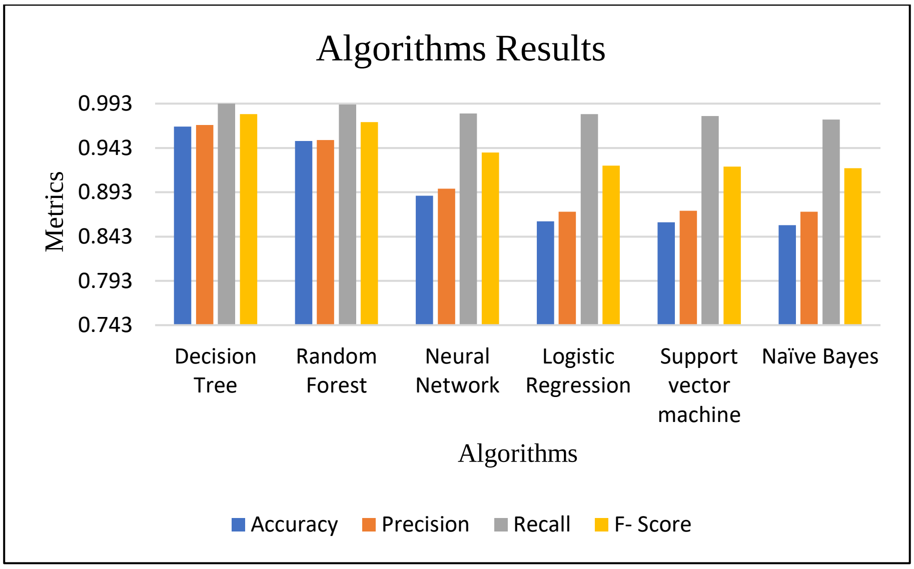

Figure 6. Algorithms Results

The performance of DT and RF algorithms was exceptional compared to the others, while NN, LR, NB, and SVM algorithms exhibited decent performance, though not surpassing the best algorithms in terms of accuracy and other metrics. After comparing DT and RF algorithms, DT was chosen as the optimal choice for application to the model due to its achievement of higher proportions in all metrics including Accuracy, Precision, Recall, and F-Score, as demonstrated in the presented results. This indicates its capability to achieve accurate predictions of the churn rate. the best performance for the DT algorithm as shown in table 7:

Table 7: Decision Tree Hypermeters

Number of leaves	Minimum leaf instances	Learning rate	Number of trees	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
19	1	0.042	86	0.913	0.940	0.959	0.950	0.894	0.238	42.877

5. Deployment

Model Deployment in machine learning is the process of converting the built model into a production environment where it can be used by users to make effective and useful predictions. Ultimately, the main goal of deploying the model is to make it accessible to users for performing meaningful tasks (Kakarla et al., 2021; Ackermann et al., 2018).

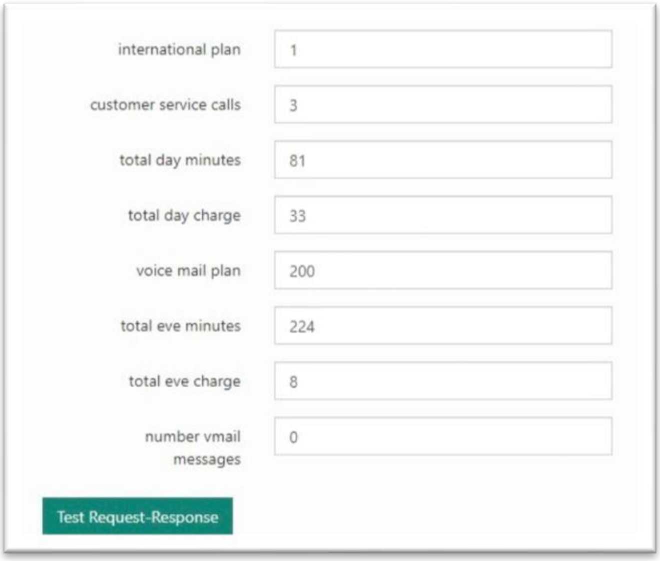

Figure 7. Input values for the second predictive experiment

Based on the input values, the churn rate was predicted:

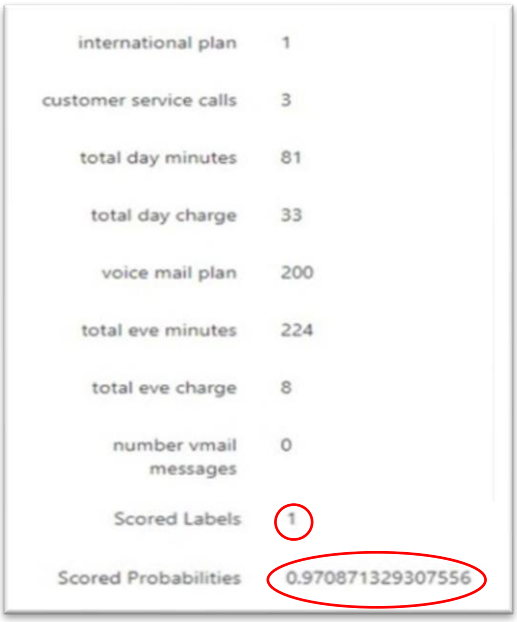

Figure 8. Output values for the second predictive experiment

After inputting values for a random customer to predict whether they will churn or not, the result was 1, indicating that the customer will not churn. However, the probability of non-churn for this customer is 97%. This suggests a high likelihood of the customer staying with the service.

6. Conclusion

The primary objective of this study is to identify the optimal model for predicting customer churn in the telecommunications sector using machine learning techniques. Given that acquiring new customers is significantly more expensive—5 to 6 times higher—companies prioritize retaining existing customers. The rise in customer churn rates, particularly from 20% to 40%, is a concerning trend for telecommunications firms. The dataset used in this research is sourced from Kaggle and includes corporate customer data. It comprises 21 columns and 3333 records, specifically designed for corporate customer analytics. The initial phase of the study involved conducting descriptive data analysis as part of the data preparation process and in the second part, the important variables have been identified, and the target variable is churn. The primary criteria used in this research were total day minutes, international plan, customer service calls, total day charge, voice mail plan, total eve minutes, total eve charge, and amount of vmail messages.

A customer churn model is based on ML methods such as DT, RF, LR, SVM, neural networks, and the NB algorithm. A comparison was made between the algorithms after tuning the model hyperparameters. The decision tree classification models produced better results compared to other algorithms, with 96.7% accuracy, 96.9% precision, 99.3% recall, and a 98.1% F-score. Random forest came in second with %95.1 accuracy, %95.2 precision, %99.2 recall, and %97.2 F-score. Also, logistic regression and SVM showed improvement in performance, but for neural networks and naïve bayes, performance slightly decreased in terms of accuracy, precision, recall, and F-Score. The decision tree algorithm was chosen as the most suitable algorithm for customer churn because it achieved the highest scores in all metrics: "accuracy, precision, recall, and F-score”. Two predictive experiments were conducted and explained in the deployment.

References

Abdallah, Z. S., and Webb, G. (2017). Encyclopedia of Machine Learning and Data Mining. Encycl. Mach. Learn. Data Min., no. September 2018. doi: 10.1007/978-1-4899-7687-1.

Ackermann, K., Walsh, J., De Unánue, A., Naveed, H., Navarrete Rivera, A., Lee, S. J., ... & Ghani, R. (2018, July). Deploying machine learning models for public policy: A framework. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 15-22).

Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning and social network analysis in big data platform. *arXiv preprint arXiv:1904.00690*.

Alwis, P. K. D. N. M., Kumara, B. T. G. S., & Hapuarachchi, H. A. C. S. (2018). Customer Churn Analysis and Prediction in Telecommunication for Decision Making. International Conference On Business Innovation (ICOBI), 25-26 August 2018, NSBM, Colombo, Sri Lanka.

Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine learning from theory to algorithms: an overview. In *Journal of physics: conference series* (Vol. 1142, p. 012012). IOP Publishing.

Bäck, T., Kok, J. N., & Rozenberg, G. (2012). *Handbook of natural computing*. Springer, Heidelberg.

Bardenet, R., Brendel, M., Kégl, B., & Sebag, M. (2013, May). Collaborative hyperparameter tuning. In *International conference on machine learning* (pp. 199-207). PMLR.

Bilişik, Ö. N., & Sarp, D. T. (2023). Analysis of Customer Churn in Telecommunication Industry with Machine Learning Methods. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *11*(4), 2185-2208.

Brandusoiu, I., & Toderean, G. (2013). Churn prediction modeling in mobile telecommunications industry using decision trees. *Journal of Computer Science and Control Systems*, *6*(1), 14.

Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems*, *192*, 105361.

Ebrah, K., & Elnasir, S. (2019). Churn prediction using machine learning and recommendations plans for telecoms. *Journal of Computer and Communications*, *7*(11), 33-53.

Dahouda, M. K., & Joe, I. (2021). A deep-learned embedding technique for categorical features encoding. *IEEE Access*, *9*, 114381-114391.

Dridi, S. (2021). Supervised learning-a systematic literature review. *preprint, Dec*.

Green, F.. (2018). *Business intelligence and data analytics*. doi: 10.4324/9781315108537-4.

Gürsoy, U. Ş. (2010). Customer churn analysis in telecommunication sector. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, *39*(1), 35-49.

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, *5*(2), 1.

Hu, L., Chen, J., Vaughan, J., Aramideh, S., Yang, H., Wang, K., ... & Nair, V. N. (2021). Supervised machine learning techniques: An overview with applications to banking. *International Statistical Review*, *89*(3), 573-604.

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *15*(4), 531-538.

Joy, T. T., Rana, S., Gupta, S., & Venkatesh, S. (2016, December). Hyperparameter tuning for big data using Bayesian optimisation. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 2574-2579). IEEE.

Kakarla, R., Krishnan, S., Alla, S., Kakarla, R., Krishnan, S., & Alla, S. (2021). Deploying Machine Learning Models. *Applied Data Science Using PySpark: Learn the End-to-End Predictive Model-Building Cycle*, 361-388.

Kavitha, V., Kumar, G. H., Kumar, S. M., & Harish, M. (2020). Churn prediction of customer in telecom industry using machine learning algorithms. *International Journal of Engineering Research & Technology (2278-0181)*, *9*(05), 181-184.

Khalid, S., Khalil, T., & Nasreen, S. (2014, August). A survey of feature selection and feature extraction techniques in machine learning. In *2014 science and information conference* (pp. 372-378). IEEE.

Kisioglu, P., & Topcu, Y. I. (2011). Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey. *Expert Systems with Applications*, *38*(6), 7151-7157.

Kizilkan, Z. B., Sivri, M. S., Yazici, I., & Beyca, O. F. (2022). Neural Networks and Deep Learning. In *Business Analytics for Professionals* (pp. 127-151). Cham: Springer International Publishing.

Konieczny, R., & Idczak, R. (2016). Mössbauer study of Fe-Re alloys prepared by mechanical alloying. *Hyperfine Interactions*, *237*, 1-8.

Krause, R. W., Huisman, M., Steglich, C., & Snijders, T. (2020). Missing data in cross-sectional networks–An extensive comparison of missing data treatment methods. *Social Networks*, *62*, 99-112.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and*

Research (IJSR).[Internet], *9*(1), 381-386.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., ... & Flach, P. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE transactions on knowledge and data engineering*, *33*(8), 3048-3061.

Model, L. R. (n.d). "Logistic Regression", doi: 10.1007/978-1-4614-7792-1.

Mohamed, A. E. (2017). Comparative study of four supervised machine learning techniques for classification. *International Journal of Applied*, *7*(2), 1-15.

Lozano, J. A., & Santafe, G., Inza, I. (2015). Dealing with the evaluation of supervised classification algorithms. *Artificial Intelligence Review*, *44*, 467-508.

Sattler, K. U., & Schallehn, E. (2001, July). A data preparation framework based on a multidatabase language. In *Proceedings 2001 International Database Engineering and Applications Symposium* (pp. 219-228). IEEE.

Seyed, H., Iranmanesh, M., Hamid, M., Bastan, H., Shakouri, G., and Nasiri, M .M., (2019, July). Customer churn prediction using artificial neural network: An analytical CRM application. In *Proceedings of the International Conference on Industrial Engineering and Operations Management, Pilsen, Czech Republic* (pp. 23-26).

Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), 130-135.

Soofi, A. A., & Awan, A. (2017). Classification techniques in machine learning: applications and issues. *Journal of Basic & Applied Sciences*, *13*, 459-465.

Sullivan, E.,"Understanding from machine learning models," *Br. J. Philos. Sci.*, vol. 73, no. 1, pp. 109–133, 2022, doi: 10.1093/bjps/axz035.

Ting, K. M. (2011). Confusion matrix. *Encyclopedia of machine learning*, 209-209.

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*, 60134-60149.

Umayaparvathi, V., & Iyakutti, K. (2016, March). Attribute selection and customer churn prediction in telecom industry. In *2016 international conference on data mining and advanced computing (sapience)* (pp. 84-90). IEEE.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1-9.

Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. *Maics*, *710*(1), 120-127.

Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*, *14*, 100342.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-39).