SEPTEMBER 22, 2024

CUSTOMER CHURN ANALYSIS FOR TELECOMMUNICATIONS COMPANY

FINAL PROJECT REPORT

MANIRAM LUITEL, BHAVESH CHAUDHARY, KUSHAL BASYAL AND MUJTABA ALI KHAN
ADVANCE CONSULTANT SERVICES

Introduction

Project Overview

Despite advance development in the digital technologies, telecommunication company suffers the customer loss issue persistently. To better understand the reason contributing to the loss of customers and find effective strategies for their retention, this project utilized data analytic techniques on customers data from the telecommunication company. The main goal of investigating the customer data is to improve the customer retention in the company by 20%.

Methodology

First, the provided customer data was pre-processed to obtain dataset ready to be used in analytic approach. To perform the clustering of the telecommunication customer and obtain the characteristic of the customer churn, k-mean algorithm was applied on the processed dataset. Similarly, an Artificial Neural Network (ANN) architecture was designed based on the dataset features and data was trained to develop a model that predicted the customer churn. The performance of the ANN model was evaluated using confusion matrix and, other techniques like ROC curve and Precision recall curve.

Data Overview

Data Description

The provided dataset contained ten features and 7043 records of customer data like demographic features, telecommunication services that customer used and their cost. Categorical features like *gender*, *SeniorCitizen* and *Dependents* are customer's demographic attributes whereas categorical features like *PhoneService*, *MultipleLines* and *InternetService* are telecommunication services. Similarly, features like *tenure*, *Contract* and *MonthlyCharges* are cost attributes of the services. On the other hand, customer either leaving the telecommunication service or not i.e. *Churn* is the target variable.

Preprocessing

The preparation step involved initial data cleaning, encoding categorical variable, scaling the dataset and splitting it into training and testing sets. To begin, the provided dataset was loaded and checked for missing values to eliminate them. However, there was no missing values present in the provided dataset. Secondly, all the categorical features were transformed into numerical data using one-hot encoding technique. This step is essential to avoid inaccuracies in analysis and model predictions. The following table clearly represents the encoding applied on the provided dataset.

Feature	Categories	Encoding	Column Name After Encoding
Gender	Female, Male	0, 1	gender_Female, gender_Male
SeniorCitizen	No, Yes	0, 1	SeniorCitizen
Dependents	No, Yes	0, 1	Dependents_No, Dependents_Yes
PhoneService	No, Yes	0, 1	PhoneService_No, PhoneService_Yes
MultipleLines	No, Yes	0, 1	MultipleLines_No, MultipleLines_Yes
InternetService	DSL, Fiber optic, No	0, 1, 2	InternetService_DSL, InternetService_Fiber optic
Contract	Month-to-month, One year, Two year	0, 1, 2	Contract_Month-to-month, Contract_One year, Contract_Two year
MonthlyCharges	(Numerical)	No encoding	MonthlyCharges
Churn	No, Yes	0, 1	Churn_No, Churn_Yes

Next, some new features namely *Charges_Per_Tenure* and *TotalCharges* were engineered based on the discrete data like *tenure* and *MonthlyCharges*. The table shows the engineering applied to obtain new features.

New Features	Engineering
Charges_Per_Tenure	Charges_Per_Tenure = MonthlyCharges /tenure
TotalCharges	TotalCharges = MonthlyCharges' **x** tenure

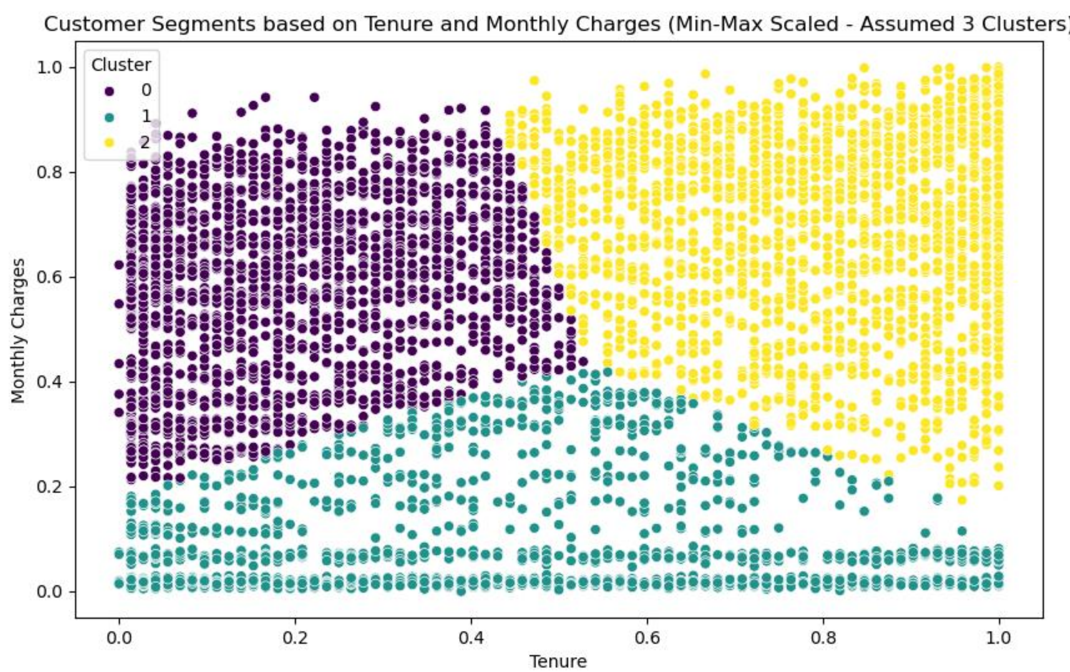
Further, two scaling technique namely Standard scaling and Min-Max scaling were used to transform the dataset for analysis. The standard scale transformed the customer data such that they have a mean of 0 and standard deviation of 1. This scaling ensured that all features remain on the same scale and avoided features like *tenure*, *MonthlyCharges* from dominating the learning process due to its larger magnitude. Similarly, Min-Max scaling transformed the features such that they range between 0 and 1 ensuring that all features value contribute equally.

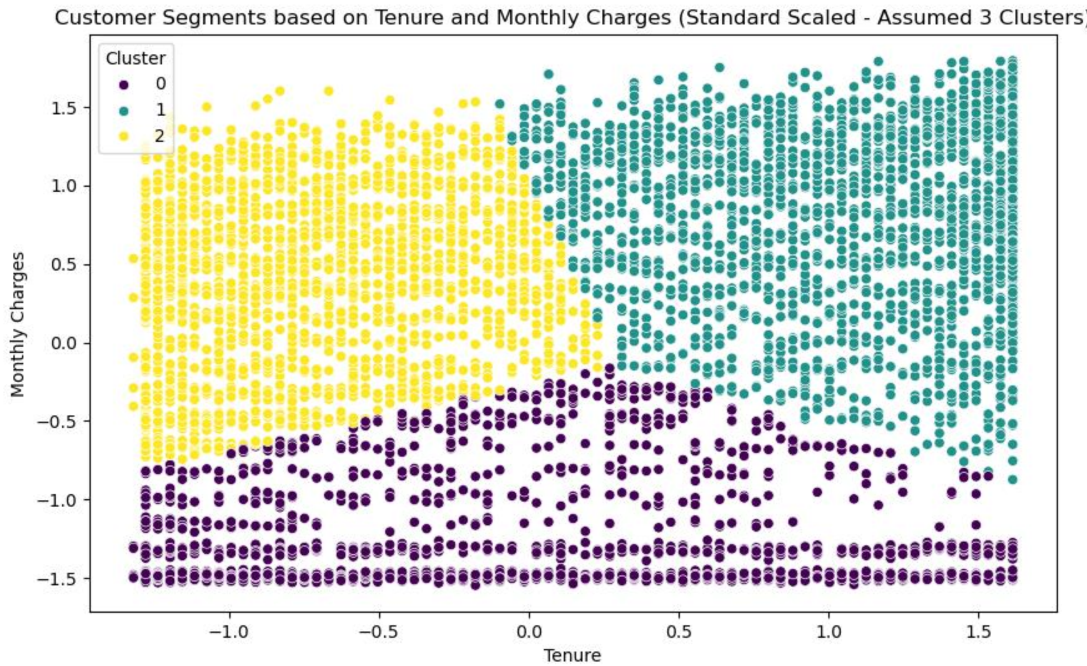
Customer Segmentation

Clustering Approach

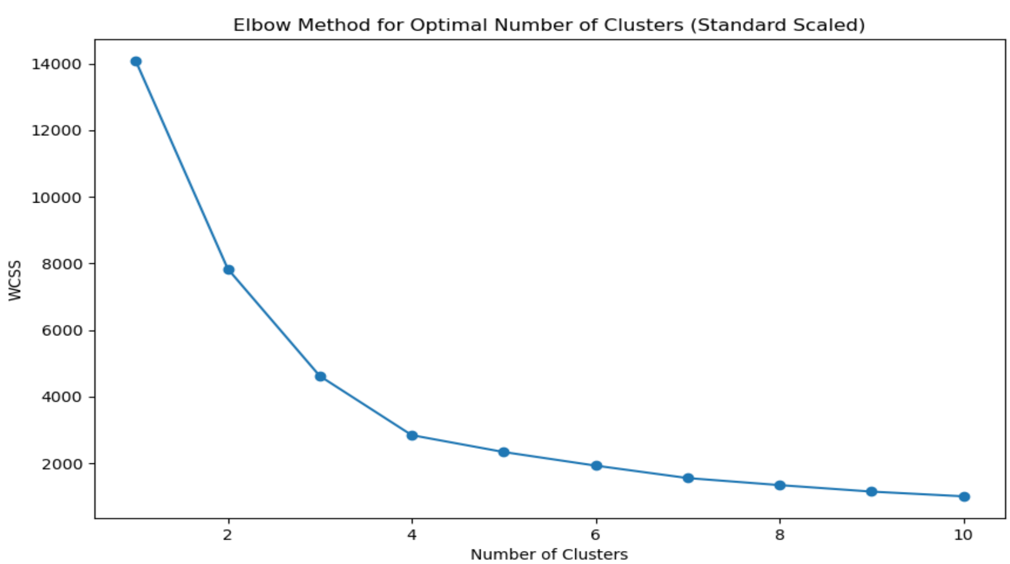
K-Means was chosen as the primary clustering algorithms. The main reason behind choosing K-Means algorithm was its efficiency with large datasets, its ability to create well-separated clusters, and its simplicity and interpretability. K-Means optimizes the centroids by minimizing the variance within clusters, and make sure about distinct and meaningful groupings.

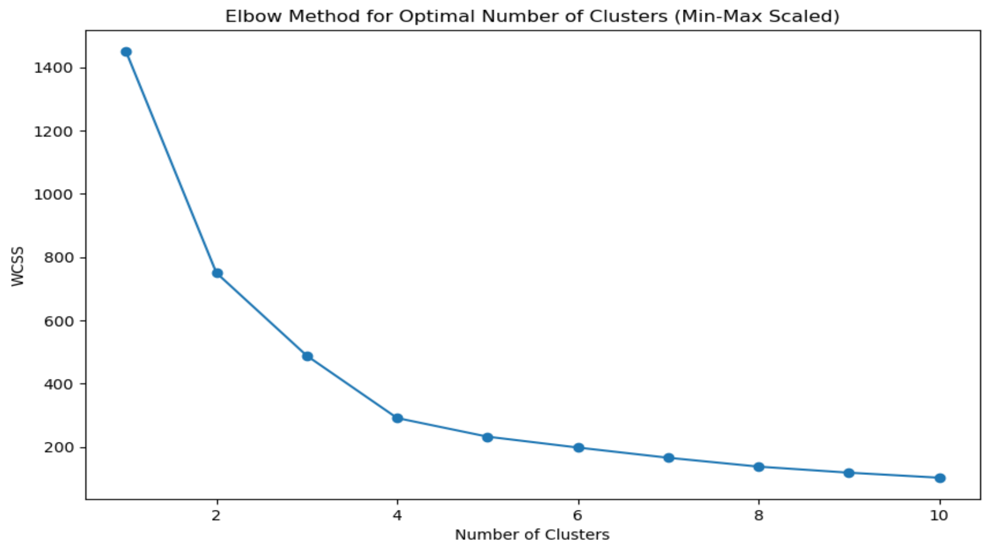
The analysis started by applying the K-Means algorithms, assuming the number of clusters equal to 3. The K-Means clustering was applied in both min-max scaled and standard scaled datasets on customer segmentation based on the features tenure and monthly charges. The clusters for the scaled dataset were visualized, providing different insights into different customer segments.



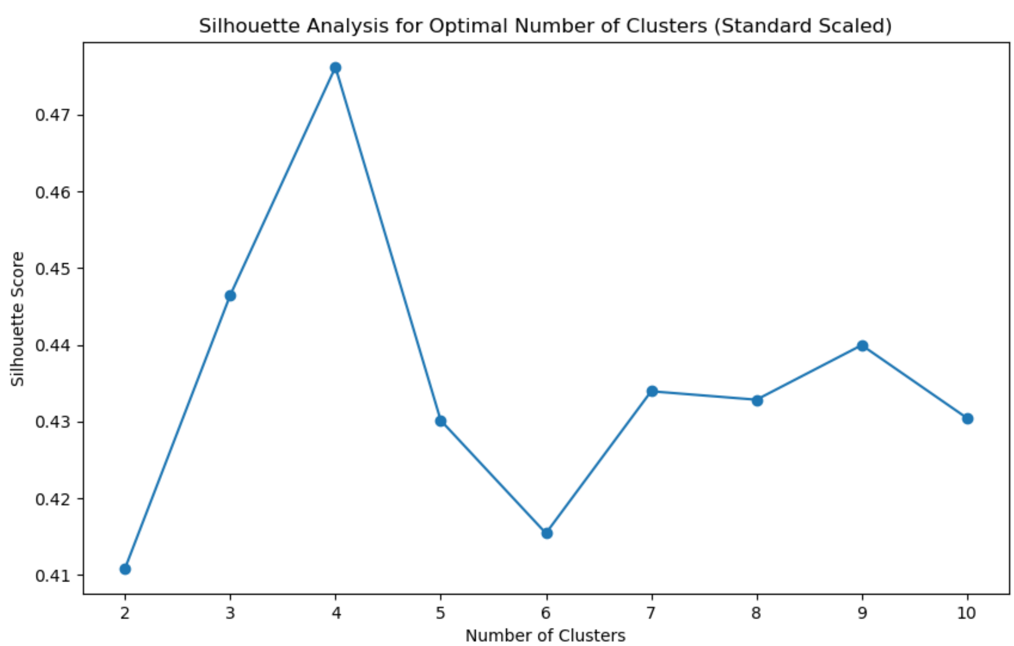


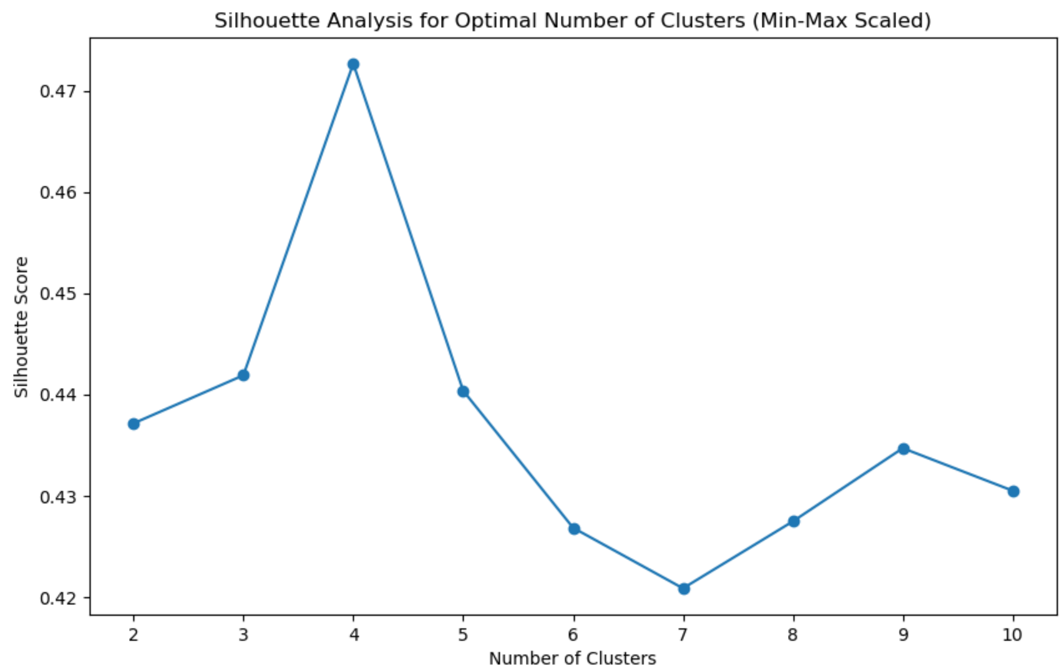
Next approach utilized the elbow method to plot sum of squared distance between each data point and its assigned cluster centre, identifying the point where the curve starts to level off, indicating the optimal number of clusters. The elbow points where additional clusters provide minimal improvement in reducing variance, ensuring the model remains both interpretable and effective without overfitting.



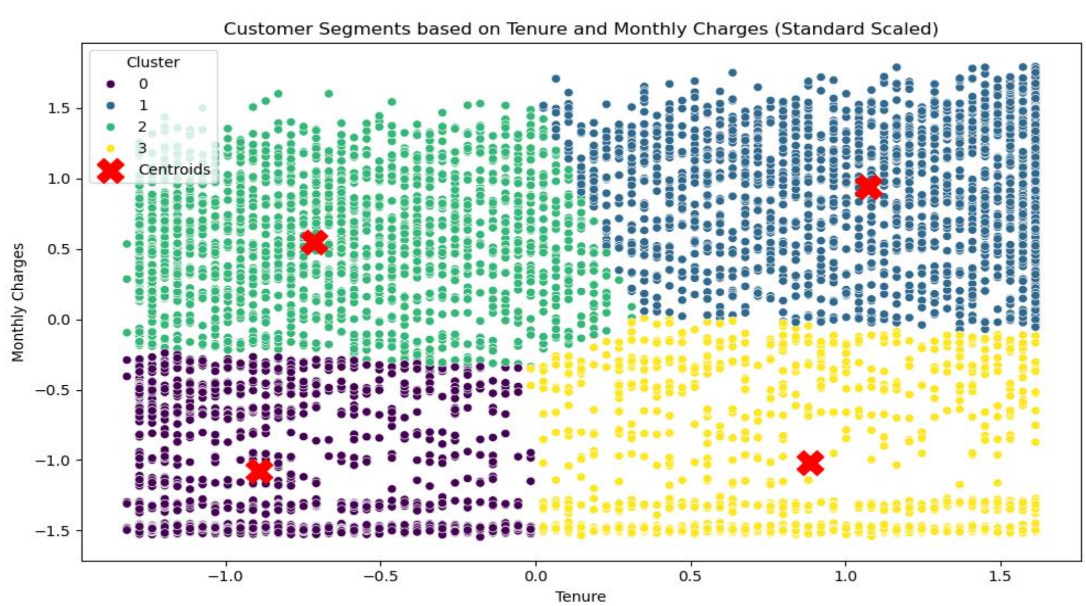


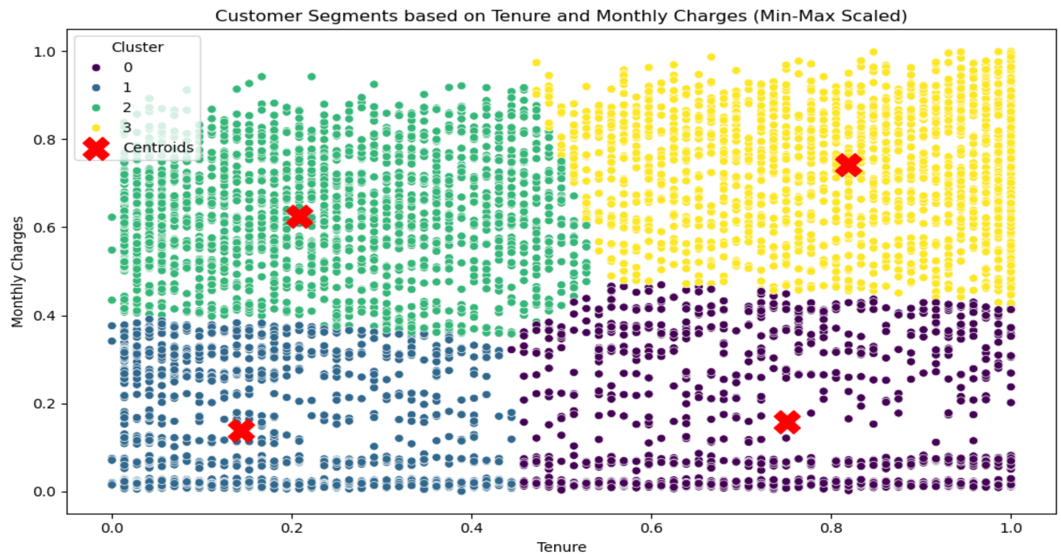
Silhouette Analysis was implemented to evaluate the cohesion and separation of clusters, offering insights how well each data points fits within its assigned clusters compared to others. A higher silhouette score indicates that clusters are well-defined and distinct, making it a valuable metric in determining the optimal number of clusters. By obtaining these scores, it was ensured that the clusters were not only compact but also clearly separated, enhancing the model accuracy and interpretability. This analysis was crucial in validating the chosen clusters, reinforcing the overall robustness of the clustering solution.





Based on these both methods, the optimal number of clusters was determined as 4. The consistency across both the scaling methods suggests that the natural grouping within the data is robust to the type of scaling used. Now by using determined number of clusters, we trained the clustering model on the dataset by using K-Means algorithm. Clustering the customers into 4 groups, the segments were identified based on their tenure and monthly charges.





Segment Profiles:

Careful observation of the characteristics of each cluster by analysing the mean and median values of the key features tenure and monthly charges, lead to a clear understanding of the distinct customer segments. Based on this analysis, it was confirmed that certain clusters represent customers with high tenure and high charges, while the others are characterized by low tenure and low charges.

Based on distinct characteristics, each cluster was labelled to reflect the specific customer's profile. For instance, each cluster identified as high tenure and high charges indicating the group of long term, premium customers. While another cluster labelled low tenure and low charges representing customers with basic plans. These labels provide the summary of the customers, making easier to tailor business strategies to each group. The following table provides detail on the identified segments of the customer.

Cluster	**Min-Max Scaled**	**Standard Scaled**	**Characteristics**
High Tenure, High Charges (Premium Customers)	Cluster 2	Cluster 1	Long-term, high-value customers likely subscribed to premium plans. Crucial for retention and upselling opportunities.
Low Tenure, Low Charges (New or Basic Customers)	Cluster 3	Cluster 0	New or lower-value customers. Prime targets for engagement strategies aimed at increasing value through promotions and upsell opportunities.

High Tenure, Low Charges (Loyal but Economical Customers)	Cluster 0	Cluster 3	Loyal customers who have opted for more economical plans over time. Potential to increase lifetime value through premium services or rewards.
Moderate Tenure and Charges (Mid-Tier Customers)	Cluster 1	Cluster 2	Mid-tier customers with stable tenure and charges. Potential for growth through targeted marketing and value-added services.

Churn Prediction Model

Model Selection

Artificial Neural Network (ANN) was selected as primary algorithm. The main reason behind choosing ANN is that the layers of interconnected neurons excel in capturing non-linear patterns in data. As features like telecommunication services, charge, tenure and customer demography have non-linear relation with churn outcome, ANN was used to evaluate the customer churn.

Model Training

To begin with, the architecture of the ANN model was defined which consists of the input layer, hidden layers, output unit and activation functions that define how it sequences data to make predictions. The input layer comprised customer-relevant attributes, such as demographic data or the type and duration of service usage. Also, the input shape was equal to the number of features in dataset. This made sure that all the features are treated and used to predict the input layer so that it has one neuron for each column in our dataset, meaning that this provides enough data to supply all its neurons with inputs.

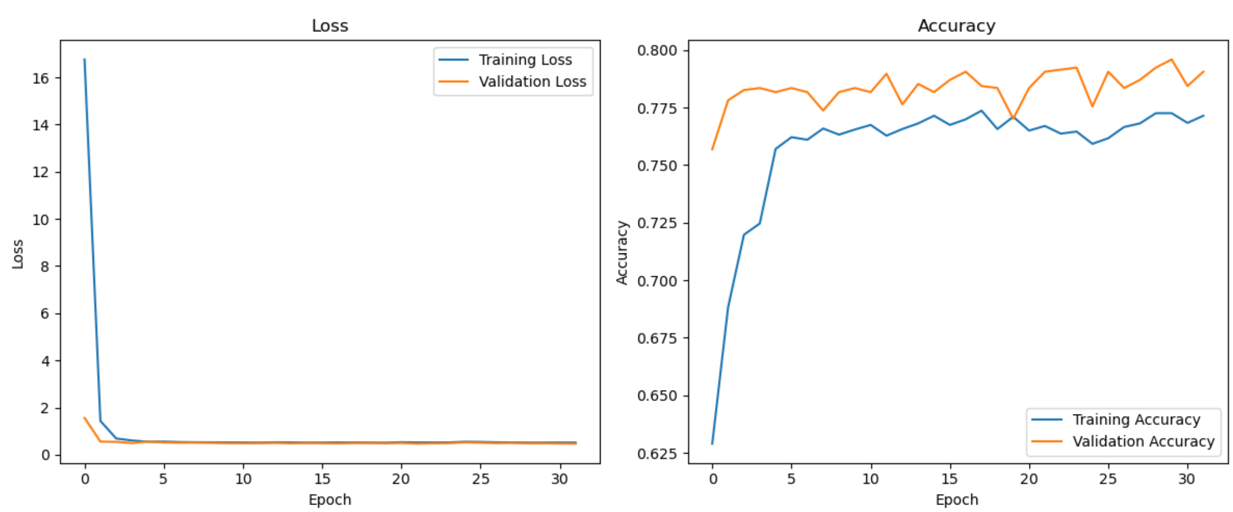
Next, the actual calculations and learning in the neural network was done by hidden layers, which had two layers. The first hidden layer consisted of 64 neurons. This layer is designed to capture the initial complexity of the data by learning high-dimensional feature representations. Similarly, the second layer consisted of 32 neurons, serving to refine the learned features and reduce dimensionality while preserving important information. Both layers used Rectified Linear unit (ReLU) activation function. ReLU was used in both layers as it was computationally efficient and helped introduce non-linearity into the model, allowing it to capture complex patterns in the data.

The output layer produced the final predictions. As it was a binary classification problem, a single neuron was used in the output layer to represent the binary outcome whether the customer churn or not. Using Sigmoid Activation Function, the output was produced between 0 and 1 which was interpreted as the probability that a customer will churn. Sigmoid activation was chosen because it is right for binary classification task. Based on the probability threshold at 0.5 to predict final class, the model predicted if a customer would churn or not. Finally, combined with Adam optimizer and binary cross-entropy loss, the model was trained to recognize data patterns and achieve less error in prediction.

The Adam optimizer was chosen for its efficiency and adaptive learning rate capabilities, which help in converging to an optimal solution faster and more reliably than standard gradient descent. Binary Cross-Entropy was used as the loss function, which is suitable for binary classification problems like churn prediction. It measures the dissimilarity between the predicted probability and the actual class label (churn or no churn), guiding the optimization process to minimize this difference.

In the training process, the ANN model was trained using the pre-processed training dataset to predict customer churn. Training involves adjusting the weights of the model to minimize the error between predicted and actual outcomes. The model was trained over 50 epochs, which was determined to be sufficient for the model to learn from the data without excessive training that could lead to overfitting. And a batch size of 32 was used, meaning the model processed 32 samples at a time before updating its weights. This size was chosen to provide a balance between the efficiency of training and the stability of the model's learning process.

Model Performance



The training of ANN model was conducted over 50 epochs with early stopping to prevent overfitting. During training, the model demonstrated a steady improvement in both training and validation accuracy.

Training Accuracy: The model achieved a training accuracy of approximately 79% by the 20th epoch. The accuracy continued to improve during early epochs, but early stopping halted the training after no significant improvement was observed beyond this point.

Training Loss: The training loss consistently decreased during the training process. The final training loss stabilized at around 0.44, demonstrating that the model was able to effectively minimize the error between predictions and actual outcomes.

Validation Results: Throughout the training, the model's validation accuracy also improved but at a slower rate compared to the training accuracy. The validation loss showed some fluctuations but generally decreased. The final validation accuracy was 78%, and the validation loss was 0.45. These figures indicate that the model was able to generalize reasonably well without significant overfitting.

Early Stopping: The model employed early stopping with a patience of 10 epochs, halting the training at epoch 20. This approach ensured that the model did not continue training beyond the point where it stopped improving on validation data.

Metrics: Accuracy was selected as the primary evaluation metric during training to monitor how well the model correctly classifies customers. Additional metrics such as Precision, Recall, and F1 Score were later computed to provide a comprehensive assessment of the model's performance, especially in handling imbalanced classes often present in churn datasets.

ROC-AUC (Receiver Operating Characteristic - Area Under Curve): The ROC-AUC score evaluates the model's ability to distinguish between classes. A higher AUC indicates better performance across various threshold settings.

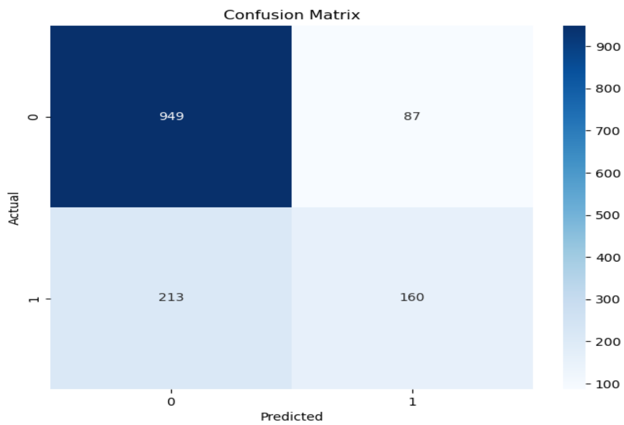
Evaluation on the Test Dataset:

Precision, Recall, and F1-Score:

- The precision score of approximately 0.6478 indicates the proportion of true positive predictions among all positive predictions made by the model. High precision is important in contexts where false positives are more costly.
- The recall score of approximately 0.4290 reflects the proportion of actual positives that were correctly identified by the model. High recall is crucial in scenarios were missing a positive case (false negatives) is more detrimental.

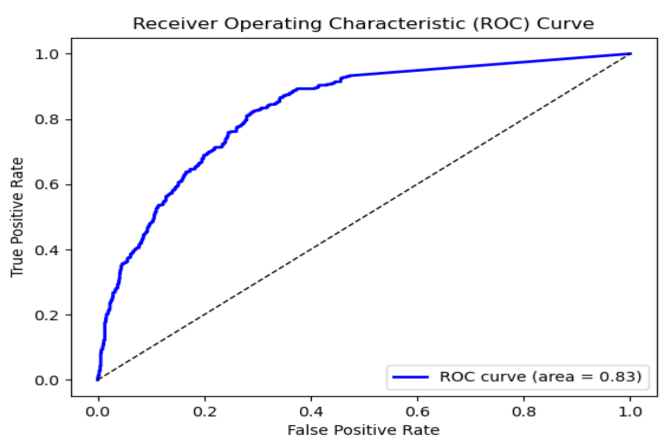
- The F1 score of approximately 0.5161 provides a balance between precision and recall, especially useful when the data has imbalanced classes. It helps gauge the model's effectiveness in handling both false positives and false negatives**.**

Confusion Matrix: The confusion matrix revealed how well the model distinguished between churn and non-churn customers:



- True Positives (TP): 160 Correctly predicted churn customers.
- True Negatives (TN): 949 Correctly predicted non-churn customers.
- False Positives (FP): 87 Non-churn customers incorrectly predicted as churn.
- False Negatives (FN): 213 Churn customers incorrectly predicted as non-churn.

ROC Curve: The ROC curve visualizes the trade-off between true positive rate (recall) and false positive rate. The AUC was observed to be 0.79, indicating good model discrimination between churn and non-churn customers.



Key Findings and Recommendations

The main aim of this report is to provide the stakeholders with comprehensive understanding of the key factors influencing customer churn and retention. This report highlights actional insights to enhance retention and helps in the business growth. It is very important to understand the factors leading to customer churn for developing the essential improve retention. The following are the essential factors identified:

1. **Customer Tenure**

   Churn: Overall, the likelihood of churning is higher for customers staying with the company for a relatively short period. That is reflected in the model performance metrics, where higher false negatives are associated with shorter tenure, meaning the model often misses churn predictions for these customers.

   Confusion Matrix Data:

   True Positives (TP): 160

   False Negatives (FN): 213

   Implication for Strategy: Strategies should be focussed on retaining new customers by minimizing their early attrition. This might be achieved through the enhancement of onboarding processes and continuous support to truly engage these customers.

2. **Monthly Charges**

   Churn: It can be seen from the model that the higher the monthly charges, the more is the chance of churning. Many false positives indicates that the model sometimes incorrectly predicts the higher-paying customers as churning.

   Confusion Matrix Data:

   TP: 160

   FP: 87

   Strategy Implication: A company is better off reassessing and probably revising the pricing strategies in the face of probable churning by higher monthly charges customers. The companies can retain such high-value customers by offering flexible pricing plans or giving better discounts so that the number of incorrect predictions of customer churning is cut down.

3. **Contract Type**

Churn: The customers who have a month-to-month contract have a higher tendency towards churn. From the performance metrics of the model, it can be gathered that a month-to-month contract leads to higher true positives, which correctly identify a customer as a churner.

Confusion Matrix Data:

True Negatives TN: 949

False Negatives FN: 213

Implication for Strategy: Companies are required to encourage longer contract commitments to stabilize the customer base. In the case of reward or special incentives when customers commit to longer contracts, companies succeed in reducing their churn rates and increasing customer retention through further stability.

4. **Service Subscription**

Churn: Customers who have subscribed to more than one service have fewer chances of churning. Higher precision in predicting non-churners for people who have more than one service on subscription suggests that these people will not churn.

Precision Data:

Precision: 64.78%

Implication for Strategy: Companies should market bundled services or the selling of additional subscriptions to improve retention. By educating clients about the value of multiple services, companies can offer discounted rates for bundled subscriptions that will help them retain a more loyal customer base.

5. **Senior Citizen Status**

Churn: The model has a significant impact on senior citizens' churn predictions. While the model identifies a major portion of senior citizens as cases of churn, its recall rate is low, therefore missing many in this demographic group.

Recall Data:

Recall 42.90%

Strategic Implication: Companies should develop focused retention strategies for senior citizens in the form of personalized offers, additional support, or tailored engagement efforts that may contribute to an improved retention rate in this group and reduce missed cases of churn.

Limitations and Challenges

The entire project had to face many limitations and challenges that were encountered at different stages, starting from data collection and moving all the way up to predictive modelling. The primary challenges that we encountered were in the quality of the dataset i.e. inconsistencies, missing values and outliers throughout the data had a major influence on its dependability. There issues would demand a lot of cleaning and verification, for the dataset to be as clean as possible. So, in this step, the missing values are filled to Features like Tenure, MonthlyCharges, SeniorCitizen; as we know that Missing Value of these Features can cause a strong bias in downstream analyses.

The application of coding and normalisation also presented difficulties-extension. There was no uniformity in data interpretation because different analyses required different approaches - standard scaling, min-max scaling. This made the analysis of results from different models more convoluted and influenced one's overall take of experiments.

We faced significant challenges with model overfitting, particularly while building predictive models especially complex ones like Artificial Neural Networks (ANN) The models performed well on the training data but failed miserably when tested on new unseen data, it was a clear overfit to the training set. This problem, in turn, reduced the robustness of the model to allow correct prediction on new data (completely unseen before).

Furthermore, the target variable being imbalanced (Churn) only made the problem worse. Most of the data were samples from the vast class (customers who didn't churn), so the model leaned greatly towards predicting such category and detected eventually badly on minority classes (costumers possible to churn). Due to this imbalance, we got above results for the metrics that accuracy is very high, but it does not represent true prediction of model.

Model evaluation metrics posed an additional challenge. While traditional metrics like accuracy showed promising results, they did not provide an accurate picture due to the class imbalance. Precision, recall, and AUC scores were needed to better assess the model’s performance, particularly in predicting the minority class. These alternative metrics revealed that while the model could predict the majority class well, it struggled with correctly identifying customers who were likely to churn.

These challenges necessitated a combination of advanced techniques, from data resampling and model regularisation to rigorous validation processes. Despite the issues, these adaptive strategies ensured the project’s findings remained robust. The following sections will explore these challenges in detail, outlining their implications and the solutions implemented to overcome them, with the goal of improving future processes and outcomes.

Key Challenges and Mitigated Solutions

Challenges	**Implemented Solutions**
Data Quality Issues	
• Missing values, inconsistencies, and outliers	• Automated data cleaning pipelines to handle inconsistencies and missing values
• Data entry errors and unstandardized methods	• Regular data validation checks and audits were conducted to ensure data consistency
Scaling and Normalization Issues	
• Different analyses required different scaling methods	• Unified scaling approach was established (used both standard and min-max scaling as required by analysis)
Cluster Interpretability and Validation Issues	
• Disagreement between validation methods (silhouette vs. elbow)	• A combination of silhouette analysis and elbow method was used for validation
• Clusters sensitive to initial parameters	• Multiple clusters run with varied initializations were performed to find stable clusters
Model Overfitting	
• ANN models performed well on training data but poorly on new data	• Regularization techniques (dropout and batch normalization) were applied during model training
• Complex models required larger datasets	• The model architecture was simplified to reduce complexity and the risk of overfitting
Low Data Size	
• The dataset size was small, which limited the model's ability to generalize effectively	• Data augmentation techniques were applied to increase the size of the training data
Imbalanced Data	
• Target variable Churn highly imbalanced	• SMOTE (Synthetic Minority Over-sampling Technique) was applied to balance the dataset
• Models biased toward the majority class	• Class weights were adjusted in the model to handle minority class (churn) more effectively
Team Member Departure and Role Redistribution	
• Departure of the Business Analyst led to gaps in documentation and communication	• Business Analyst tasks were redistributed among the team members

Project Limitations

Despite the steps taken to address the challenges, certain limitations were encountered throughout the project that constrained the final outcomes:

1. **Data Size:**

   They used a rather small dataset which restricted the model capacity for any generalization to unseen data (i.e., ANN). Better insights and model accuracy could have come by upgrading to larger datasets.

2. **Time Constraints:**

   The project was a relatively short-lived one, they could not spend months experimenting with different models or tuning technique and feature engineering that may have improved the model.

3. **Role Redistribution:**

   In the absence of one Business Analyst, which was leaving the team (the key person for such activities in this team) other members were forced to take over his responsibilities. Although this provided stability, it also meant that some BAs take on the responsibilities for which we had hoped to have a dedicated expert and were therefore covered by other members of the team before.

4. **Imbalanced Data:**

   Even with the use of resampling techniques to deal with dataset imbalance (Churn), some amount of problem was present, and this affected model predictions for minority classes in general.

5. **Model Overfitting:**

   Overfitting was a particular concern in the prediction modelling stage, especially for the ANN model despite applying regularization techniques. This restricted the model to be utilized on new data.

6. **Variability: Scaling and Normalization**

   Scaling and normalization methods were different for clustering and predictive modelling, consequently making the final interpretations slightly more challenging to equate across models.

7. **Cluster Validation Uncertainty**:

   Despite using multiple validation methods such as the elbow method and silhouette analysis, determining the optimal number of clusters remained uncertain. Different methods produced conflicting results, and further expert review may have been beneficial to confirm the final number of clusters.

Conclusion

The analysis conducted in this project provided valuable insights into customer churn behaviour within the telecommunications industry. By utilizing both clustering techniques and predictive modelling, we were able to segment customers effectively and predict their likelihood of churning. Key findings highlighted the importance of factors such as tenure, monthly charges, contract type, service subscription, and senior citizen status in determining churn risk.

The implementation of an Artificial Neural Network (ANN) model achieved reasonable accuracy in predicting customer churn, but there were limitations related to data size, imbalance, and model overfitting. Despite these challenges, the project successfully identified actionable strategies to improve customer retention, such as targeting new customers with enhanced onboarding processes, revising pricing strategies for high-value customers, and encouraging longer contract commitments.

Future work could focus on expanding the dataset, applying more sophisticated techniques to address class imbalance, and experimenting with different model architectures to further improve prediction accuracy. Additionally, further exploration of customer behaviour trends and testing retention strategies in a real-world setting will provide more concrete recommendations for reducing churn and improving business growth.