BACS2003/BMCS2003/BACS3074

(Remain only your course code before submission)

ARTIFICIAL INTELLIGENCE

202401 Session, Year 2023/24

Assignment Documentation

Project Title: ***Customer Churn Prediction using machine learning***	
Programme: ***RDS Y2S2***	
Tutorial Group: ***1***	
Tutor: ***Mr Ho***	

Team members' data

No	Student Name	Student ID	Module In Charge
1	Chai Zhi Xiang	22WMR13533	Support Vector Machines/ K-Nearest Neighbour/ Neural Networks
2	Jason Yap	22WMR13564	Random Forest / Decision Tree / Naive Bayes

3	Choong Zhi Yang	22WMR13552	XGBoost/ Logistic Regression/ ADA Boost

1. Introduction

1.1. Problem Background

The telecommunications sector has become one of the main industries in developed countries including Malaysia. The telecommunications market has grown exponentially with the emergence of advanced technologies and widespread use of mobile devices, driving intense competition among service providers. In this rapidly growing and highly competitive market, maintaining consumers has become an important challenge for telecommunication companies. This is because when the number of operators and consumer demands is increasing, telecommunications companies face the ongoing risk of customer churn (Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K., 2017). Customer churn is a term that refers to the rate of customers leaving the business or the services. Churn could be due to various factors, such as switching to a competitor company, canceling their existing subscription or services because of poor customer service (Mustafa, N., Sook Ling, L. & Abdul Razak, S. F., 2021). Customer churn problem not only reduces the revenue of a company but also will damage the brand reputation and market position. The impact of customer churn is more than just financial, as churn requires more investment in customer acquisition efforts, which will increase the operating costs and reduce the profitability as losing of customers will increase the need for attracting the new customers and this will cost five to six times more expensive than customer retention (Van Den Poel, D., & Larivière, B., 2004).

In addition, the rapid growth in technology has totally altered customer needs and behaviors within the telecommunications landscape. With a wide range of choices at their fingertips, customers enjoy flexibility in selecting service providers. Therefore, telecom companies must constantly improve and differentiate their offerings to retain customers among higher rivals. At the same time, changing customer preferences driven by economic and social changes, and cultural factors require telecommunications companies to adjust rapidly (Kumar, A., Shankar, R., & Debnath, R. M., 2015). Failure to adjust with changed needs of customers puts at risk of customer dissatisfaction and increased churn rates.

Next, telecom sectors could reach a saturation point in developed nations such as Malaysia, where most potential customers are already subscribers (Bhd, S. P. H. S., 2015). In these highly competitive environments, attracting new clients is becoming more difficult and expensive. As a result, in order to achieve continued revenue growth and maintain market share, telecom companies need to redirect their efforts toward retaining current customers.

Given these challenges, proactive churn identification and targeted retention strategies are essential. Telecommunications companies can gain insights into consumer patterns, estimate churn risks, and take early retention steps through using data analytics and predictive modeling. By doing this, telecom companies can maintain long-term profitability in the face of a constantly changing competitive environment while also increasing customer happiness and brand loyalty.

1.2. Objectives/Aims

The primary objective of this project is to develop an accurate churn prediction application that is particularly suitable for the telecommunications industry. Our aim is to provide telecom companies with an efficient tool that can accurately forecast cases of customer churn, allowing early action to reduce the risk of losing important customers. By using historical customer data and predictive analytics techniques, we want to identify hidden trends and signals that may lead to customer churn. This early detection ability is essential for telecom companies as it allows them to carry out immediate and specific actions to retain at-risk customers before they decide to switch to competitors or cancel their subscriptions.

Furthermore, we aim to use analytics techniques to extract useful information from the identified patterns and indicators. We try to find the hidden correlations and trends in the data that might give telecommunications companies important strategic direction. These insights could include usage patterns, service preferences and other important factors that affect customer behavior. With the knowledge gained from these insights, telecommunications companies will be able to better retain and attract at-risk customers, which will decrease churn rates and increase customer lifetime value.

In essence, our objective is not only to develop a churn prediction application but also to provide telecom companies with a strategic toolset for effectively managing customer churn and fostering long-term customer loyalty.

1.3. Motivation

The motivation behind this project stems from the critical importance of customer retention in the telecommunications industry. As telecom companies strive to differentiate themselves in a crowded market, retaining existing customers emerges as a strategic imperative for maintaining competitiveness and driving sustainable growth.

Moreover, customer retention initiatives are inherently more cost-effective than customer acquisition efforts, offering a higher return on investment and bolstering long-term profitability (E Pfeifer, P., 2005). By proactively identifying and addressing churn risks, telecom companies can optimize resource allocation, minimize revenue leakage, and maximize customer lifetime value.

Furthermore, effective churn prediction not only enhances financial performance but also enhances brand loyalty and customer satisfaction. By delivering personalized experiences and value-added services for the at-risk customer, telecom companies can build deeper connections with their customer base, driving advocacy and loyalty in an increasingly competitive landscape.

In conclusion, developing a robust customer churn prediction model and integrating it with a customer lifecycle strategy provides strategic opportunities for telecommunications companies to gain a competitive advantage, drive sustainable growth, and build lasting relationships with customers.

1.4. Timeline/Milestone

Week 3	Finding dataset
Week 4	Research existing work
Week 5	EDA
Week 6	
Week 7	Preprocessing
Week 8	Machine Learning Evaluation
Week 9	

Week 10	Deployment and Report
Week 11	

2. Research Background

2.1. Background of the applications

2.2.1 Trend Customer Churn

In today's highly competitive business environment, customer churn has become a major concern for companies across various industries, including telecommunications, banking, and e-commerce (Li & Zhou, 2020). Customer churn refers to the phenomenon of customers ceasing their relationship with a company, either by cancelling a subscription or switching to a competitor. This phenomenon poses significant challenges for businesses, as it not only leads to a loss of revenue but also increases the cost of acquiring new customers to replace those that have churned (A. P. Patil, M. P. Deepshika, S. Mittal, S. Shetty, S. S. Hiremath and Y. E. Patil, 2017). Therefore, the identification and prediction of customer churn have become crucial for businesses in order to retain existing customers and reduce the churn rate.

2.2.2 Techniques

1. Imbalance Dataset Handling

   Handling an imbalanced dataset is important when doing model training in machine learning. This is due to the reason that an imbalanced dataset will lead to a huge negative impact on performance of model training (Kumar, P., Bhatnagar, R., Gaur, K., & Bhatnagar, A., 2021). The imbalance between majority and minority would lead to a bias model and produce inaccurate results if the imbalance dataset is not resolved (Goel, G., Maguire, L., Li, Y., & McLoone, S., 2013). In order to deal with the imbalance dataset, there are two methods that can be used which are oversampling and undersampling (Ahmad, A. K., Jafar, A., & Aljoumaa, K., 2019). The techniques used based on the research are Synthetic Minority Over-sampling Technique (SMOTE) and random undersampling with a result that the oversampling and undersampling method give a similar result in terms of area under curve (AUC) (Shumaly S., Neysaryan P. and Guo Y., 2020). There is another research using the same sampling techniques but the result states that the random undersampling worsened the accuracy (Kimura, T., 2022). Therefore, experiment for each different technique used is needed to find out which techniques may have better accuracy.

2. Feature Scaling

   The research paper shows the performance of the techniques used which are min max scaler, standard scaler and normalization with the result that min max scaler and standard scaler have a similar performance but normalization is the worst among all (Chauhan, N. K., & Singh, K., 2022). Moreover, there is another research that compare the feature scaling methods with Standard Scaler, Min-max Scaler, Maximum Absolute Scaler, Robust Scaler and Quantile Transformer, the result of this research is Standard Scaler have the best performance and the non scaled data have the worst performance (Amorim, L., Cavalcanti, G. D. C., & Cruz, R. M. O., 2023). Based on the research the methods that can be used are Normalizer, Robust Scaler, Min-Max Scaler and Standard Scaler with the conclusion that Robust scaling has better performance because it is not sensitive to outliers (Thakker, Z. L., & Buch, S. H., 2024).

2.2.3 Machine Learning Algorithm

The models that used to make comparison on customer churn prediction are Artificial Neural Networks (ANN), Decision Tree, Support Vector Machine (SVM), Naive Bayes, Logistic Regression, k-Nearest Neighbor (k-NN), Extreme Gradient Boosting (XGBoost), Cox Regression and Deep Learning Technique with the result deep learning techniques perform better results in more complex structures (Celik, O., & Osmanoglu, U. O., 2019). Additionally, there are another research compare the models with Logistic Regression, Naive Bayes, Support Vector Machine, Decision Tree, Random Forest, Extra Tree Classifier, Adaboost, XGBoost and Catboost. The performance of XGBoost and Adaboost have better accuracy compared to other models with an AUC score of 84% for the churn prediction (Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P., 2021). Besides, there is a research paper comparing Artificial Neural Network, Support Vector Machines, Decision trees, Naïve Bayes, Logistic Regression and Boosting algorithms. The conclusion of the research state that there are two top model perform well which are two-layer Back-Propagation Network with 15 hidden units and the Decision Tree classifier with 94% accuracy and 77% F measure, by boosting the algorithms except Naïve Bayesa and Logistic Regression the performance of the models improved and the best classifier is the boosted SVM (SVM-POLY with AdaBoost) with accuracy of almost 97% and F-measure over 84% (Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C., 2015).

2.2. Analysis of selected tool with any other relevant tools

Tools comparison	**Remark**	**numpy**	**pandas**	**matplotlib**	**seaborn**	**imblearn**	**sklearn**	**scipy**

Type of license and open source license	State all types of license	BSD-3-Clause (Open Source)	BSD-3-Clause (Open Source)	Modified BSD (Open Source)	BSD-3-Clause (Open Source)	BSD-3-Clause (Open Source)	BSD-3-Clause (Open Source)	BSD-3-Clause (Open Source)
Year founded	When is this tool being introduced?	2005	2008	2003	2012	2014	2007	2001
Founding company	Owner	Travis Oliphant (Core Developer)	Wes McKinney	John Hunter (Core Developer)	Michael Waskom (Lead Developer)	Fernando Nogueira	David Cournapeau (Core Developer)	Enthought (Original Sponsor)
License Pricing	Compare the prices if the license is used for development and business/commercialization	Free for both Development and Commercial Use	Free for both Development and Commercial Use	Free for both Development and Commercial Use	Free for both Development and Commercial Use	Free for both Development and Commercial Use	Free for both Development and Commercial Use	Free for both Development and Commercial Use
Supported features	What features that it offers?	Numerical computations, array manipulation, linear algebra	Data structures (Series, DataFrame), data cleaning, time series analysis	Visualization (static, animated, interactive), plotting various data types	Statistical data visualization built on top of matplotlib	Techniques for imbalanced datasets (oversampling, undersampling)	Machine learning algorithms, model selection, preprocessing	Scientific computing algorithms, optimization, integration
Common applications	In what areas this tool is usually used?	Scientific computing, machine learning, data analysis	Data analysis, finance, econometrics	Scientific computing, data exploration, presentations	Machine learning, data exploration, statistical graphics	Machine learning, handling imbalanced datasets	Machine learning, data mining, classification, regression	Scientific computing, engineering, signal processing
Customer support	How the customer support is given, e.g. proprietary, online community, etc.	Active online community, extensive documentation	Active online community, extensive documentation	Active online community, extensive documentation	Active online community, leverages matplotlib support	Active online community, documentation	Active online community, extensive documentation	Active online community, leverages NumPy support
Limitations	The drawbacks of the software	Can be memory intensive for very large datasets	Less efficient for very large datasets compared to specialized tools	Limited customization compared to some commercial plotting libraries	Relies on matplotlib for underlying functionality	Limited to imbalanced learning problems	May require coding expertise for advanced use cases	Can be complex for beginners due to broad functionality

2.3. Justify why the selected tool is suitable

NumPy provides the foundation for numerical operations with its efficient handling of large arrays, which is essential when processing and analyzing customer data. Pandas complements this by offering robust data structures for indexing, slicing, and reshaping large datasets, making it incredibly useful for data preparation tasks such as feature selection and data cleaning in churn analysis.

Matplotlib and Seaborn are invaluable for data visualization, allowing for the exploration of data and the understanding of key patterns and relationships that could influence churn. Matplotlib offers granular control over every aspect of plotting, while Seaborn provides a more streamlined, high-level interface for generating more complex statistical visualizations with ease, aiding in the intuitive representation of customer behaviors and characteristics.

Scikit-learn and imblearn specifically cater to the machine learning aspect of churn prediction. Scikit-learn provides a vast array of machine learning algorithms including regression, clustering, and classification which can be applied to predict customer churn. It also offers tools for model evaluation and selection, which help in fine-tuning model performance. Imblearn is particularly crucial when dealing with imbalanced datasets, a common issue in churn prediction, where the number of churned customers is typically much smaller than those who stay. It offers various techniques to handle this imbalance, ensuring that the model accurately predicts churn without bias.

Lastly, SciPy can be used for additional statistical testing and optimization tasks that may arise during the analysis, making sure that the interpretations and conclusions are based on solid statistical evidence. Together, these tools provide a comprehensive framework not only to build and refine predictive models but also to derive insights that can inform business strategies to reduce churn.

3. Methodology

3.1. Description of dataset

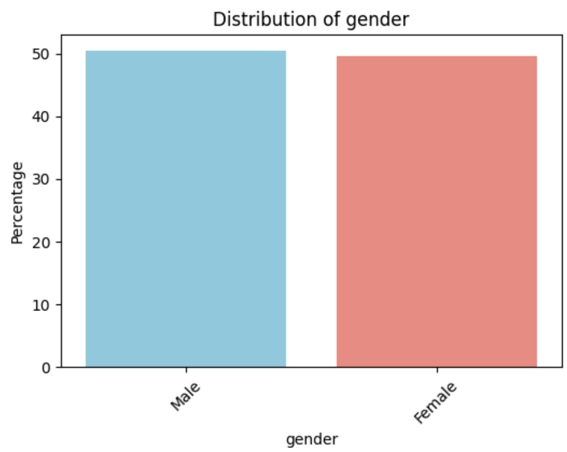
The dataset used for customer churn prediction was sourced from Kaggle and it includes a wide range of customer-related variables. This dataset contains a total of 7043 rows (customers) and 21 columns (features), where the “Churn” column is our target. Each entry is identifiable by a customerID and contains demographic information such as gender, age, and whether they have partners or dependents. Additionally, tenure denotes the length of their employment with the organisation. Customers' subscription services, including phone, internet, and other add-ons such as online security and tech assistance, are mentioned. Contract details such as length and payment method are stored, along with billing settings such as paperless billing. Financial factors are shown as monthly and total charges. Finally, the critical churn column indicates if a consumer has ended their relationship with the organisation. This extensive dataset combines numerous aspects of customer interactions and traits, providing a solid foundation for predicting churn and influencing retention tactics.

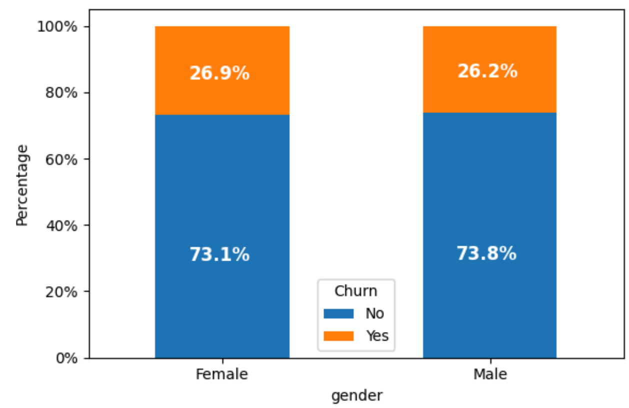
Name	Type	Dependent / Independent variable	Description
customerID	Categorical	Independent	Unique identifier for each customer.
gender	Categorical	Independent	Gender of the customer (e.g., Male, Female).
SeniorCitizen	Categorical	Independent	Whether the customer is a senior citizen or not (0 for No, 1 for Yes).
Partner	Categorical	Independent	Whether the customer has a partner or not (Yes/No).
Dependents	Categorical	Independent	Whether the customer has dependents or not (Yes/No).
tenure	Continuous	Independent	The length of time (in months) that the customer has been with

			the company.
PhoneService	Categorical	Independent	Whether the customer has a phone service or not (Yes/No).
MultipleLines	Categorical	Independent	Whether the customer has multiple lines or not (Yes, No, or No phone service).
InternetService	Categorical	Independent	Type of internet service subscribed by the customer (DSL, Fiber optic, or No).
OnlineSecurity	Categorical	Independent	Whether the customer has online security service or not (Yes, No, or No internet service).
OnlineBackup	Categorical	Independent	Whether the customer has online backup service or not (Yes, No, or No internet service).
DeviceProtection	Categorical	Independent	Whether the customer has device protection service or not (Yes, No, or No internet service).
TechSupport	Categorical	Independent	Whether the customer has tech support service or not (Yes, No, or No internet service).
StreamingTV	Categorical	Independent	Whether the customer has streaming TV service or not (Yes, No, or No internet service).

StreamingMovies	Categorical	Independent	Whether the customer has streaming movies service or not (Yes, No, or No internet service).
Contract	Categorical	Independent	The type of contract the customer has (Month-to-month, One year, Two year).
PaperlessBilling	Categorical	Independent	Whether the customer has opted for paperless billing or not (Yes/No).
PaymentMethod	Categorical	Independent	Payment method used by the customer (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)).
MonthlyCharges	Continuous	Independent	The amount charged to the customer monthly.
TotalCharges	Continuous	Independent	The total amount charged to the customer.
Churn	Categorical	Dependent	Whether the customer churned or not (Yes/No).

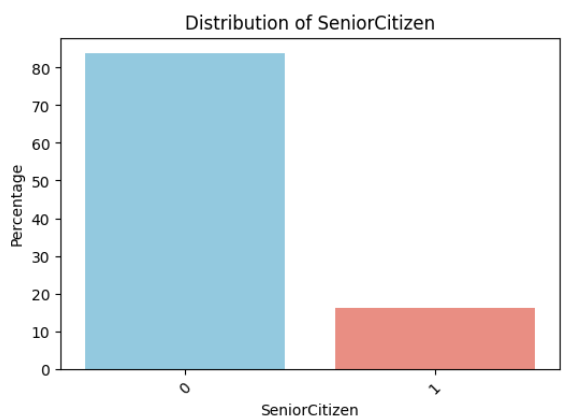
3.1.1 Gender

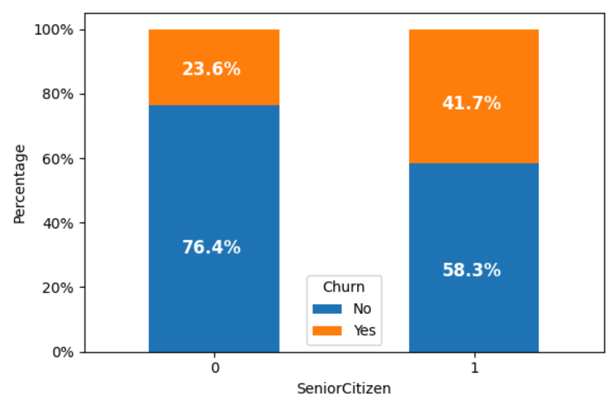




Based on the chart above, we can see that the gender distribution is nearly same, with both males and females possessing roughly 50%. Another chart displaying the churn percentage by gender clearly shows that the churn rate is not significantly different between males and females. Female has a little higher churn rate (26.9%), which is 0.7% greater than male.

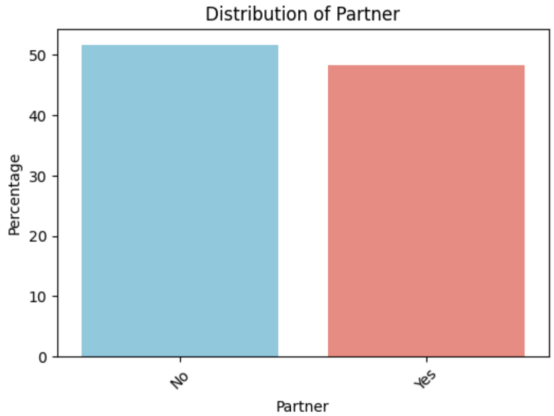
3.1.2 Senior Citizen

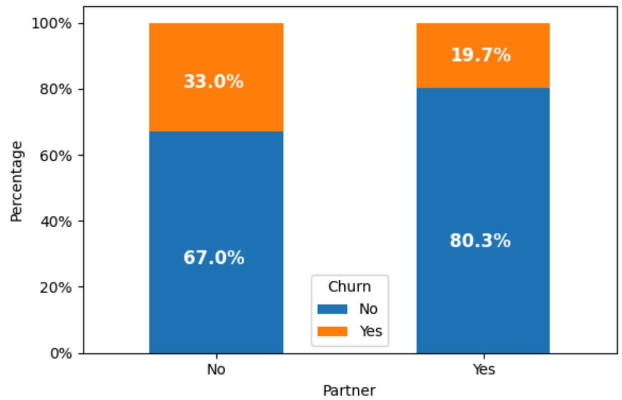




It is evident from the above diagram that just a relatively tiny percentage of the dataset—roughly 15 to 20%—are older citizens. This suggests that younger citizens who require more telecom services make up the majority of a telecom company's client base. Examine the chart; the group of elderly citizens is probably going to have a greater rate of churn than the non-senior citizen category. According to the churn table, the turnover rate for non-senior citizens is just 23.6%, but it is 41.7% for senior citizens.

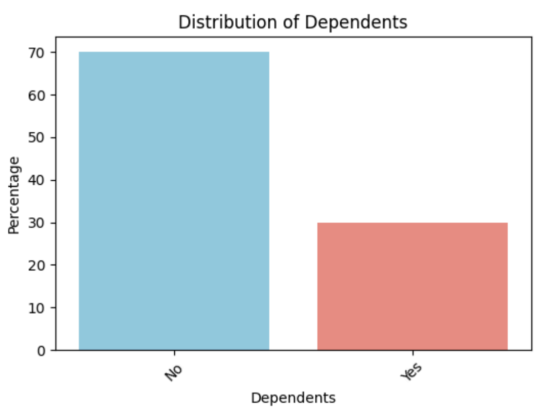
3.1.3 Partner

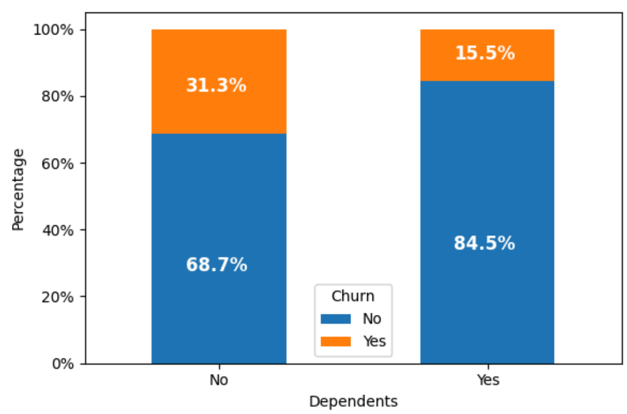




The ratio of having a partner to not having a partner in this dataset is about equal, with the proportion of not having a relationship being slightly higher at 50–55% and that of having a partner at 45–50%. Customers with partners have a lower churn rate (19.7%), whereas customers without partners have a higher churn rate (33%).

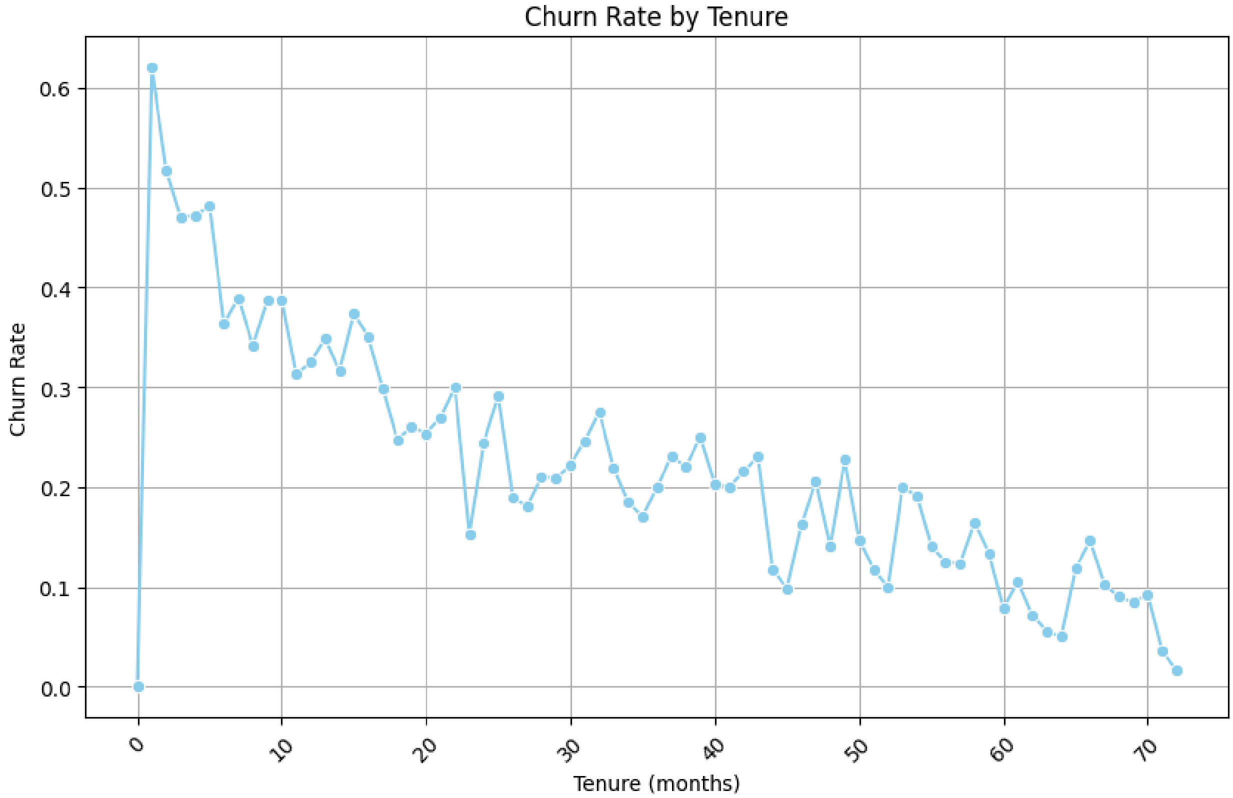
3.1.4 Dependents





Within the dependents group, up to 70% of customers are single, while the remaining 30% are related to someone else. The client without dependents will have a larger likelihood of churning (31.3%), whereas the customer with dependents will only have half the churn rate compared to the customer without dependents (15.5%). This is evident from the churn chart.

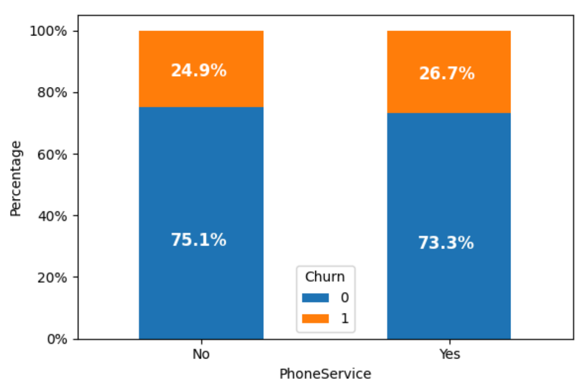
3.1.5 Tenure

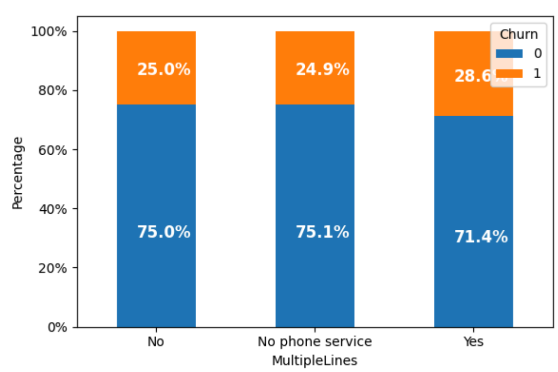


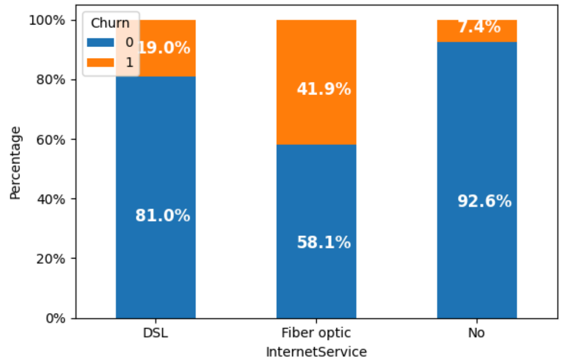
The graph above titled “Churn Rate by Tenure” provides insightful data on customer retention over time. Initially, the churn rate is at its peak, indicating a high loss of customers at the beginning of their tenure. This suggests that the first few months are critical for customer engagement and satisfaction. As customers stay longer, the churn rate decreases significantly, reflecting an increase in customer loyalty. However, the churn rate doesn’t stabilize completely but shows moderate fluctuations throughout the tenure. This could imply that there are ongoing factors affecting customer satisfaction even after the initial period. Businesses can use this information to focus on improving the early customer experience and to investigate the causes of churn throughout the customer lifecycle to enhance retention strategies. The overall trend suggests that the longer customers stay, the less likely they are to leave, highlighting the importance of long-term customer relationship management.

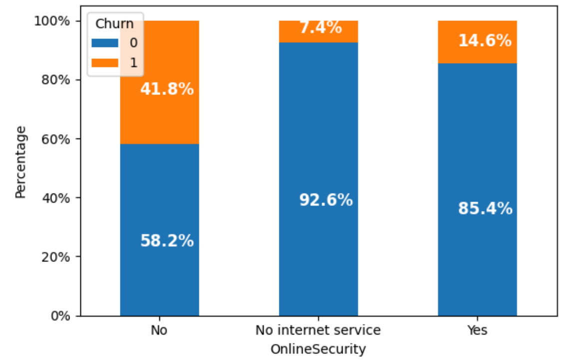
3.1.6 Services

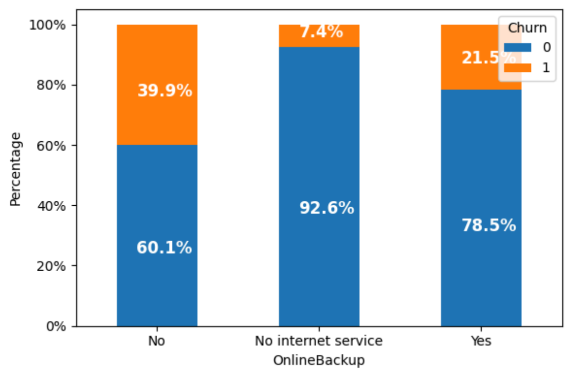
The telecom company offers nine different services, each represented in a donut chart as seen in the above figure. To better comprehend the percentage of customers consuming each service, the donut chart displays the distribution of each service.

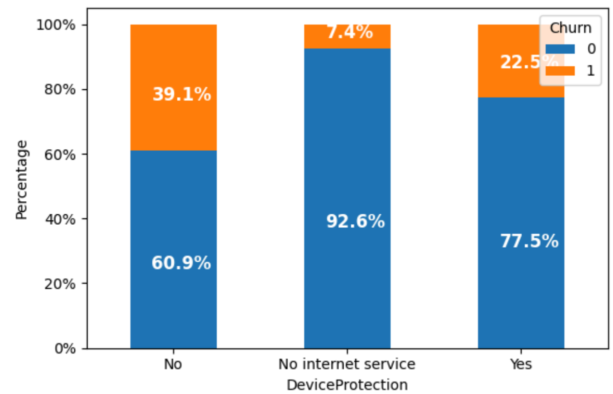


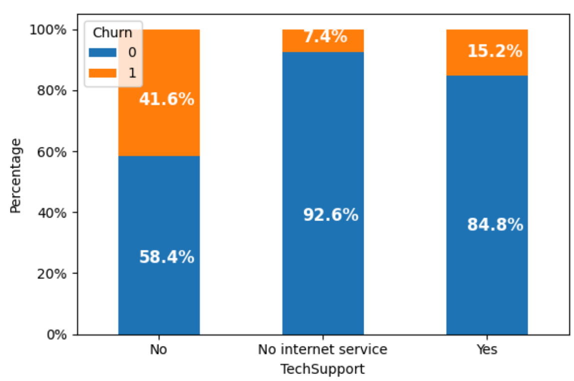


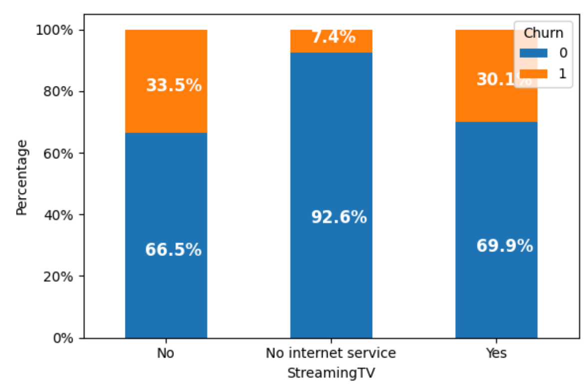


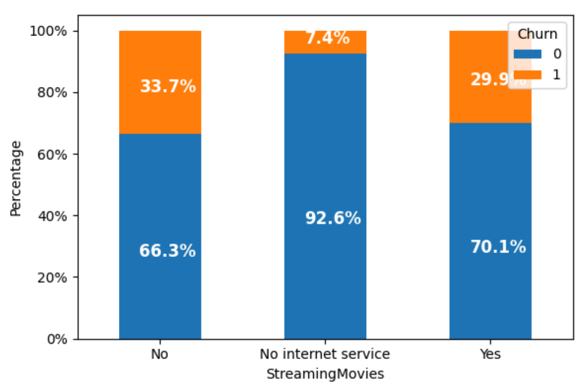






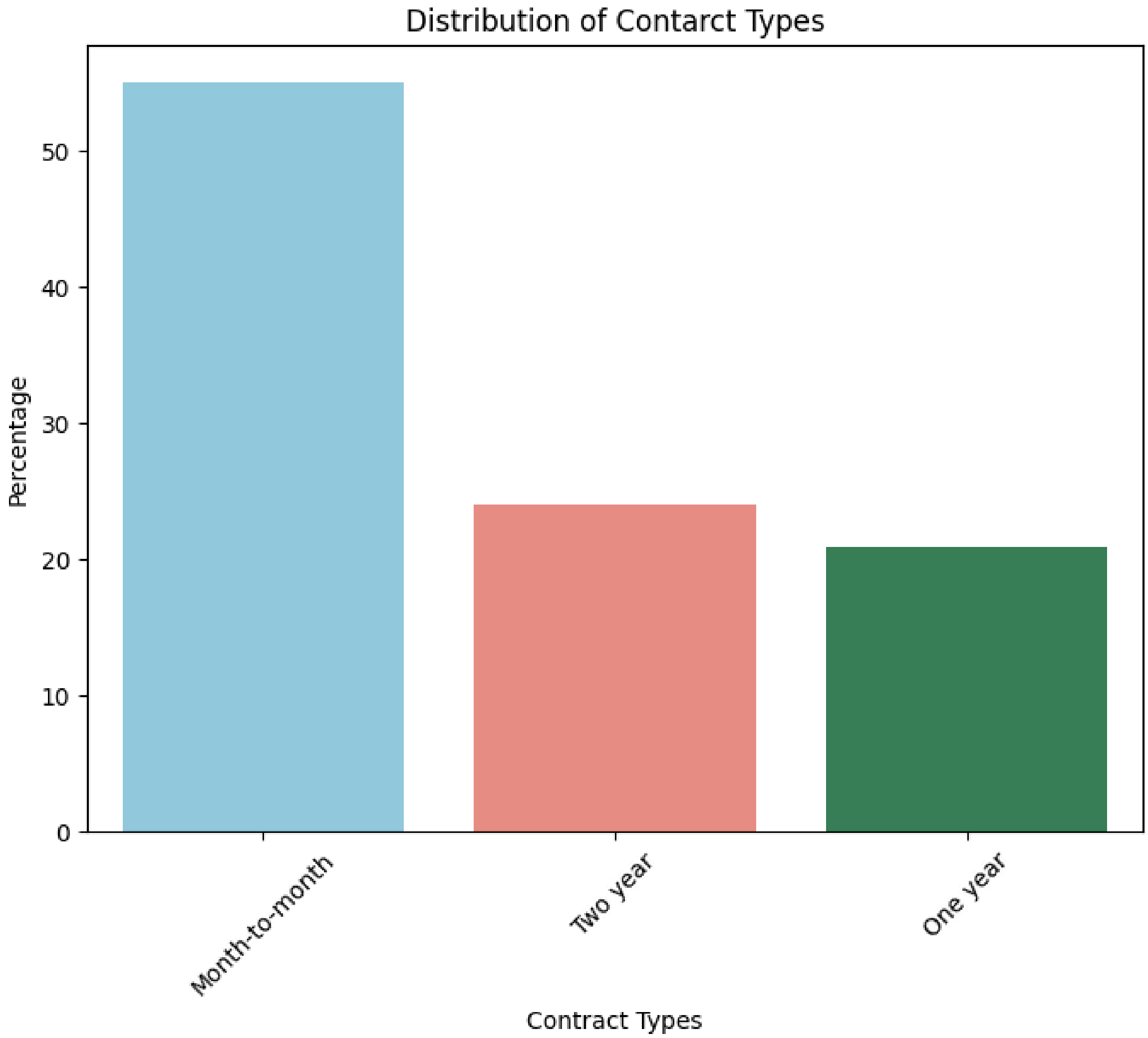




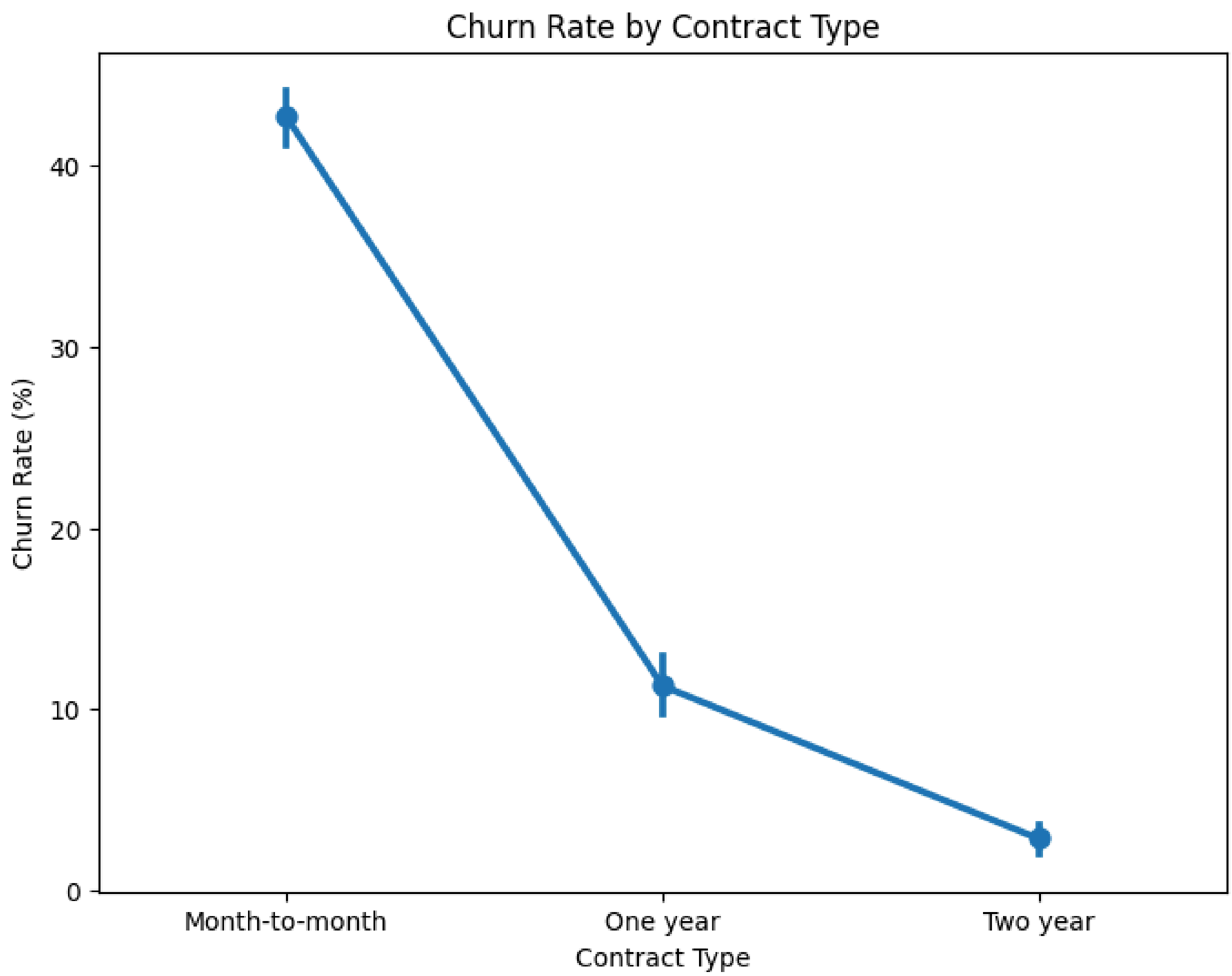


There is a commonality across all nine of the churn charts for the services mentioned above: the percentage of churn rate will decrease for customers who subscribe to the services, while the churn rate will remain greater for those who do not.

3.1.7 Contract

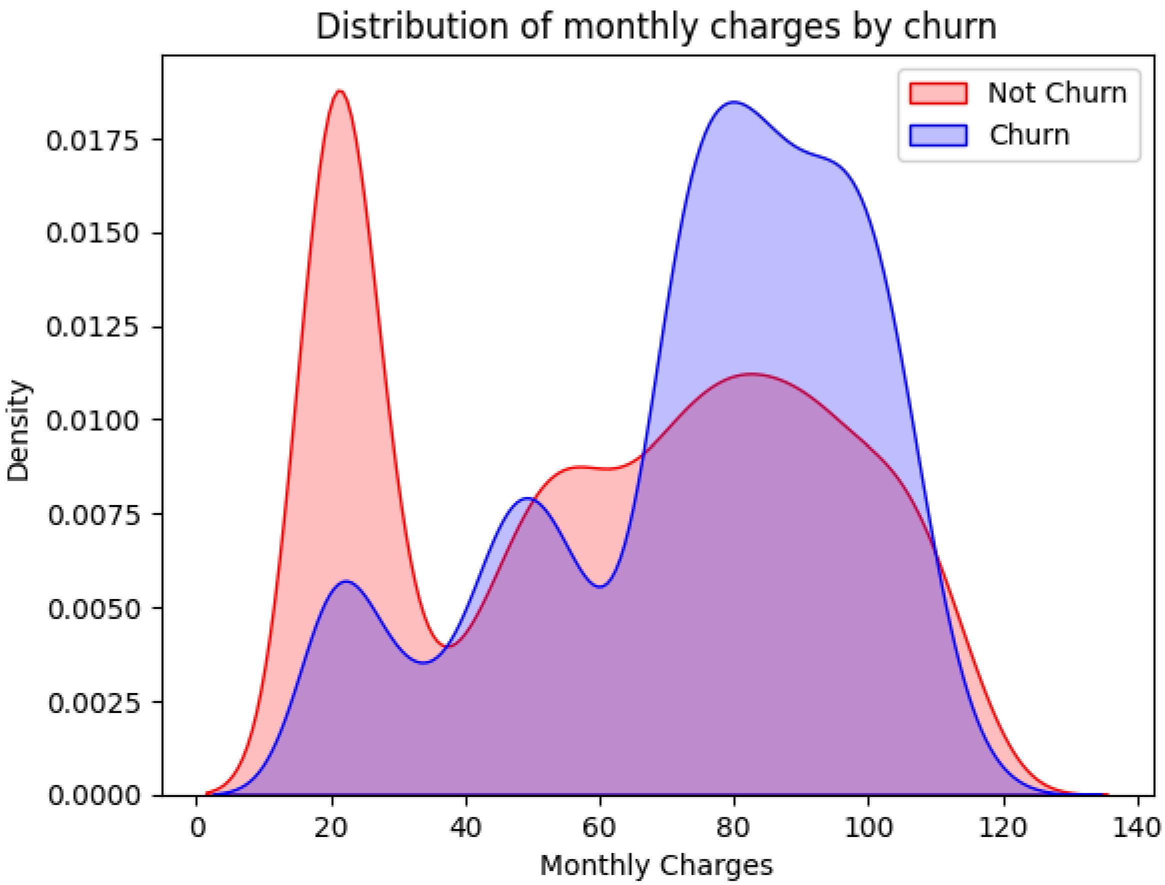


From the diagram above, the most common contract type is “Month-to-month,” indicated by a tall blue bar surpassing 50%, suggesting that over half of the contracts are short-term and lack long-term commitment. This is followed by “Two-year” contracts, represented by a red bar just slightly above 20%, showing a moderate preference for longer stability. The least common are “One-year” contracts, with a green bar around 20%, indicating a smaller segment opting for a mid-term duration.



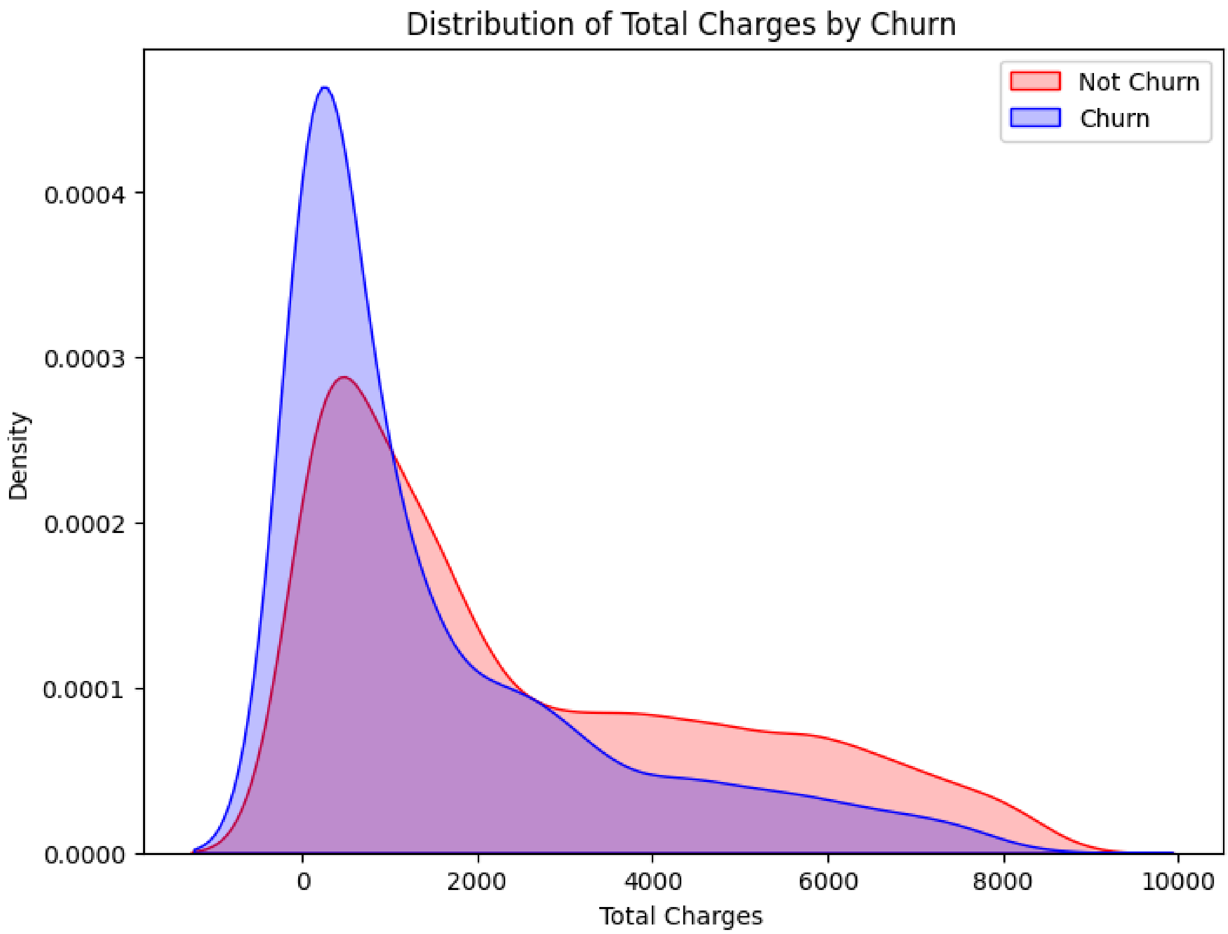
According to the line graph, month-to-month contracts have the greatest churn rate—just over 40%—which suggests that a sizable portion of consumers end their subscription early. On the other hand, the churn rate for one-year contracts drops significantly to about 10%, indicating that lengthier commitments result in higher client retention. With a churn rate of almost 0% for two-year contracts, the client base for long-term contracts is extremely stable. The significance of contract length in client retention efforts is shown by this graphic data, since longer contracts are probably associated with lower customer churn and greater customer loyalty.

3.1.8 Monthly charges



The image features a graph that illustrates the distribution of monthly charges among customers who have churned and those who have not. The red line prominently peaks at lower monthly fees, indicating that non-churning clients are more concentrated in the lower price range. On the other hand, the blue plot peaks at higher monthly fees, suggesting that increased costs are a typical reason for consumers to choose to leave. The pricing range that includes both kept and churned clients is indicated by the purple region where the plots overlap.

3.1.9 Total charges



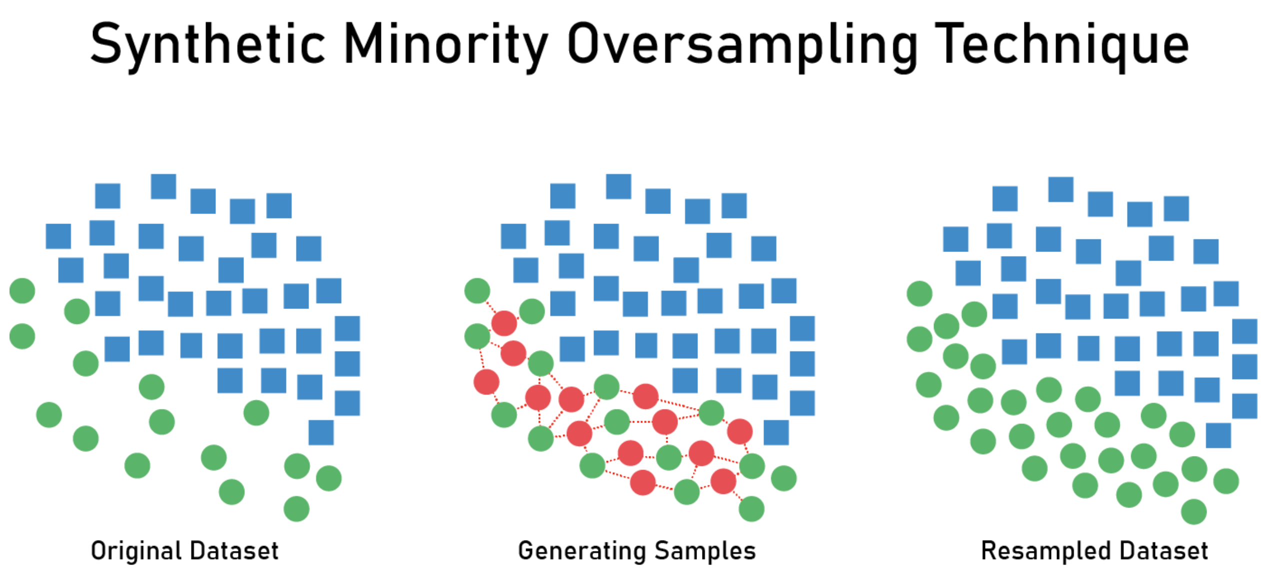
The image displays a graph titled “Distribution of Total Charges by Churn,” which compares the density of total charges for customers who have churned versus those who have not. The “Not Churn” group, shown in blue, has a sharp peak at lower total charges, indicating a high density of customers with lower total expenses who remain with the service. In contrast, the “Churn” group, depicted in red, shows a more gradual increase and extends further along the total charges axis, suggesting that customers with higher accumulated charges are more likely to leave the service. This pattern could imply that higher total charges over time contribute to customer churn.

3.2. Applications of the algorithm(s)

Techniques

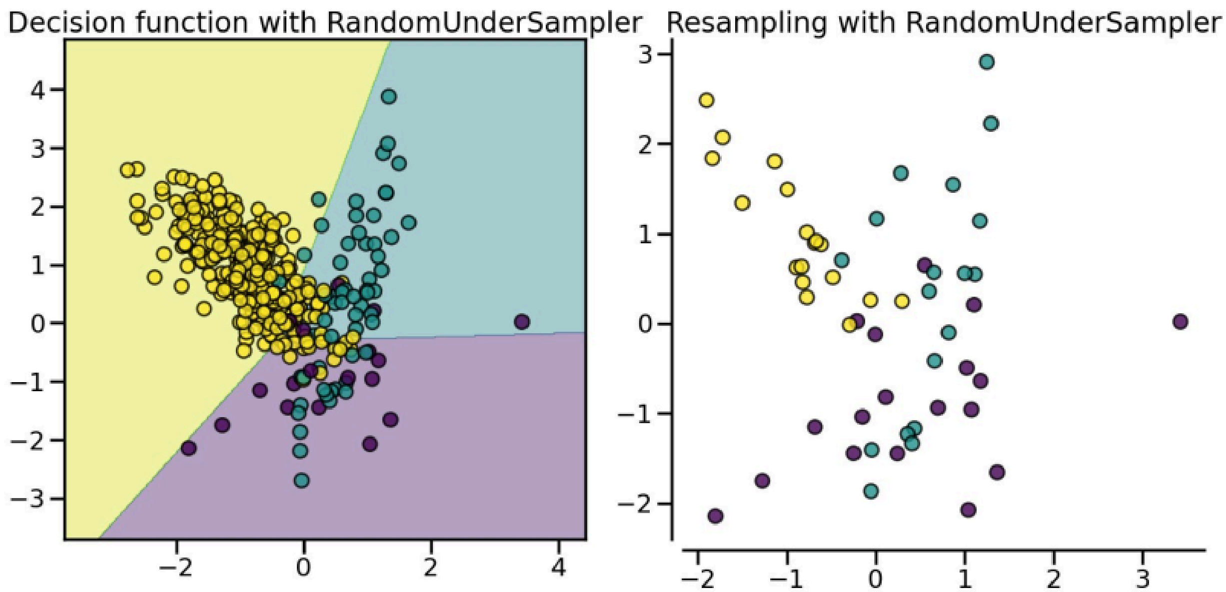
1. Upsampling with SMOTE (Synthetic minority over-sampling technique)

   In the dataset under consideration, the number of elements forming the class that interests us (for example, churned customers) is significantly less than those that do not (for example, non-churned customers), in such cases upsampling becomes the only feasible solution to the problem. Oversampling of minority class is typically one of the ways to produce new synthetic samples for the minority class data set by interpolating a new point between the existing minority class instances. Using the sampling method is advantageous as it helps eliminating imbalance of classes by generating better class distributing. This reduction of the concentration of training data in the minority class ensures that the classesRatio classifier learns the differences between categories with less possibilities for bias toward the majority class.



2. Downsampling with RandomUnderSampler

   We also conduct our data downsampling using RandomUnderSampler algorithm in our case. By downsampling we refer to the process by which the number of samples in the majority class is reduced in order to achieve the balance within the dataset. RandomUnderSampler randomly picks a sample subset of the majority class so as to balance the number of them with those of the minority class. This method precludes the classifier from being swamped by big size of the data present in most of the class; instead, it gives each class an equal weight. Downsampling offers a strong capability of dealing with datasets that are utterly skewed and have a huge number of the dominant class. Slicing down the size of the data, can bring faster training time and the improved computer usage at the same time.



3. Min max scaler

   This procedure of scaling scale the features by adjusting each one for some definite range, usually for the range between 0 and 1. MinMaxScaler is especially efficient under the distribution of the features is not Gauss (other words, not normally distributed) and the reproduction of the shape of the initial distribution is not needed. Through mapping features uniformly, MinMaxScaler or the process of feature scaling helps in ensuring that all the attributes are contributing equally to the model building process. Through applying MinMaxScaler and its dimensions 'tenure', 'MonthlyCharges', and 'TotalCharges' we can be sure their values are on the same scale. Now features whose magnitudes are bigger don't dominate learning process anymore.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

4. Standard scaler

   Unlike MinMaxScaler, StandardScaler scales the features to have a mean of zero and a standard deviation of one, therefore the data is in a normal distribution. StandardScaler is most effective when the distribution of the parameters tends to be Gaussian and there are outliers in the data. The StandardScaler maintains the same distribution by aligning the data around zero, and scaling it with a standard deviation, making the features supply about the exact same magnitude. The StandardScaler is applied to spectral features like 'tenure', 'MonthlyCharges', and 'TotalCharges' in order to overcome the variations of their magnitude, allowing these magnitudes to be compatible with feature scale-sensitive algorithms such as logistic regression and support vector machines.

$$X_i' = \frac{X_i - \mu}{\sigma} = \frac{X_i - X_{\text{mean}}}{X_{std}}$$

StandardScaler

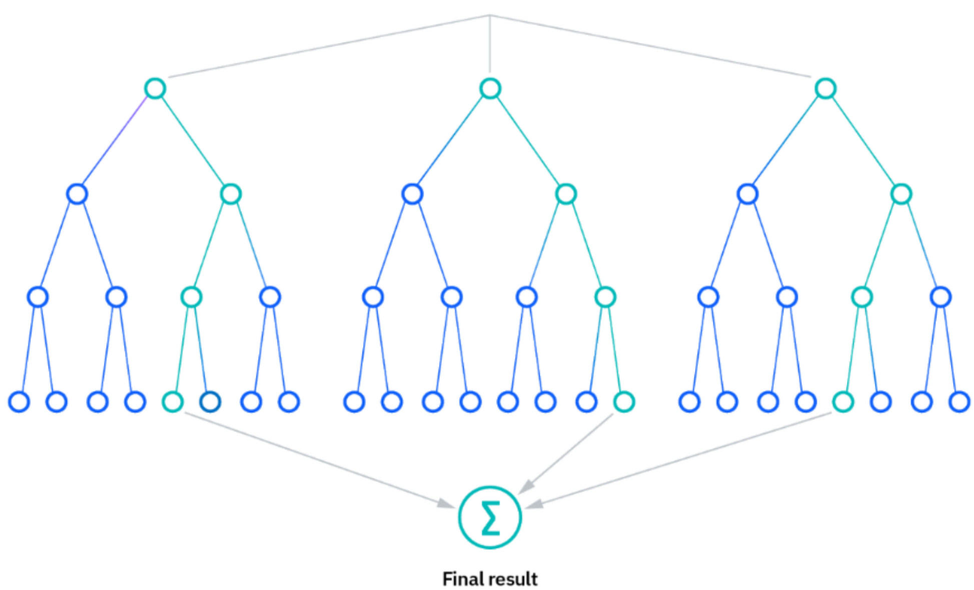
Hyperparameter Tuning

Grid Search CV (Cross-Validation) is a technique that helps to identify the best combination of hyperparameters for a model. In order to use this technique, a grid of possible values for each hyperparameter must be set up. Grid Search CV will evaluate every possible combination of the specified values in the grid for the model. Grid Search CV searches through the grid detailly, computes the performance metric for each combination, and finally selects the parameters that yield the best performance.

Algorithms

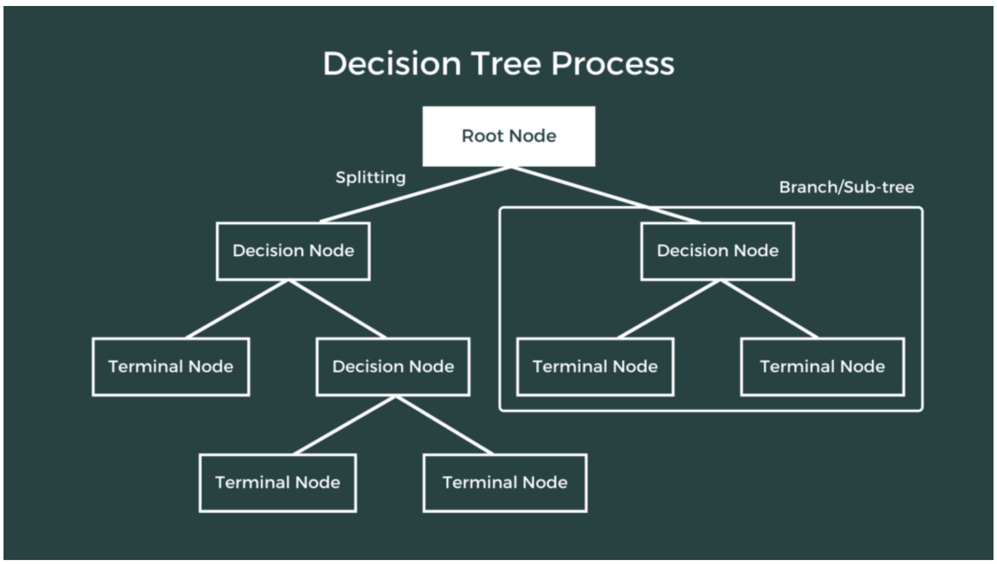
1. Random forest (RF)
   In churn of customers prediction, Random Forest has been an algorithm that is strong and very popular among machines learning algorithms; it has being able to maintain the reputation because of its effective of working. The ensemble principle of learning is the basis of RF which use the collective expertise of several distinct decision trees to produce only one highly accurate answer. First, the algorithm will build a bootstrap sample in which the points are randomly chosen from the training dataset, possibly with repeated exposure. The next step

is the creation of decision trees out of the samples that were generated with the help of bootstrapping. This is a recursive hierarchical process, where the data is continuously partitioned into sub-groups until the sub-group that consists only of the class that represents churned customers or retained customers is obtained. What singles Random Forest out is that it goes through the process of tree construction with an element of randomness during tree construction. Mainly, the selection of the best feature, for node splitting is taken place by the algorithm choosing the optimal feature, randomly, in turn, that adds to diversity and extendes the robustness of the model. Hence, the process is repeated and multiple trees are generated. Then, in consequence, the forest is formed where the decision tree reveals the structure that is indispensable for the classification. The result of the final churn prediction in the ensemble Random Forest is a summation of each tree's vote or input, particularly taking up diverse perspectives from multiple trees. Eventually, that model not only endures but also becomes very strong and accurate so Random Forest model becomes a perfect option for customer churn prediction projects.



2. Decision tree (DT)
   The Decision Trees, used in the customer churn prediction, tactics are praises by their high interpretability and feasibility. These models keep splitting the dataset with regard to the customers' characteristics, aiming to group the customers into two groups propitiously—churned or retained. Decision Trees apply hierarchical segmentation of the feature space where at particular node (customers' feature) each branch represents a decision (based on that feature value). Within model construction, the algorithm takes the most distinguishing features and segments

the dataset such, so that it has homogeneous or the next to the homogeneous matrices in the leaf nodes corresponding to the churn status. This way of presentation features a distinctive graphical representation of the churn prediction process which can be used by stakeholders for grasping such relationships between specific customer characteristics to the propensity to churn. As the Decision Trees are prone to be overfitted, particularly when tackling complex data sets with a high degree of features which may result in worse generalization compared to the more advanced machine learning algorithms. This notwithstanding, simple, interpretable and nonlinear relationship capture feature of Decision Trees makes them an integral part of a customer churn prediction journey.



3. Naive bayes (MNB)
   Multinomial naive Bayes (MNB) being one of the most frequently used probabilistic machine learning algorithms that are known for the simplicity yet effectiveness, finds application in customer churn prediction tasks. Matched for the features serving for the counts, MNB is good at providing chronological details for those datasets that the features being the feature either frequencies or occurrences, making it suitable for customer churn problem based on either categorical or numerical attributes.

   In churn customers prediction problem scenario, MNB formulates that of the probability that one given customer is in a churn class (e.g., is churned or retained). Building on Bayes' theorem with a default assumption of independence across features (customer attributes), a churning probability corresponding to a customer's status is estimated using their attributes. Simplicity of his assumption is not the guarantee of success. Crucial results are often modelled, even though they may be inaccurate.

MNB models are being run using the dynamic of customer attributes as a churn frequency while also search for the strong association between those customer attributes and the churn status. It builds a probabilistic model by applying logistic regression to categorize the attributes along with that churn class based on the training data. Prediction part of MNB determines the probability of a new customer belongs to either of churn classes using the attribute values together and with the application of Bayes' theorem to find out the most likely churn class.

Though MNB is simple and fast at data learning, it may be unable to fully capture the notions for more than two attributes, as well as have some difficulties with rare or never seen attribute values that were not provided with in the training data. However, despite the very high effectiveness of Multinomial Naive in comparison to other algorithms, it stays a baseline algorithm which is often used to approach the categories problem having high dimensional datasets.

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

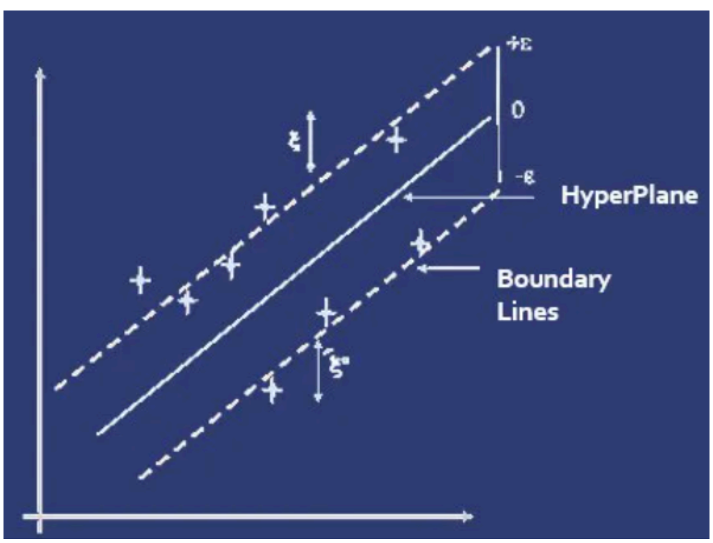
$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

4. Support Vector Machines (SVM)
   Among the supervised learning algorithms, Support Vector Machines (SVMs) can be found as the most robust and versatile that lead as a classification method for both linear and non-linear problems. However, the description is about Support Vectors Regression(SVR) not SVM; although they resemble the fundamental principles and approach.

For the predictive modeling of customer churn in SVM, complex computational process is applied to a layer of numerical data that is ideally split to two classes namely the churned and retained customers. The objective is to find the supporting vectors, which are explained as the closest data points to the hyperplane that sets the margin or distance between the latter and the hyperplane. Linear SVMs, that is similar to linear regression, try to split the high dimensional space with a straight-line-boundary to classify the customers into churned and retained.

Nevertheless, forecasting customer churn cases engages plenty of non-linear connections and complex patterns in data that would be impossible to interpret manually. Under this, SVMs with kernel functions (e.g. polynomial, radial basis function, or sigmoid) maps the data in a higher space to help the creation of nonlinear decision boundaries. Such arc describes by an SVM enables it to deal with both real and multiple level dependencies within the customer churn data, thereby its accuracy increases.

The limits on SVM algorithms used to estimate customer slippage are illustrated by the margin lines or the boundaries within which the system assumes the customer will slip. The SVR looks for margin planes, and the comprehensive line representation is Y=mx+c=a and Y=mx+c=-a, where a is the margin width. The system is working within a predefined set of tolerances (satisfying - a < Y = mx + c < a), which includes the data points that are considered by the SVR in predicting the churn of the customers.

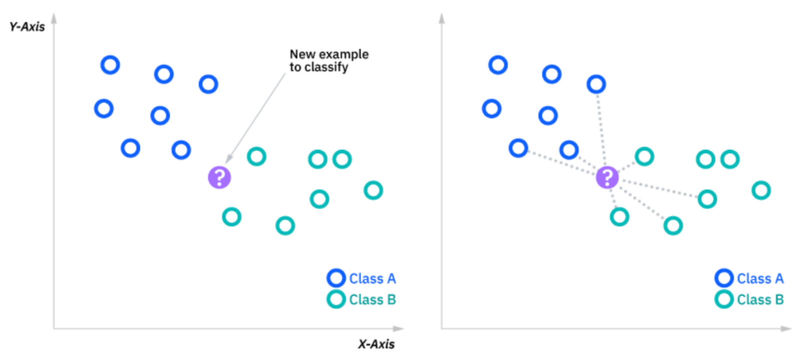


5. K-Nearest Neighbors (KNN)

K-Nearest Neighbor (KNN) algorithm is a versatile supervised machine learning algorithm often used for churned customer prediction in this respect, it operates both is the classification and regression issues. In the area of customer churn classification, KNN provides the class membership determination according to the majority vote among its K nearest neighbors that are distanced with applied metrics like Euclidean, Manhattan, or Minkowski metrics. Such as, taken K=1 the customer is designated as churned or retained class about the label of its nearest neighbor.

In customer churn regression, we predict the probability or likelihood of churn with the close neighbors who have the same churn probabilities. In the case of KNN classifier as well as regression, distance functions are utilized to determine the similarity between observations.

The algorithm works by choosing the K customers which are the most identical with a new customer among those customers who use the churn labels for predicting the current customer's churn status. Users choose K, the number of k nearest neighbors, which affects the model's bias and variance. A larger K value leads to less variance at the same time bias is increased, but a lower K value is linked to high variance and lower bias. KNN is highly successful in handling classifications which come with outliers or noise as a higher K value serves to boost robustness. For classifying task, it is better to use equal K value for avoiding equality during the majority vote.

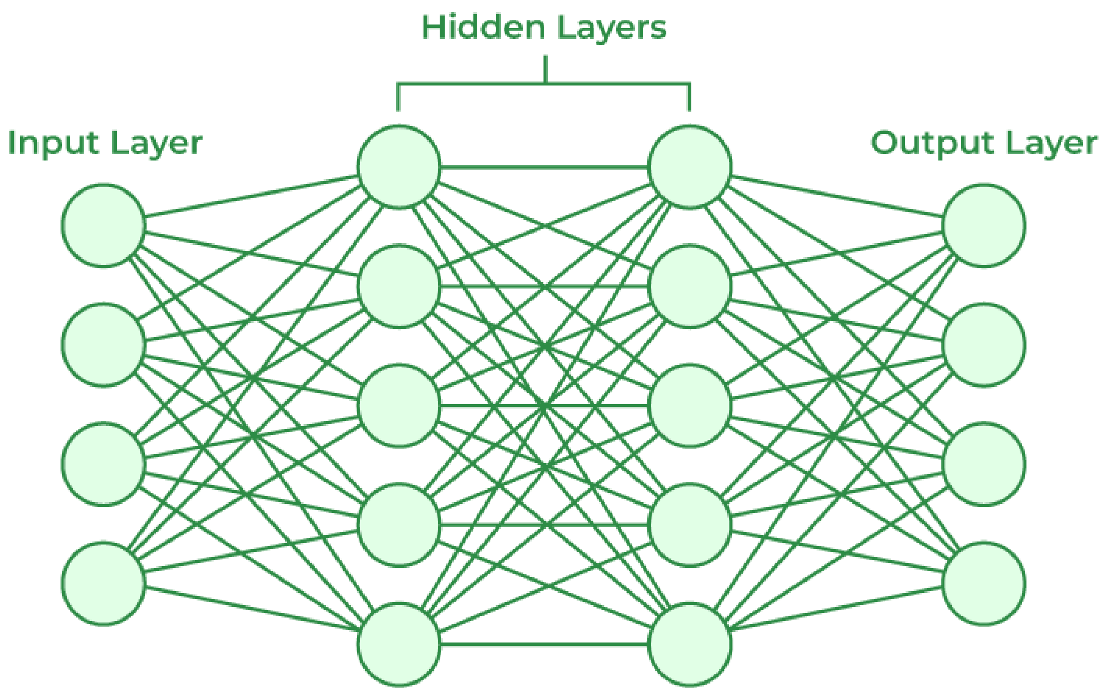


6. Neural networks (NN)
   NN's, or Neural Networks, are an example of the many models that perform very well in customer churn prediction. They provide powerful and versatile tools which help capture complex relationships and patterns in the data. Deep Neural Networks is a kind of machine learning model which is inspired by the structure and functioning of the human brain's neural networks; therefore, it is also called brain-inspired machine learning. When we consider churn prediction for customers Neural Networks are used based on the principle of training through historical customer data resulting in future churn prediction.

   An important component of the Neural Networks is the layers of the nodes that are connected mutually, each of which performs some mathematical transformation. These layers comprise of the input layer, hidden layers (one or more), and output layer. The agent performs training by adjusting the weights and biases carried by connections of neurons between them to fiddle the differences between the predicted churn probabilities and the actual churn labels in the training data.

For customer churn prediction, Neural Networks can take in multiple features associated with customers which are length of service, charges every month, monthly bill, and demographic information respectively. Through a variety of nonlinear types of transformation, Neural Networks enable deep understanding of nuanced patterns and dependencies in the data and therefore lead to fairly exact churn predictions.

The design of a Neural Network applied for customer churn prediction can be different depending on the complexity of the analyzed dataset and the expected precision level. The popular techniques such as dropout regularization, batch normalization and the activation functions like ReLU (Rectified Linear Unit) are widely used in advancements of this concept and for prevention of overfitting.



7. XGBoost (XGB)
   Customer churn prediction has become one of the most popular application areas of machine learning. XGBoost (Extreme Gradient Boosting) has become the algorithm most frequently used in such applications due to its ability to handle structured data and deliver high prediction accuracy. XGBoost belongs to the ensemble extension model and

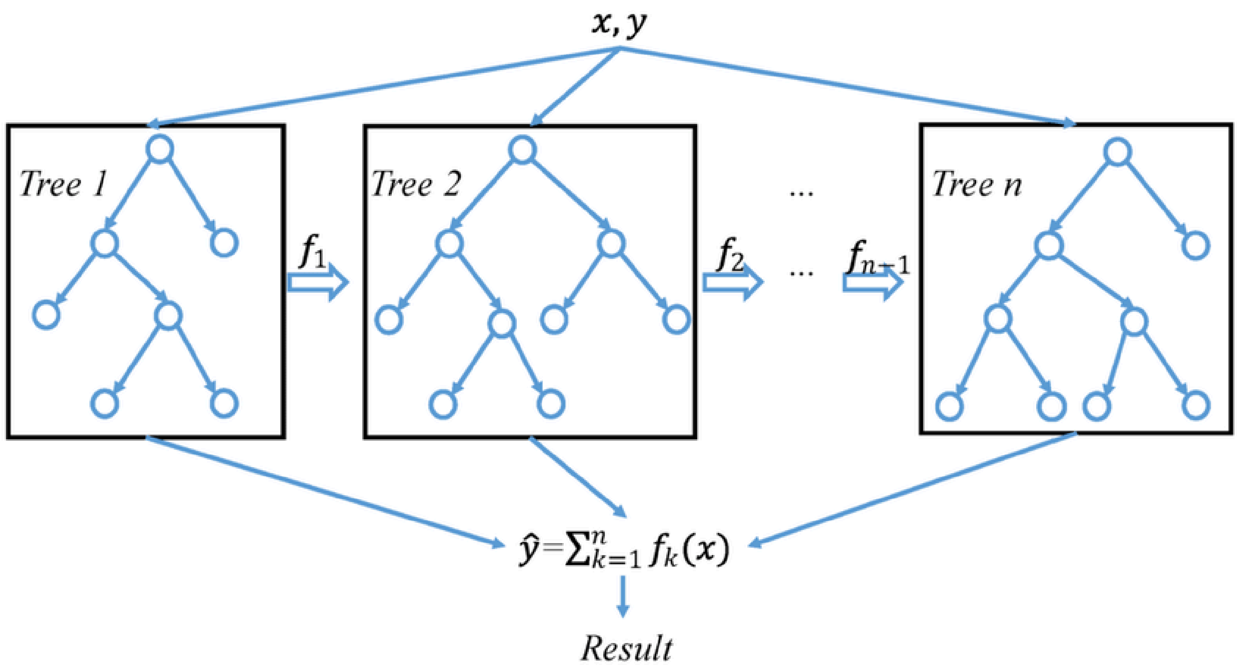
works by building successive decision trees, each correcting the mistakes of the preceding one of which creates a strong predictive model.

In the presence of churn prediction customer model, XGBoost will process a variety of customer attributes including the length of tenure, monthly charges, type of contract, and demographic data. Using repetitive structure of decision trees XGBoost makes the system learn complicated models and dependencies, consequently it gets closer to the prediction.

The algorithm manufacture entails a weighted sum of multiple decision trees, with each tree covers a different region of the feature space. XGBoost uses both learning rate for shrinkage and subsampling of features to free model from overfitting and improve the model generalization. To continue, it uses a method called gradient boosting that increases the model's loss function by randomly reducing the error made by previous trees so as to improve the prediction accuracy.

XGBoost's flexibility beckons us as it is able to address both classification and regression problems powerfully. It estimates probability that a customer will dissatisfied and churn by creating scores that show the possibility of customer churn. These rankings will be employed to rate the customers by their level of churn and assisting business to concentrate on retaining high churn risk customers.

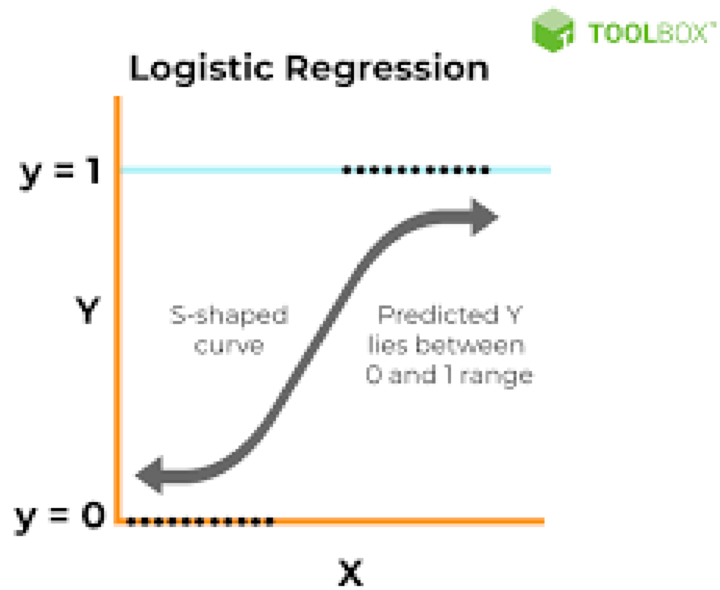
High efficiency of XGBoost is the core reason for its popularity: it has good performance, it is scalable, and it is user-friendly. The feature importance analysis of the system provides the information about which customer attributes are of greater importance in predicting churn. Thus, it assists the data-driven decision-making and strategy formulation. In sum, XGBoost is an efficient tool for predicting customer churn for businesses to take a preemptive action to prevent churn and grant customer relationships.



8. Logistic Regression (LGR)
   Logistic Regression, despite its name, is a classification model that is normally used in customer churn prediction tasks because of its ease of use, reasoning ability. In contrast with linear regression which involves predictions in continuous type of outcomes, logistic regression deals with discontinuous outcomes such as retained or churned customers. It supposes that the probability of a binary output is a function of customer's attributes, which is then transformed into an output value using a logistic function (sigmoid function) and the likelihood of a customer churning between 0 and 1.

   For example, in the aspect of customer churn prediction, there are various features that may include customer data attributes such as tenure, number of charges, and kind of contract type. The Logistic Regression is implicitly based on these features to calculate the log-odds of churn occurring, where contrasting churn is the probability of retention. The churn probability thus is calculated via the given logistic function, where the values closer to 1 denote increased confidence when churn occurs, and the values closer to 0 indicate increased confidence when retention takes place.

   The algorithm will learn the relationship of its input features and churn labels by adjusting its parameters iteratively in optimization which will be concluded by methods such as gradient descent in this case. Although it is a linear model, Logistic Regression has the ability to capture non-linear relationships in the data, possibly with feature engineering techniques or perhaps by employing other models. Its effectiveness in simple models, quick algorithms, and the ability to provide probabilities scores make the Logistic Regression a widely used method in customer churn prediction with big attention to understanding the models.



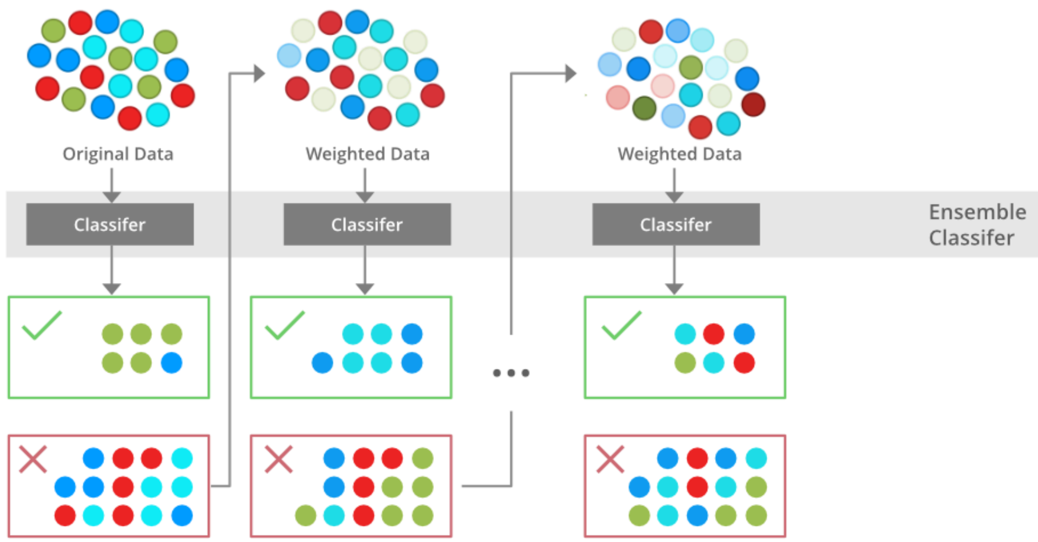
9. ADA Boost (ADB)
   ADA boost (Adaptive Boosting) is a machine learning algorithm used for classifying weak learners into a strong classifier as its using generation and adding weak learners can make a strong ensemble model. Boosting by ADA involves the greedy manner, and each weak learner has the attention on the failed examples so that the coming models are trained more on these examples, hence, make the weight to the challenging instances really important.

   ADA Boost is a technique whereby the system performs a set of weak learners, such as shallow decision trees usually, on the input data and trains the decision trees iteratively. A weak learner is an individual who studies the misclassifications of the previous models, thereby, developing the model systematicality have access with the complex patterns and dependencies within the churn data.

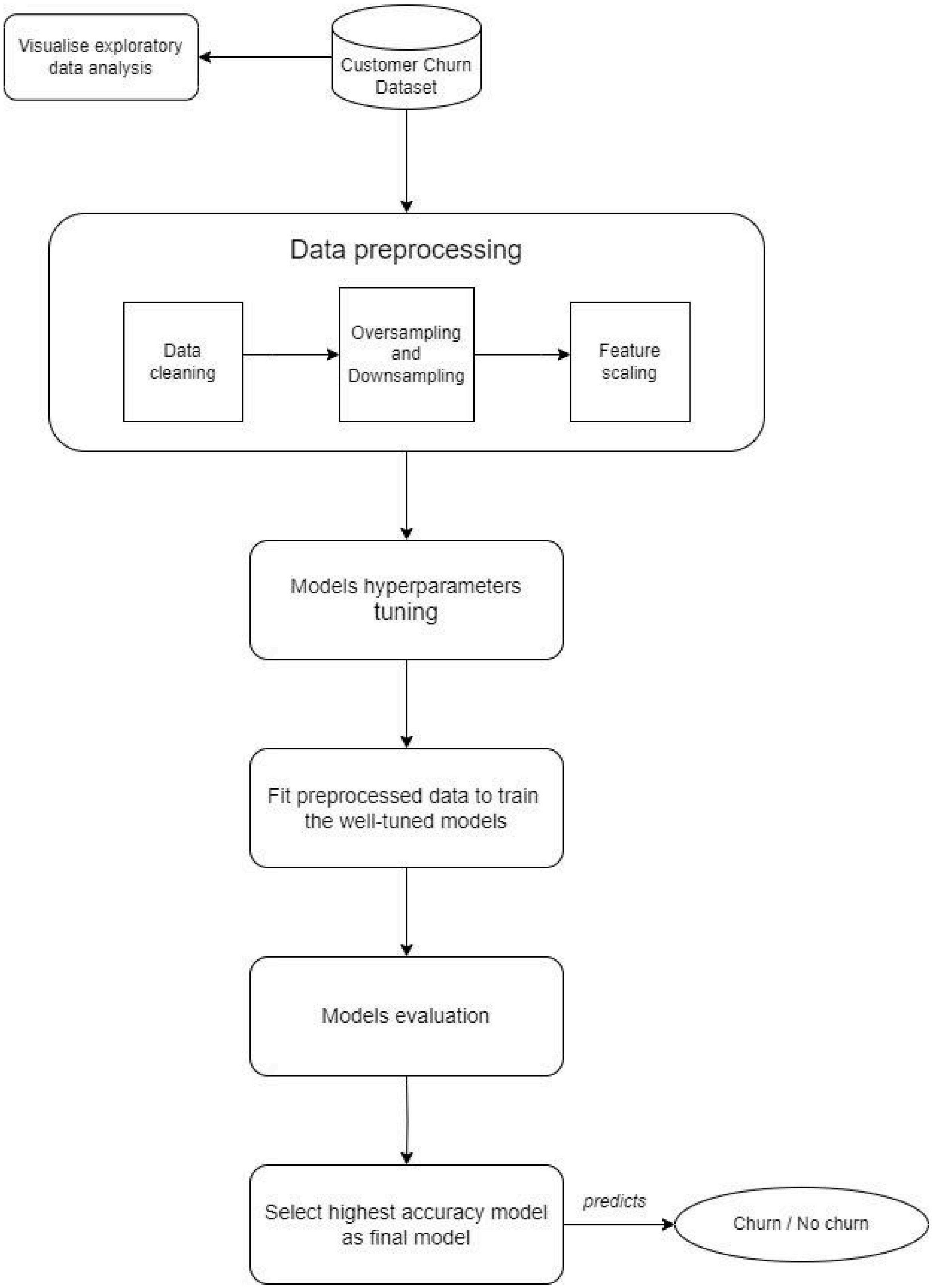
   By doing so, it boosts the more misclassified instances during each round, thus giving them the highest weights and, thereby, becoming the most influential. This iterative process, ADA Boost, gradually produces a robust ensemble model that displays a high rate of predictive exparctul, eventually leading to accurate customer churn.

   It is considered one of the main benefits of ADA Boost that it is capable of working out well with unbalanced data such as customer churn tracking, which makes it an even stronger model in those prediction tasks where the number of churned customers is lower than steady customers. Differential Analysis Formulation Boost is one of the most effective machine learning algorithms that focuses on a certain type of classified instances, therefore helping the model to learn from the class that is less represented and becomes better at recognizing at-risk customers.

Elaborating further, ADA Boost stand supreme in ease of implementation and simplicity as the reason of why it is quite popular among practitioners and data scientists who are using it as a customer churn prediction tool. The ensemble consists of the amalgamation of multiple weak learners in which each of them brings in its own kind of prejudices into the chur data; therefore, the powerful predictive model has the ability of translating those prejudices into the most accurate churn prediction.



3.3. System flowchart/activity diagram



3.4. Proposed test plan/hypothesis

From the research background in 2.1 we can do some hypothesis of :
Hypotheses:

1. Upsampling and downsampling techniques will improve the prediction accuracy of the churn model compared to using the original dataset but up-sampling will perform better than down-sampling.

2. Scaling techniques, such as MinMax scaling and Standard scaling, will have the same effects on model performance.

3. Ensemble methods, such as Random Forest or Gradient Boosting, will yield superior performance compared to individual classifiers like Decision Trees or Logistic Regression.

To test whether the hypothesis is true, we need no do the following testing :
Experiment/Test Design:

1. Divide the dataset into training and testing sets using a suitable split ratio (80% training, 20% testing).
2. Implement various preprocessing techniques (e.g., upsampling, downsampling, scaling) individually and in combination.
3. Train multiple classification algorithms and ensemble methods on the preprocessed training data.
4. Evaluate the performance of each model on the testing data using standard evaluation metrics (Accuracy, Precision, Recall, F1 score, AUC-ROC)
5. Compare the performance of models across different preprocessing techniques and algorithms.

4. Result

4.1. Results

4.1.1 Accuracy

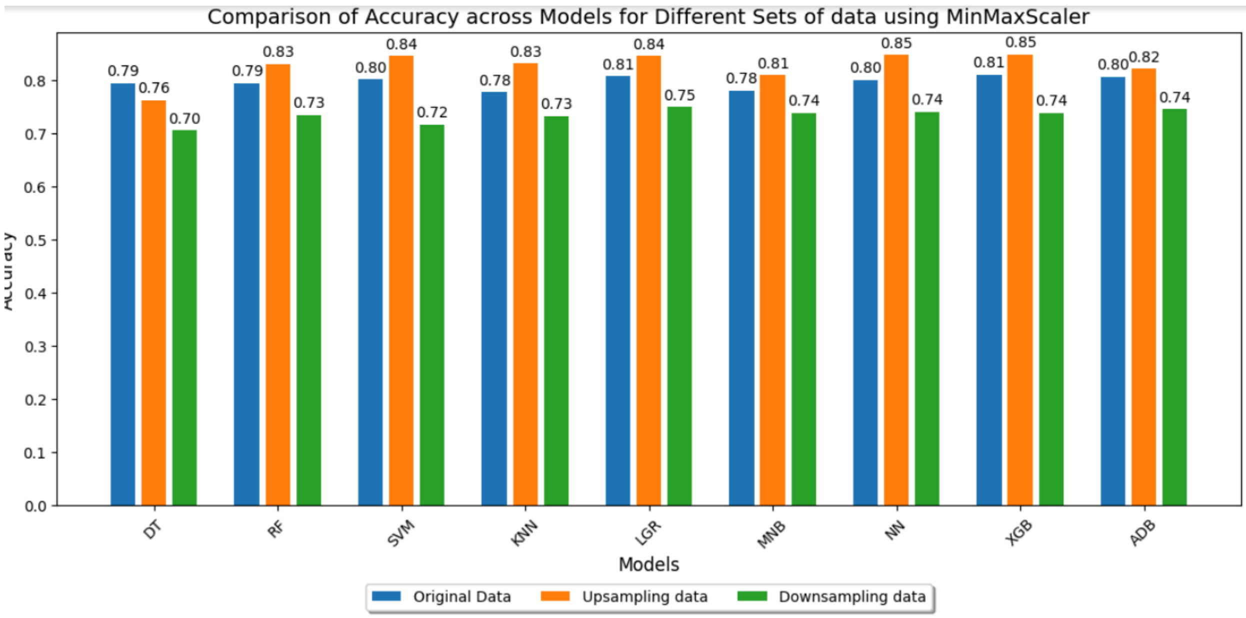


Figure 4.1 : Accuracy chart for dataset 1

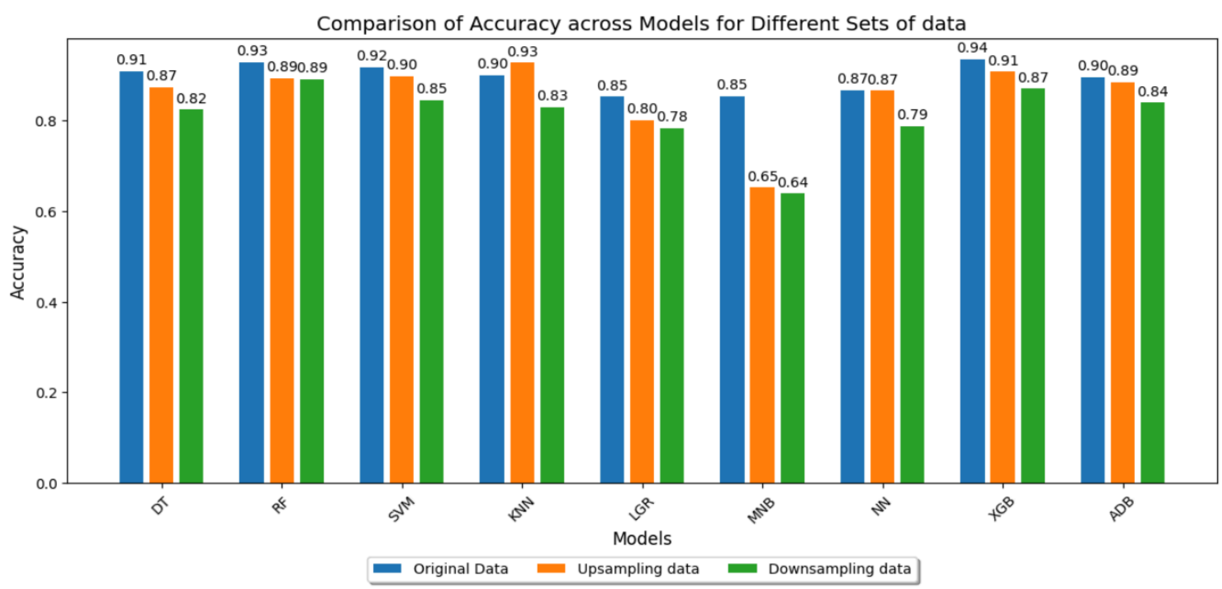


Figure 4.2 : Accuracy chart for dataset 2

	Dataset 1			Dataset 2		
	original	upsampling	downsampling	original	upsampling	downsampling
Decision Tree	0.79	0.76	0.70	0.91	0.87	0.82
Random Forest	0.79	0.83	0.73	0.93	0.89	0.89
support vector machine (SVM)	0.80	0.84	0.72	0.92	0.90	0.85
KNN	0.78	0.83	0.73	0.90	0.93	0.83
Logistic Regression	0.81	0.84	0.75	0.85	0.80	0.78
Naive Bayes	0.78	0.81	0.74	0.85	0.65	0.64
Neural Network	0.80	0.85	0.74	0.87	0.87	0.80
XGBoost	0.81	0.85	0.74	0.94	0.91	0.87
AdaBoost	0.80	0.82	0.74	0.90	0.90	0.85

Table 4.1 : Accuracy table between different dataset.

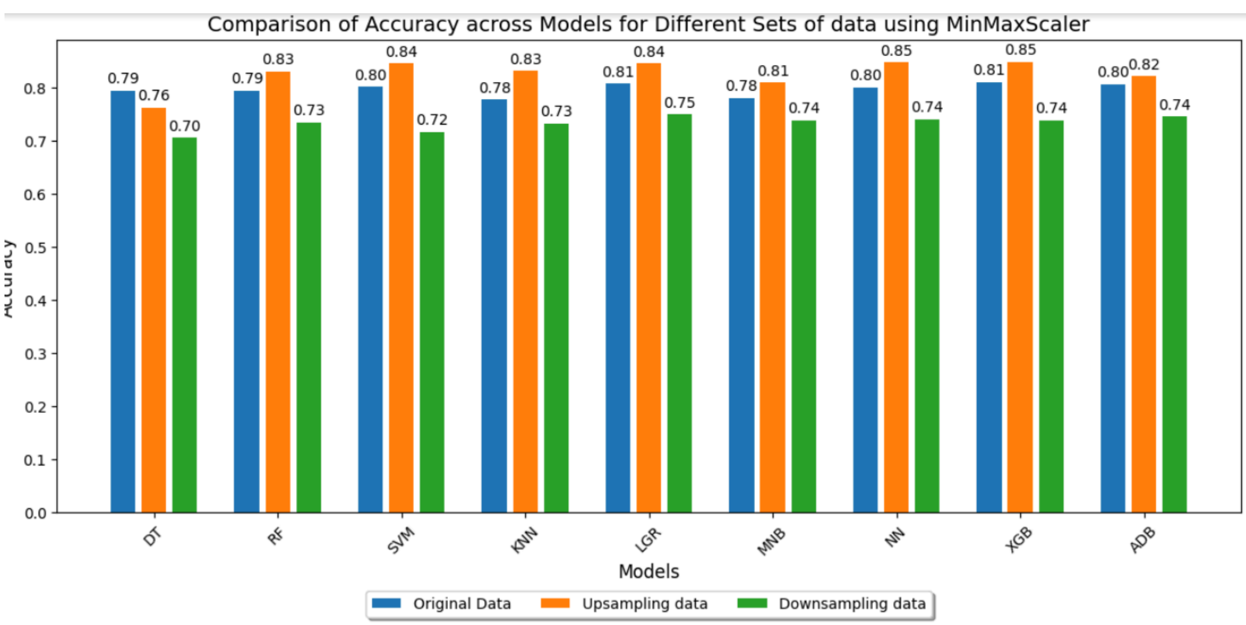


Figure 4.3 : Accuracy chart for dataset using MinMax Scaler

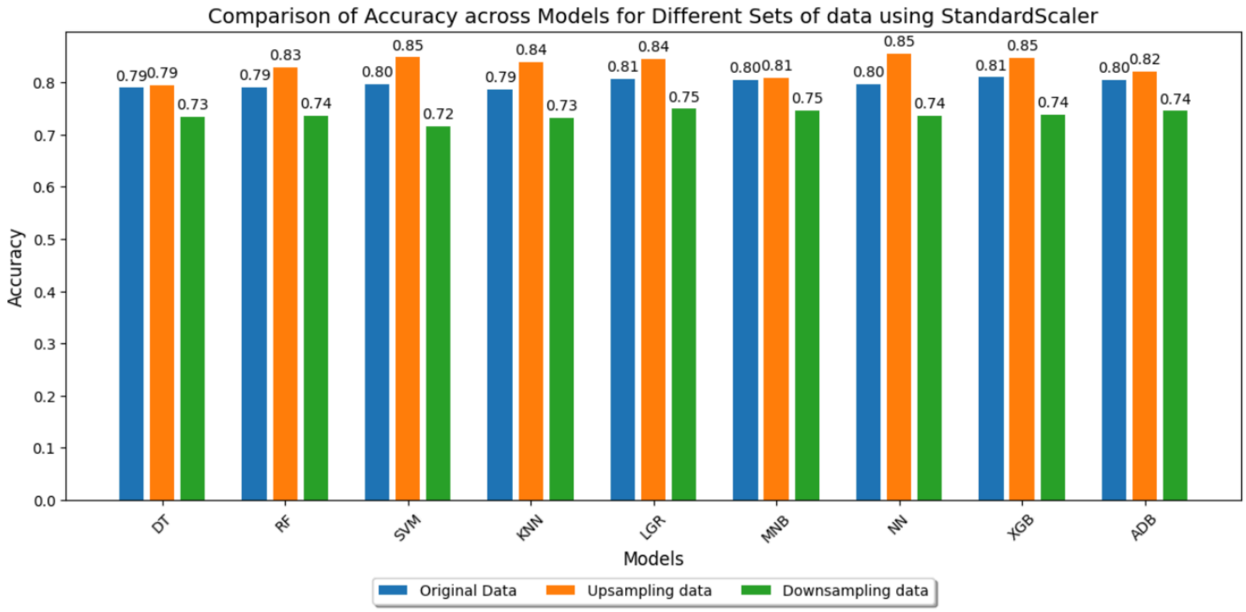


Figure 4.4 : Accuracy chart for dataset using Standard Scaler

	MinMax Scaler			Standard Scaler		
	original	upsampling	downsampling	original	upsampling	downsampling
Decision Tree	0.79	0.79	0.73	0.79	0.79	0.73
Random Forest	0.79	0.83	0.74	0.79	0.83	0.74
support vector machine (SVM)	0.80	0.85	0.72	0.80	0.85	0.72
KNN	0.79	0.84	0.73	0.79	0.84	0.73
Logistic Regression	0.81	0.84	0.75	0.81	0.84	0.75
Naive Bayes	0.80	0.81	0.75	0.80	0.81	0.75
Neural Network	0.80	0.85	0.74	0.80	0.85	0.74
XGBoost	0.81	0.85	0.74	0.81	0.85	0.74
AdaBoost	0.80	0.82	0.74	0.80	0.82	0.74

Table 4.2 : Accuracy between different feature scaling technique

4.1.2 Precision

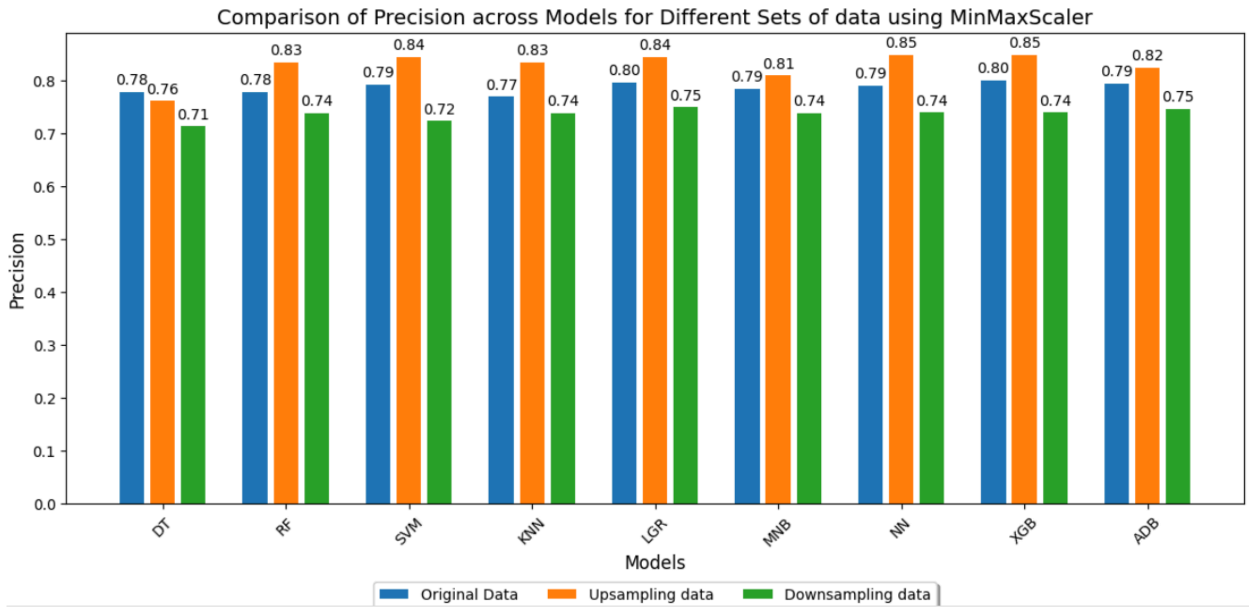


Figure 4.5 : Precision chart for dataset using MinMax Scaler

	original	upsampling	downsampling
Decision Tree	0.78	0.76	0.71
Random Forest	0.78	0.83	0.74
support vector machine (SVM)	0.79	0.84	0.72
KNN	0.77	0.83	0.74
Logistic Regression	0.80	0.84	0.75
Naive Bayes	0.79	0.81	0.74
Neural Network	0.79	0.85	0.74
XGBoost	0.80	0.85	0.74
AdaBoost	0.79	0.82	0.75

Table 4.3 : Precision table of different resampling technique

4.1.3 Recall

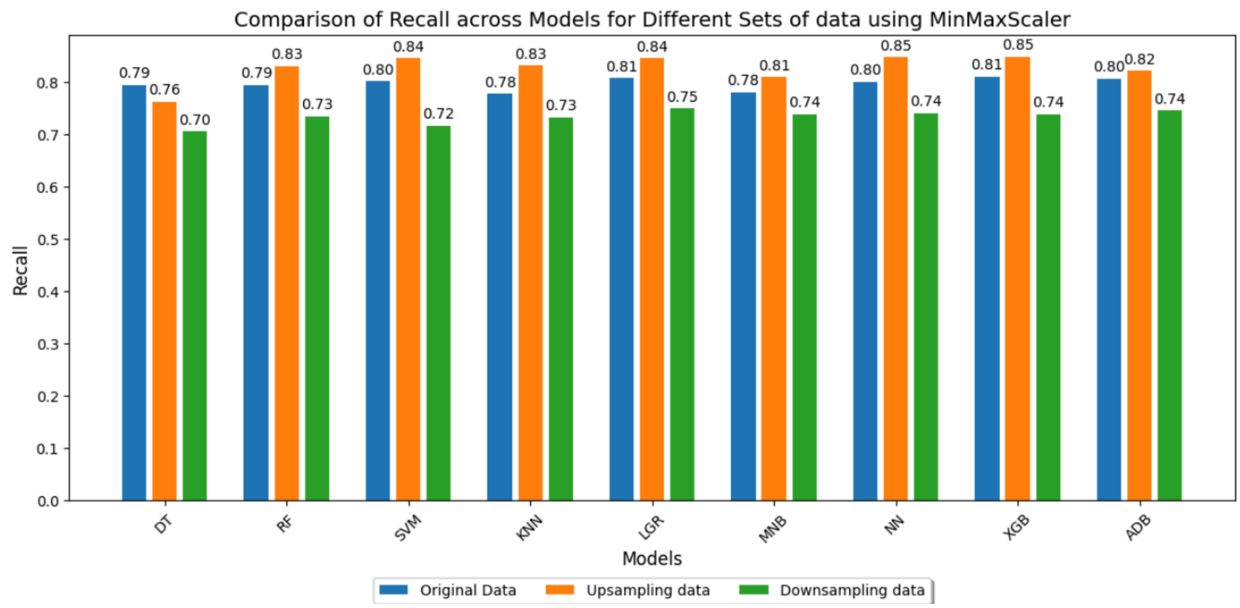


Figure 4.6 : Recall chart for dataset using MinMax Scaler

	original	upsampling	downsampling
Decision Tree	0.79	0.76	0.70
Random Forest	0.79	0.83	0.73
support vector machine (SVM)	0.80	0.84	0.72
KNN	0.78	0.83	0.73
Logistic Regression	0.81	0.84	0.75
Naive Bayes	0.78	0.81	0.74
Neural Network	0.80	0.85	0.74
XGBoost	0.81	0.85	0.74
AdaBoost	0.80	0.82	0.74

Table 4.4 : Recall table of different resampling technique

4.1.4 F1-Score

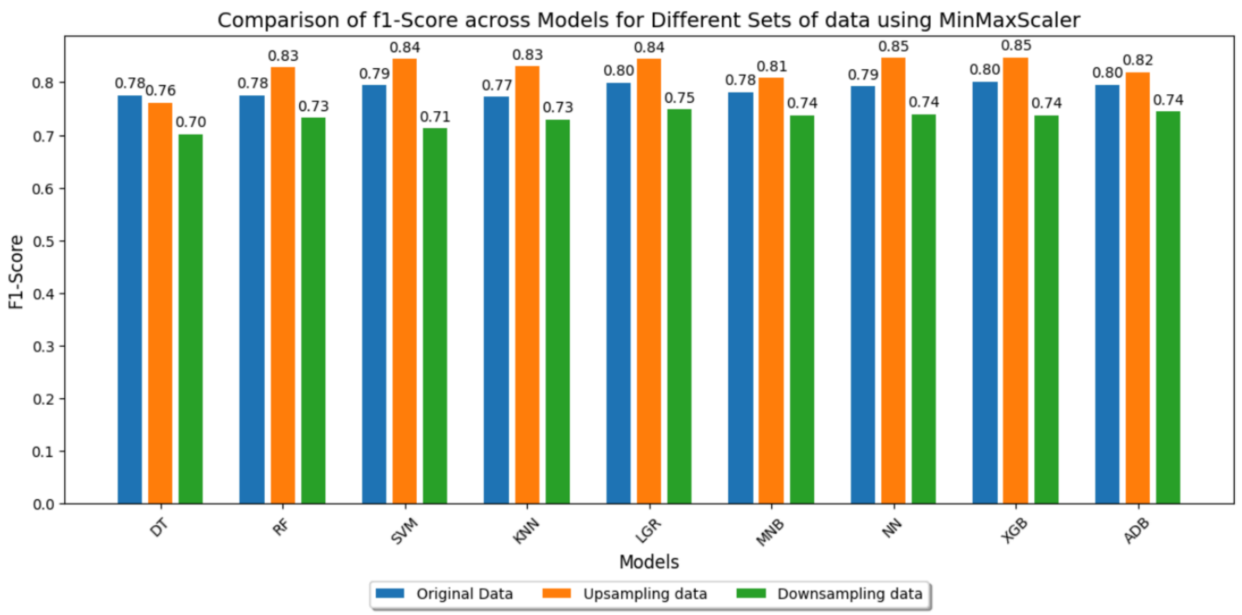


Figure 4.7 : F1-Score chart for dataset using MinMax Scaler

	original	upsampling	downsampling
Decision Tree	0.78	0.76	0.70
Random Forest	0.78	0.83	0.73
support vector machine (SVM)	0.79	0.84	0.71
KNN	0.77	0.83	0.73
Logistic Regression	0.80	0.84	0.75
Naive Bayes	0.78	0.81	0.74
Neural Network	0.79	0.85	0.74
XGBoost	0.80	0.85	0.74
AdaBoost	0.80	0.82	0.74

Table 4.5 : F1-Score table of different resampling technique

4.1.5 AUC Score

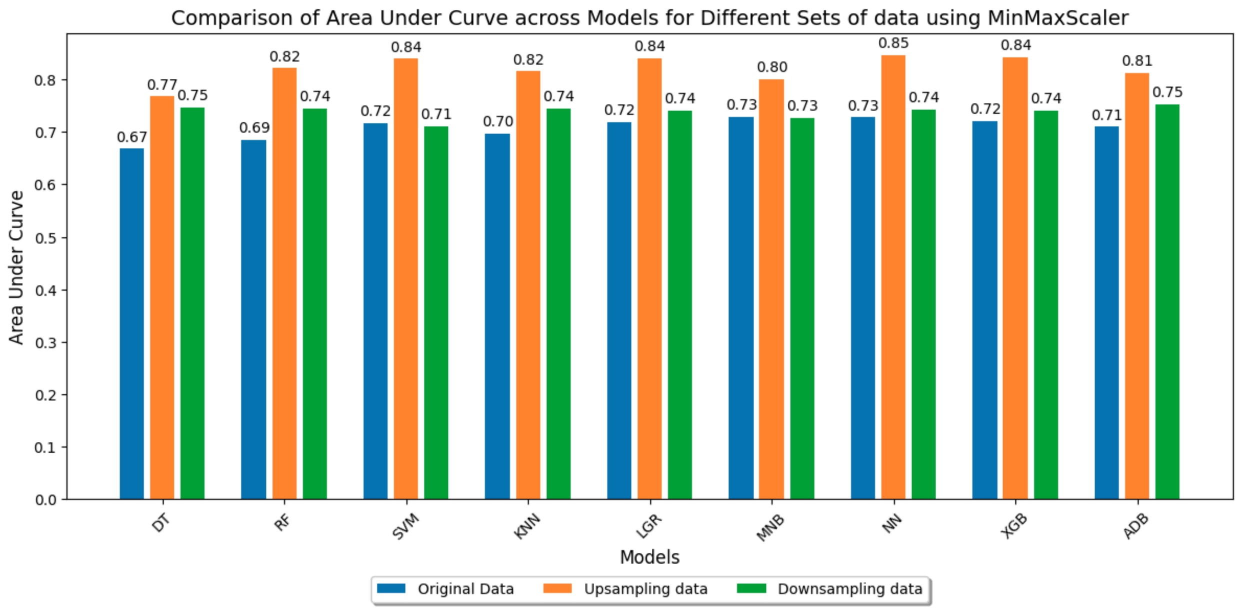


Figure 4.8 : AUC Score chart for dataset using MinMax Scaler

	original	upsampling	downsampling
Decision Tree	0.67	0.77	0.75
Random Forest	0.69	0.82	0.74
support vector machine (SVM)	0.72	0.84	0.71
KNN	0.70	0.82	0.74
Logistic Regression	0.72	0.84	0.74
Naive Bayes	0.73	0.80	0.73
Neural Network	0.73	0.85	0.74
XGBoost	0.72	0.84	0.74
AdaBoost	0.71	0.81	0.75

Table 4.6 : AUC Score table of different resampling technique

4.2. Discussion/Interpretation

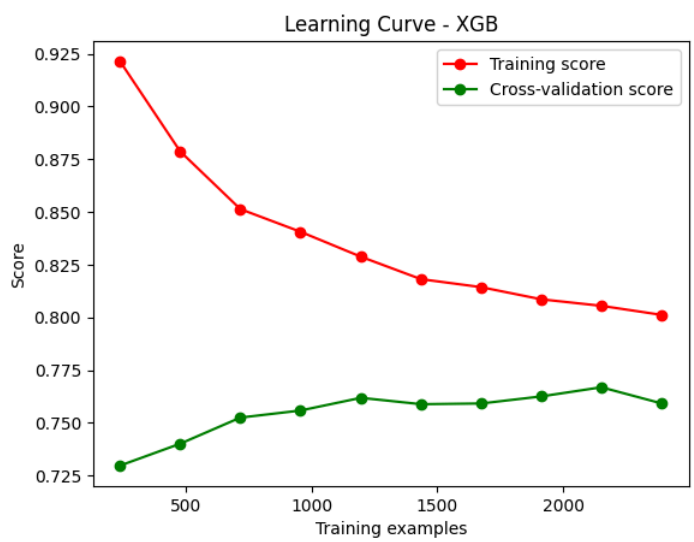
1. Effect of scaling method:

   The results presented in Table 4.2 showed that the accuracy for both minmax scaler and standard scaler have the similar result. Both MinMaxScaler and StandardScaler aim to normalize the features, making them more comparable and improving the performance of certain machine learning algorithms. Despite their different approaches to normalization, they both achieve the overarching goal of scaling the features within a certain range. MinMaxScaler scales the features to a predetermined range (usually [0, 1]), While StandardScaler scales the features to have a mean of 0 and a standard deviation of 1. In this case, the specific range or distribution of the scaled features does not significantly impact the performance of certain algorithms, leading to similar accuracy scores. In summary, the similar accuracy obtained using MinMaxScaler and StandardScaler suggests that for the specific dataset and machine learning algorithms used, the choice of scaling technique may not be a critical factor in achieving high accuracy.

2. Effect of Dataset Resampling Techniques:

   The results presented in table 4.1, table 4.2, table 4.3, table 4.4, table 4.5 and table 4.6 showed that all the accuracy, precision, recall, f1-score and AUC score of upsampling techniques is better as compared to downsampling and the original dataset across all models. This result is explained by the notable class imbalance in the original dataset, where the amount of churn instances is significantly lower than the number of non-churn cases. Specifically, the original dataset contains 5174 non-churn instances and only 1869 churn instances, thereby creating an imbalance that can significantly impact a model's performance. Due to the class imbalance, models trained on the original dataset may result in biases towards the majority class, leading to poor predictions for the minority class.

   Upsampling is the technique used to solve this issue by increasing the number of minority class instances, providing the model with more balanced training data. As a result, models trained on upsampled data show improved performance across various metrics. On the other hand, downsampling seeks to fix the class imbalance by reducing the number of majority class instances. Even though downsampling can improve the balance between classes,if majority class instances are reduced too much, valuable information that is necessary for model learning may be accidentally removed. Therefore, downsampling techniques may lead to decreasing model performance due to the reduced availability of training data. Moreover, with a reduced dataset, the model may find it easier to memorize the training examples rather than generalize from them. This can lead to overfitting, where the model captures noise or random fluctuations in the training data as if they were significant patterns.




```
Mean cross-validation score: 0.7591973244147157
Training score: 0.8003344481605351
```

Figure 4.9 Learning curve of XGBoost model using downsampling technique

From the figure x, it showed that the model is overfitting when training. A training score of 0.83 indicates that the model performs very well on the training data. It suggests that the model has learned the patterns in the training data almost perfectly, achieving a high level of accuracy. However, the lower cross-validation score of 0.75 suggests that the model does not perform as well on unseen data (validation data). This could be due to the model not able to generalize well to new and unseen instances. The lower cross-validation score indicates that the model's performance drops when applied to data it hasn't seen before, which is a characteristic of overfitting.

In summary, the preference for upsampling over downsampling and the original dataset showed the critical role of addressing class imbalance in enhancing model performance. By providing more representative training data through upsampling, models can better capture the complexities of the underlying data distribution, leading to more accurate and robust predictions.

3. Overall Model Performance Analysis:

   From table 4.1, it showed that in dataset 1 which contains more binary data and some columns of integer values, Neural Network and XGBoost achieved the highest accuracy scores of 0.85. Logistic Regression and SVM also performed well, with an accuracy of 0.84. Naive Bayes had a slightly lower accuracy of 0.81, while Decision Tree and AdaBoost showed moderate performance. Table 4.2, table 4.3, table 4.4, table 4.5 and table 4.6 also showed that Neural Network and XGBoost had better performance as compared to others.

   While in dataset2 that is used to test the consistency of the model, which consists of more integer values and some binary data, KNN emerged as the top-performing model with the highest accuracy score of 0.93. XGBoost and SVM performed consistently well in both datasets, with accuracy scores of 0.91 and 0.90, respectively. Naive Bayes experienced a significant drop in accuracy, performing the lowest among all models with a score of 0.65.

   From the result, We can conclude that Neural Network and XGBoost consistently performed well across both datasets, indicating their robustness and adaptability to different data distributions. The ensemble nature of XGBoost and the ability of Neural Networks to capture complex patterns contribute to their high accuracy scores. Next, KNN demonstrated excellent performance in Dataset 2, indicating its effectiveness in handling datasets with numerical features. The simplicity of the KNN algorithm and its reliance on local patterns contribute to its success in this scenario. While the significant decrease in accuracy observed in Dataset 2 suggests that Naive Bayes may be sensitive to changes in data distribution or feature characteristics. The assumption of independence among features in Naive Bayes may not hold true in Dataset 2, leading to poorer performance.

4. Model Selection :

From table 4.1, it showed that SVM, Neural Network and XGBoost achieved the highest accuracy (0.85) in dataset 1. To assess model consistency, we also trained and evaluated these models on Dataset 2 and apply the same preprocess technique.In dataset 2, the performance rankings differed slightly with KNN achieved the highest accuracy (0.93), followed by XGBoost(0.91), and AdaBoost(0.90). Despite the variation in performance across datasets, XGBoost demonstrated consistent performance by maintaining its position among the top-performing models in both Dataset 1 and Dataset 2.

This consistency in performance across different datasets underscores the reliability and robustness of XGBoost as a churn prediction model. Its ability to adapt to diverse data distributions and capture complex relationships between features makes it a versatile choice for churn prediction tasks in the telecommunications industry.

Overall, the selection of XGBoost as a preferred model for churn prediction is supported by its consistent performance across multiple datasets, highlighting its suitability for deployment in real-world scenarios where data distributions may vary. This demonstrates the importance of assessing model consistency and robustness to ensure reliable and effective churn prediction in the telecommunications industry.

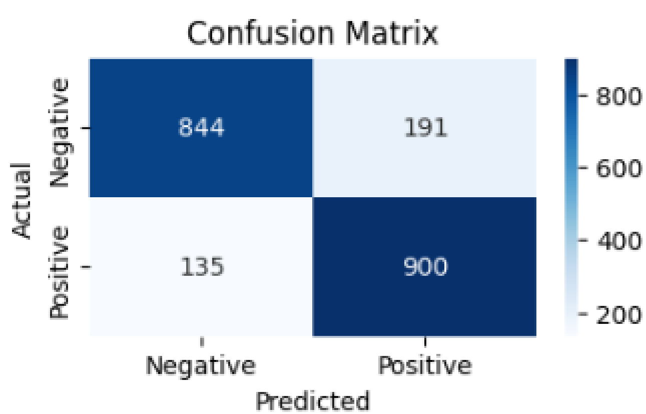


Figure 4.10 : Confusion Matrix of XGBoost trained by Upsampled and Scaled dataset

In addition, in the confusion matrix shown in figure xx, it shows that the true positive (correctly predicted churn) is 900, while the false positive (incorrectly predicted churn) is 135. This means that the model correctly identified 900 churn cases and predicted 135 cases of not churn as churn. This indicates that the XGBoost model is effective at identifying true churn cases among all actual churn instances. On the other hand , the false negative is 191 and the true negative is 844. This means that the model correctly identified 844 not churn classes but predicted 191 churn classes as not churn. This indicates that the XGBoost model is also effective at identifying true non-churn cases among all actual non-churn instances. The XGBoost model achieved high precision and recall indicating its robustness in distinguishing between churn and non-churn instances.

5. Deployment

In conclusion, based on the previous discussion, we will use the XGBoost as the machine learning model that is trained by a dataset that is resampled using upsampling technique and scaled by minmax scale to deploy a web application that can be used to predict the customer churn. This can achieve our goal to design a customer churn prediction application that can accurately predict the customer churn. The final system is shown in figure 4.11 and 4.12. From this result, we can see that the application is able to predict the churn and notchurn based on different input from the user.

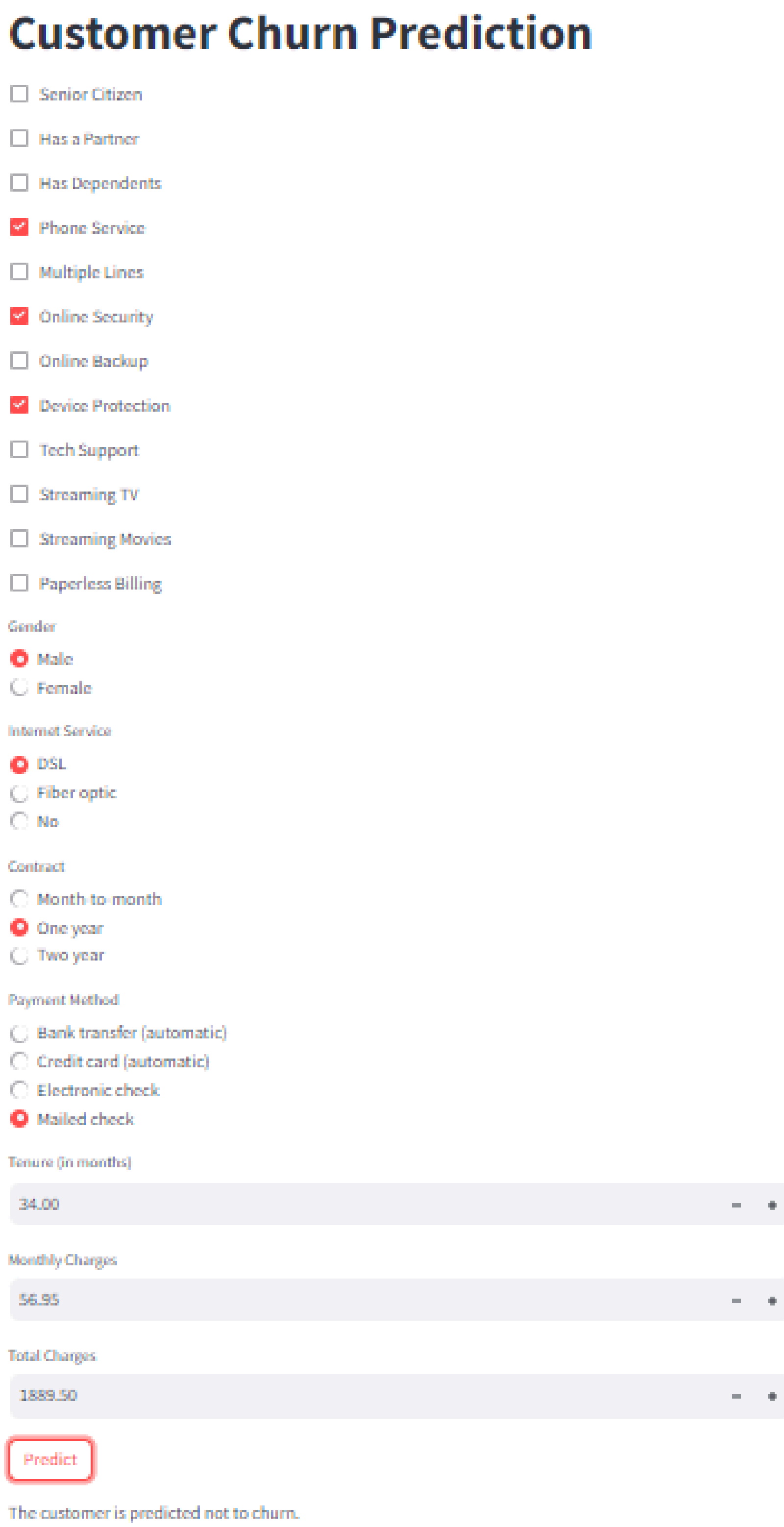


Figure 4.11: Example predicted of not to churn

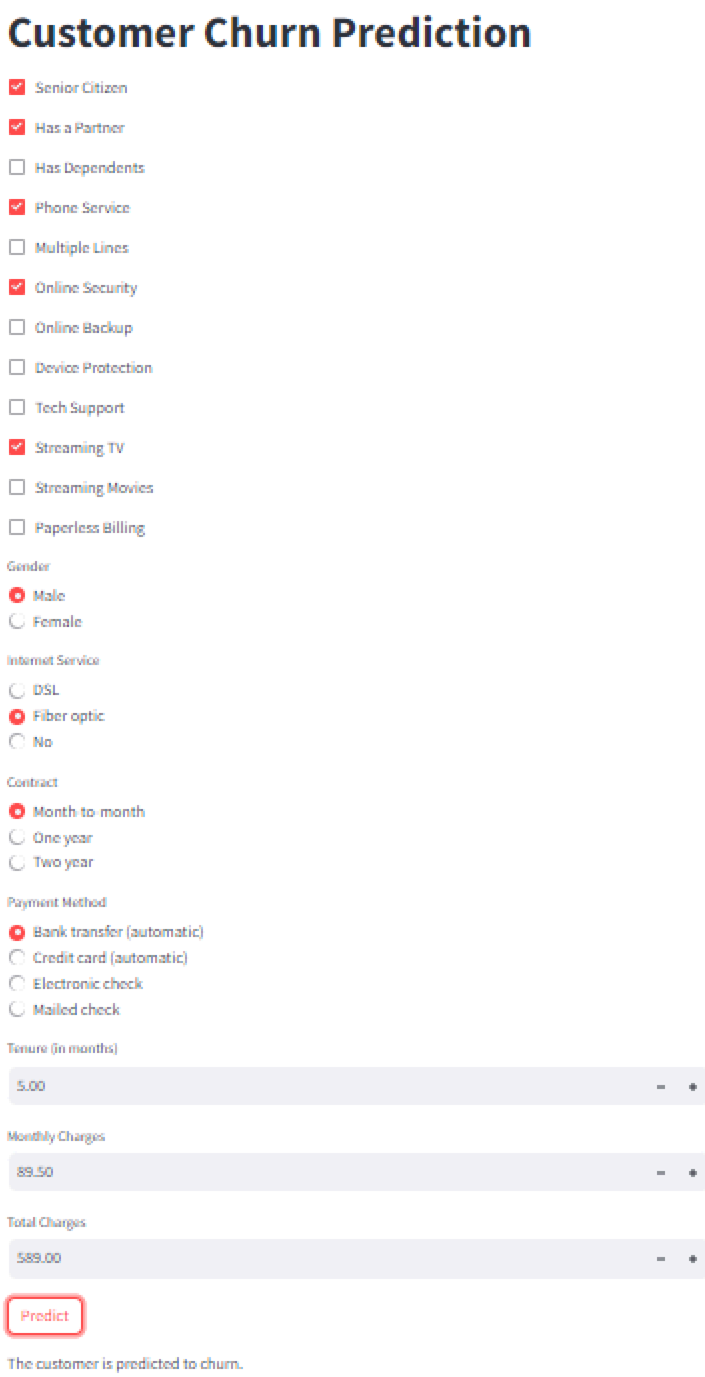


Figure 4.12: Example predicted of churn

5. Discussion and Conclusion

5.1. Achievements

The project has successfully achieved its primary objective of developing a churn prediction application that is able to accurately predict customer churn in the telecommunications company. Through the analysis of historical customer data and behaviors, including an exploratory data analysis (EDA), the project identified key patterns and variables related with possible churn. This in-depth investigation discovered previously unseen information and enabled telecom companies to understand the factors that cause churn. This recognition has given telecom companies important information to strongly interact to retain customers at-risk.

Furthermore, the project assisted in early action by telecom companies to effectively engage and retain at-risk customers. By using advanced analytics techniques, such as predictive modeling, the project provided telecom companies with useful resources to reduce churn risks and maximize customer retention strategies. This method has significantly improved the churn management capabilities of telecom companies, allowing them to make informed, data-driven decisions.

In summary, the project has effectively fulfilled the objective by developing a powerful churn prediction application for the telecommunications industry and this project's achievements have been turned into real-world advantages for telecom companies, including reduced churn rates and maximized customer lifetime value. By using the information obtained from the churn prediction application, telecom companies may reduce churn risks, create stronger customer relationships, and ultimately ensure maintained business growth and competitiveness in the telecommunications market.

5.2. Limitations and Future Works

First, One of the limitations of this project is the use of publicly available datasets from the internet. Even while these datasets provided helpful data about customer churn trends, it's possible that they missed some of the details and unique characteristics of each telecom company's customer. As a result, the predictive models trained on these datasets may not be directly applicable to all telecom companies, especially those with different service offerings or target demographics.

In future versions of this project, it would be helpful to collaborate directly with telecom companies to obtain private data specific to their customer base. The prediction models can be customized to better capture the unique characteristics and dynamics of each company's customer churn behavior by using company-specific data for model training. This approach could improve the accuracy and applicability of the churn prediction application, allowing telecom companies to make more informed decisions and implement targeted retention strategies customized to their specific customer segments. Furthermore, ongoing collaboration with telecom providers would enable constant model improvement and adjustment to changing customer preferences and market conditions.

Next, another limitation of this project is the assumption of stationary patterns in customer behavior. Customer preferences, behaviors, and external market factors could change over time, leading to changes in churn patterns. Therefore, the predictive models trained on historical data may become less effective over time if they have not been frequently updated and adjusted to reflect these changes.

In future work, implementing a system for continuous model tracking and updating would solve this limitation. By combining real-time data provides and feedback systems into the churn prediction application, telecom companies can continuously monitor model performance and adjust the predictive models to changing customer behaviors and market dynamics. This approach would ensure that the churn prediction application remains relevant and effective in capturing changing churn patterns, thus improving its long-term usefulness and effects. Additionally, exploring advanced techniques such as online learning algorithms and combination approaches could further improve the flexibility and reliability of the churn prediction models in environments that are changing.

Reference & Source

Provide the sources of the dataset and tool(s) used for development

List the articles or other references you have cited in the text using the APA Referencing system.

Li, W., & Zhou, C. (2020). *Customer churn prediction in telecom using big data analytics*. IOP Conference Series. Materials Science and Engineering, 768(5), 052070. Retrieved from https://doi.org/10.1088/1757-899x/768/5/052070

A. P. Patil, M. P. Deepshika, S. Mittal, S. Shetty, S. S. Hiremath and Y. E. Patil (2017). *Customer churn prediction for retail business*. IEEE Conference Publication | IEEE Xplore. Retrieved from https://ieeexplore.ieee.org/abstract/document/8389557?casa_token=9M88EX4iBqIAAAAA:WS9Nk8NiEPqporeZmJg4DaC5e5hbalO2HU8CpnYJB07nLZZd5OCYSY6zYR7TipQU6fQ2tNn1QCrdClE

Kumar, P., Bhatnagar, R., Gaur, K., & Bhatnagar, A. (2021). *Classification of Imbalanced Data:Review of Methods and Applications*. IOP Conference Series. Materials Science and Engineering, 1099(1), 012077. Retrieved from https://doi.org/10.1088/1757-899x/1099/1/012077

Goel, G., Maguire, L., Li, Y., & McLoone, S. (2013). *Evaluation of Sampling Methods for Learning from Imbalanced Data*. In Lecture notes in computer science (pp. 392–401). Retrieved from https://doi.org/10.1007/978-3-642-39479-9_47

Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). *Customer churn prediction in telecom using machine learning in big data platform*. Journal of Big Data, 6(1). Retrieved from https://doi.org/10.1186/s40537-019-0191-6

Shumaly S., Neysaryan P. and Guo Y. (2020). *Handling class imbalance in customer churn prediction in telecom sector using sampling techniques, bagging and boosting trees*. IEEE Conference Publication | IEEE Xplore. Retrieved from https://ieeexplore.ieee.org/abstract/document/9303698?casa_token=nb9FIYBu1EUAAAAA:MrpGO_4xXNSxzgXaRUYp-FSFr1q96z_DCASAv6eTPu6BAwUnuZ0NgKaSkeZHLx5dDGb2OYMzui1BfnQ

Kimura, T. (2022). *Customer churn prediction with hybrid resampling and ensemble learning*. Journal of Management Information and Decision Sciences, 25(1), 1-23. Retrieved from https://www.researchgate.net/profile/Takuma-Kimura-3/publication/360287935_Customer_Churn_Prediction_with_Hybrid_Resampling_and_Ensemble_Learning/links/626d6b91d49fe200e1c99823/Customer-Churn-Prediction-with-Hybrid-Resampling-and-Ensemble-Learning.pdf

Chauhan, N. K., & Singh, K. P. (2022). *Performance assessment of machine learning classifiers using selective feature approaches for cervical cancer detection*. Wireless Personal Communications, 124(3), 2335–2366. Retrieved from https://doi.org/10.1007/s11277-022-09467-7

Amorim, L., Cavalcanti, G. D. C., & Cruz, R. M. O. (2023). *The choice of scaling technique matters for classification performance*. Applied Soft Computing, 133, 109924. Retrieved from https://doi.org/10.1016/j.asoc.2022.109924

Thakker, Z. L., & Buch, S. H. (2024). *Effect of feature scaling pre-processing techniques on machine learning algorithms to predict particulate matter concentration for Gandhinagar,* *Gujarat, India.* International Journal of Scientific Research in Science and Technology, 410–419. Retrieved from https://doi.org/10.32628/ijsrst52411150

Celik, O., & Osmanoglu, U. O. (2019). *Comparing to Techniques Used in Customer Churn Analysis*. Journal of Multidisciplinary Developments. Retrieved from https://www.researchgate.net/profile/Oezer-Celik-2/publication/337103029_Comparing_to_Techniques_Used_in_Customer_Churn_Analysis/links/5dc52a29a6fdcc2d2ffc1a24/Comparing-to-Techniques-Used-in-Customer-Churn-Analysis.pdf

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2021). *Customer churn prediction system: a machine learning approach*. Computing, 104(2), 271–294. Retrieved from https://doi.org/10.1007/s00607-021-00908-y

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. Retrieved from https://doi.org/10.1016/j.simpat.2015.03.003

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

Mustafa, N., Sook Ling, L. & Abdul Razak, S. F. (2021). *Customer churn prediction for telecommunication industry: A Malaysian Case Study.* PubMed Central. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9051585/

Van Den Poel, D., & Larivière, B. (2004). *Customer attrition analysis for financial services using proportional hazard models.* European Journal of Operational Research, 157(1), 196–217. Retrieved from https://doi.org/10.1016/s0377-2217(03)00069-9

Kumar, A., Shankar, R., & Debnath, R. M. (2015). *Analyzing customer preference and measuring relative efficiency in telecom sector: A hybrid fuzzy AHP/DEA study*. Telematics and Informatics, 32(3), 447–462. Retrieved from https://doi.org/10.1016/j.tele.2014.10.003

Bhd, S. P. H. S. (2015, December 28). T*elcos to be more innovative in a saturated market.* Daily Express Retrieved from https://www.dailyexpress.com.my/news/105502/telcos-to-be-more-innovative-in-a-saturated-market/

E Pfeifer, P. (2005). *The optimal ratio of acquisition and retention costs.* Journal of Targeting Measurement and Analysis for Marketing 13(2):179-188. Retrieved from https://www.researchgate.net/publication/233555767_The_optimal_ratio_of_acquisition_and_retention_costs