### Title

# Enhancing Customer Retention through Predictive Analytics A Comprehensive Study on Customer Churn Prediction

Table of Contents

[1. Introduction 2](#_Toc177228974)

[1.1 Problem Statement 2](#_Toc177228975)

[1.2 Project Objectives 3](#_Toc177228976)

[1.3 Importance of Customer Retention in the Telecommunications Industry 3](#_Toc177228977)

[2. Literature Review 3](#_Toc177228978)

[3. Data Overview 5](#_Toc177228979)

[3.1 Dataset Description 5](#_Toc177228980)

[3.2 Data Sources 5](#_Toc177228981)

[3.3 Data Cleaning and Preprocessing 6](#_Toc177228982)

[3.4 Summary Data 6](#_Toc177228983)

[EDA 7](#_Toc177228984)

[Churn Distribution Analysis: Insights by Gender, Contract Type, and Internet Service 7](#_Toc177228985)

[Senior Citizen Distribution by Churn: 8](#_Toc177228986)

[Distribution of Monthly Charges: 9](#_Toc177228987)

[Tenure by Churn: 9](#_Toc177228988)

[Distribution of Contract Types: 10](#_Toc177228989)

[Average Monthly Charges by Month: 11](#_Toc177228990)

[K-Means Clustering 11](#_Toc177228991)

[Elbow Method for Optimal Number of Clusters: 12](#_Toc177228992)

[K-Means Clustering Result 13](#_Toc177228993)

[Summary Result 13](#_Toc177228994)

[Predictive model 14](#_Toc177228995)

[Result 15](#_Toc177228996)

[ROC Curve and AUC 16](#_Toc177228997)

[Classification Report Heatmap 16](#_Toc177228998)

[Recommendation 17](#_Toc177228999)

[Conclusion 17](#_Toc177229000)

[References 18](#_Toc177229001)

# Introduction

In today world and face paced with highly competitive business environment the customer retention has become a corseting of long term and success with particularly in the telecommunication industry and organization with the rapid good growth of technology and a abundance of option available to consumer telecommunication companies face the significant challenges in the retaining their customer process and high customer churn rates can have detrimental effects on a profitability as the most cost of acquiring with new customer often and exceeds the cost of treating the existing ones

The project that look explore the factors that drive a customer churn within the telecommunication process and develop the data driven in strategies to improve the retention rate By employing the advance s and analytics and machine learning techniques like clustering and predictive model and many more this study would find the key behavioral patterns in customer segment at the risk of churn and the underlying the main reason for dissatisfaction. This insight will gained and enable the telecommunication that providers to designer the targeted retention programs increase customer satisfaction and strengthen their competitive edge process

### Problem Statement

The primary main problem that addressed in this analysis which is high churn rate from all customer of a telecommunications organization that leading to lost the revenue and increasing the acquisitions costs and Despite the efforts to increase the service quality and customer engagement with a significant number of customer can continue to leave the service by opting for competitors. Understanding these factors driving this churn to finding the patterns that can help to predict customer departures in crucial for the each company to implement effective retention strategies

# Project Objectives

* **To analyze** customer data to identify key factors contributing to churn.
* **To segment** customers using clustering techniques to group similar behaviors.
* **To develop** a predictive model for identifying at-risk customers with high accuracy.
* **To provide** actionable insights for targeted customer retention strategies.
* **To recommend** data-driven interventions to reduce churn rates and improve customer loyalty.

### Importance of Customer Retention in the Telecommunications Industry

Customer retention is a critical with metric and in the telecommunication industry due to the high cost of acquiring the new customer that compared to retaining existing once in these kind of saturated market it is often more cost effective to focus on the nurturing and arranging with maintain relationship with current customer than constantly pursing the new one and loyal customer that tend to generate more revenue over time through upwelling and cross selling opportunities by Understanding churn dynamic and proactively addressing them can lead to reduces customer attrition

* 1. **Tools and Technologies**

In this project a variety of tools and technologies have been employed to facilitate efficient data analysis and visualization Python serves as the primary programming language, utilizing libraries such as **Pandas** for data manipulation, **NumPy** for numerical computations, and **Matplotlib** and **Seaborn** for creating insightful visualizations. For machine learning and predictive modeling, **Scikit learn** has been used, incorporating algorithms like clustering and regression analysis. The **Elbow Method** is implemented to determine the optimal number of clusters, while **StandardScaler** is applied to normalize data for model consistency. **Jupyter Notebook** is used for interactive coding and seamless presentation of results.

### Literature Review

Customer Churn prediction that has been a critical areas of focus in the telecommunication industry because of its direct impact and revenue and portability As competition increases the ability to retain customer which becomes a significant challenges for telecom operator and Numerous studies have explored the application of machine learning algorithms to predict customer churn and allowing proactive measure to improve retention rates

Rajendran and Devarajan (2023) discuss the use of various machine learning algorithms to construct a churn prediction model. Their research highlights the importance of predicting customer churn in the telecom industry and compares multiple techniques. Among the tested models, the Random Forest algorithm combined with the SMOTE ENN technique yielded the highest performance with an F1 score of 95%. This demonstrates the effectiveness of ensemble models when coupled with data balancing techniques for improving predictive accuracy .

Similarly, Kumar and Chandrakala (2016) conducted a comprehensive survey on customer churn prediction using machine learning techniques across multiple sectors, including telecom. Their review emphasizes the increasing importance of retaining customers due to the rising cost of customer acquisition. The study identifies commonly used algorithms such as decision trees, support vector machines, and neural networks, noting that telecom operators benefit most from churn prediction by applying advanced machine learning techniques

Prabadevi et al. (2023) extended this research by focusing on early client churn prediction. They experimented with four algorithms stochastic gradient booster, random forest, k nearest neighbors (KNN), and logistic regression the random forest algorithm again proved effective, with a prediction accuracy of 82.6%, followed closely by stochastic gradient booster at 83.9%. Their study underscores the importance of early detection of churn, allowing businesses to take preventive actions to retain customers

Dahiya and Bhatia (2023) approached churn prediction by developing a framework using WEKA the data mining software their research tested decision tree and logistic regression models, comparing their efficiency in predicting customer churn in the telecom industry their findings suggest that decision tree algorithms which are more interpretable and provide clearer decision rules that can be particularly useful for telecom operators seeking to understand the underlying factors contributing to churn .

Overall, the literature consistently highlights the value of machine learning in predicting customer churn, with ensemble models like Random Forest frequently emerging as top performers early detection and the application of various techniques, including decision trees and logistic regression that provide valuable insights into churn behavior, helping telecom companies design effective retention strategies

# Data Overview

* 1. Dataset Description

The dataset that consist the 7043 customer records and each has the 10 columns that showing the various customer attributes and these attributes which include the demographic information like gender and SeniorCititzen and Dependents, the service related details are tenure, Phone, Mutiplelines with internet and the target variable is Churn that indicate wheatear the customer has left the service. The data is a mix of different categorical numerical variable process which providing the overview of customer profiles and their service which they has used

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **gender** | **SeniorCitizen** | **tenure** | **MonthlyCharges** | **Churn** | **Dependents\_Yes** | **PhoneService\_Yes** | **MultipleLines\_Yes** | **InternetService\_Fiber optic** | **Contract\_One year** | **Contract\_Two year** |
| 0 | 0 | 0 | -1.292434 | -1.182533 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0.056731 | -0.276264 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 2 | 1 | 0 | -1.251550 | -0.379933 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 1 | 0 | 0.506452 | -0.766184 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 4 | 0 | 0 | -1.251550 | 0.183559 | 1 | 0 | 1 | 0 | 1 |  |  |

# Dataset Information

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **Gender** | Represents the gender of the individual (e.g., Male, Female, Non-Binary). |
| **Age** | Represents the age of the individual, typically in years. |
| **Tenure** | Represents the length of time (in years) the individual has been with the company or in a specific role. |
| **Job Role** | Represents the individual's position or title within the organization (e.g., Manager, Analyst, Engineer). |
| **Salary** | Represents the individual's annual salary or monthly income, typically in the local currency. |
| **Department** | Represents the department or division the individual works in (e.g., Marketing, Finance, HR). |
| **Education Level** | Represents the highest level of education the individual has completed (e.g., High School, Bachelor’s Degree). |
| **Performance Rating** | Represents the individual's performance score, typically on a scale (e.g., 1-5) based on annual evaluations. |
| **Satisfaction Score** | Represents the individual's satisfaction level with their job, rated on a scale (e.g., 1-5). |
| **Attrition** | Represents whether the individual has left the organization (Yes/No or 1/0). |

This table organizes the information clearly, allowing for a quick understanding of the dataset's structure and the features

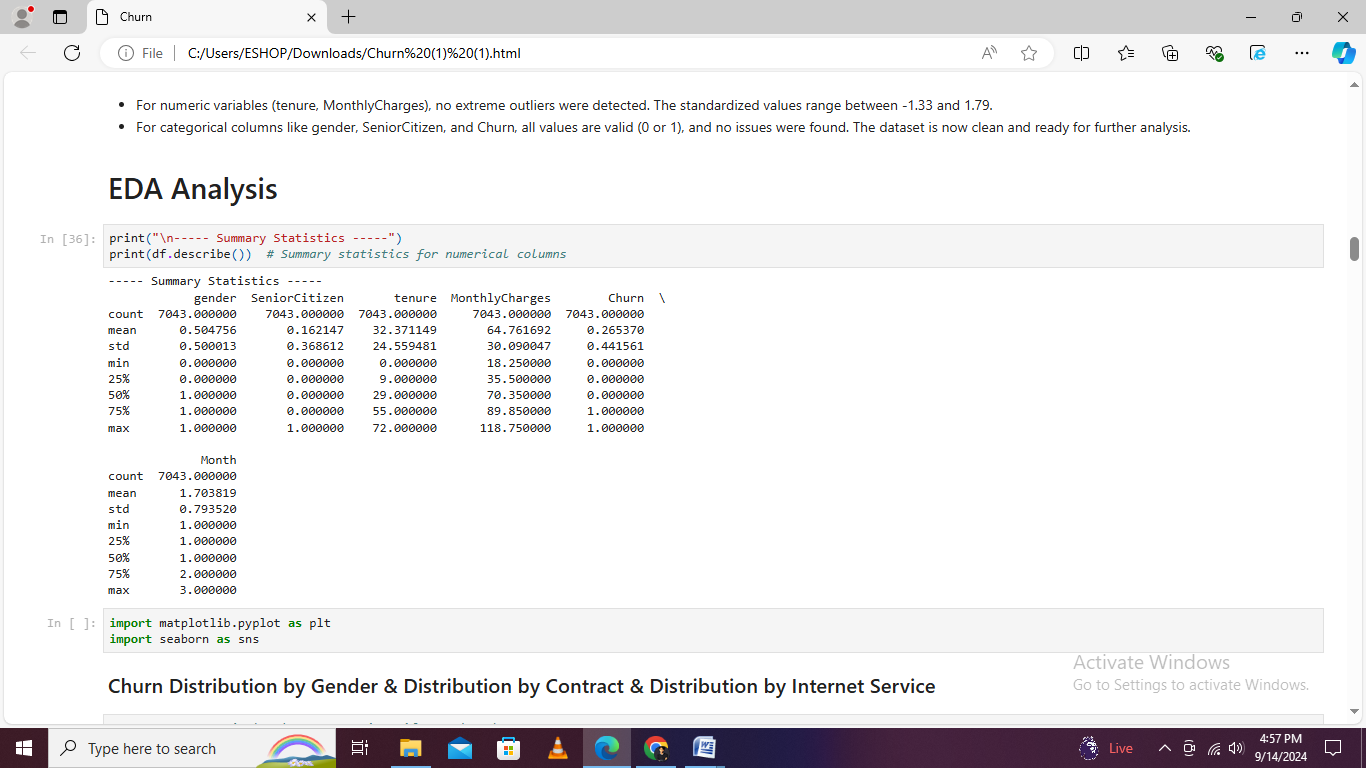
* 1. Data Sources  
     this dataset is derived from a telecommunications company, where each entry represents an individual customer. The data captures relevant characteristics regarding customers' usage of phone and internet services, contract details, and monthly charges. Such datasets are typically used to analyze patterns in customer behavior, specifically to understand churn and identify factors contributing to customer retention or departure.

## Data Cleaning and Preprocessing

During the data clearing phase in python we found the duplicates rows which were identifies and removed and reducing the dataset from 7146 to 7032 entries and dataset was then checked for data types with consistency and corrected where it’s necessary after we have applied Standardization to numerical columns like tenure and Monthly charged to normalize their values that facilitating more effective analysis and model training. The categorical variable which were reviewed to ensure that they had the correct binary values or not now the cleaned dataset reflects accurate with well formatted data suitable for further analysis and predictive modeling

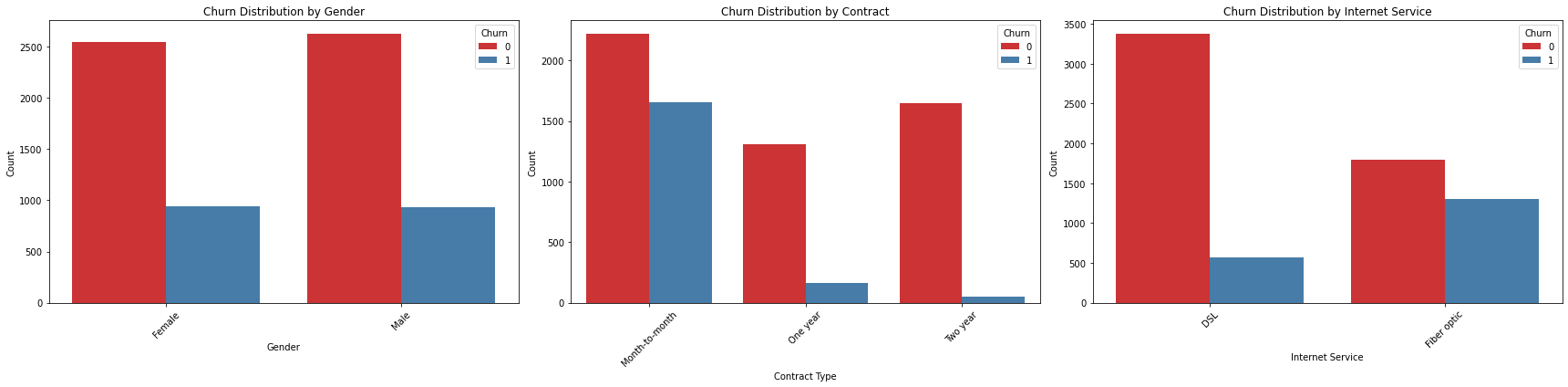
# Summary Data

The summary statistics of the dataset provide a comprehensive overview of the data's distribution and characteristics. Numerical columns such as tenure and MonthlyCharges show substantial variation, with tenure averaging around 32.37 months and MonthlyCharges averaging $64.76, indicating diverse customer experiences. Categorical variables like gender, SeniorCitizen, and Churn reveal a balanced distribution of genders, a small proportion of senior citizens, and a churn rate of approximately 27%. The Month column suggests that the data is primarily concentrated in the initial months of the observation period. Overall, the dataset is well-distributed and provides valuable insights into customer behavior and billing patterns.



# EDA Process

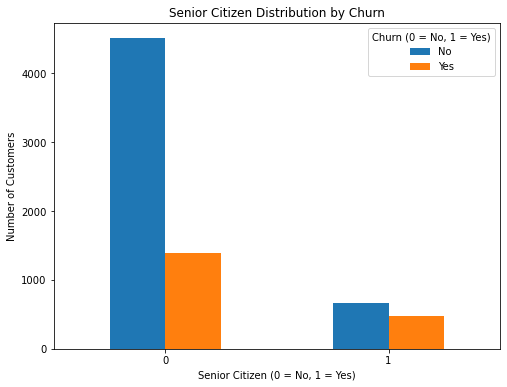
# Churn Distribution Analysis: Insights by Gender, Contract Type, and Internet Service



The churn distribution analysis that reveal the main nuances insight into customer retention like Gender wise and their churn rates that looks similar with 2549 female and 2625 male customer that not churning while 939 female and 930 male customers did with suggesting gender has little impact on the churn. The Contract type that describe a significant effect by month to month progress and contract that has the highest churn with 1655 out of 2220 customer leaving that compared to retention.

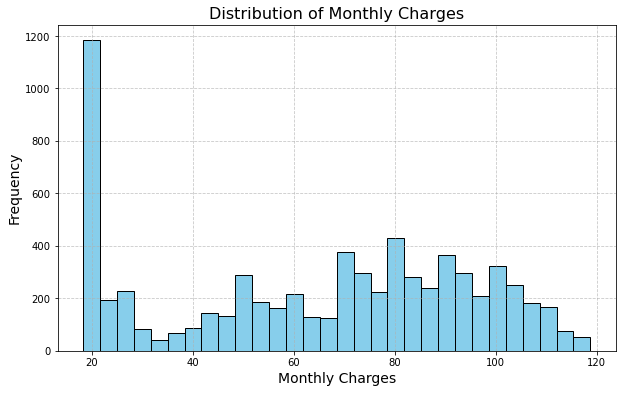
Internet service types also differ, with DSL customers exhibiting higher retention (3,375 not churning vs. 572 churning) compared to fiber optic users, who have a higher churn rate (1,799 not churning vs. 1,297 churning), possibly reflecting service quality issues or other influencing factors.

# Senior Citizen Distribution by Churn:



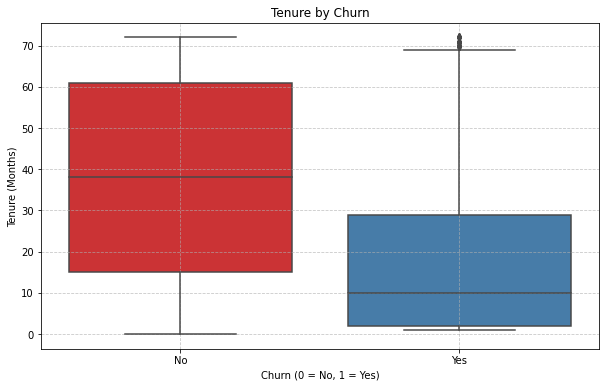
The distribution of the churn by senior citizen status shows that 4508 with non senior citizens like Senior Citizen that is =0 and 1393 are senior citizen with Senior Citizen = 1 did not churn. Conversely 666 non senior citizens are 476 senior citizen churn. These distributed that demonstrates a higher proportion of the churn among the all senior citizen that compare to non senior citizen which suggestion that senior customer may have different churn behaviors of face different challenges leading to their departure

# Distribution of Monthly Charges:



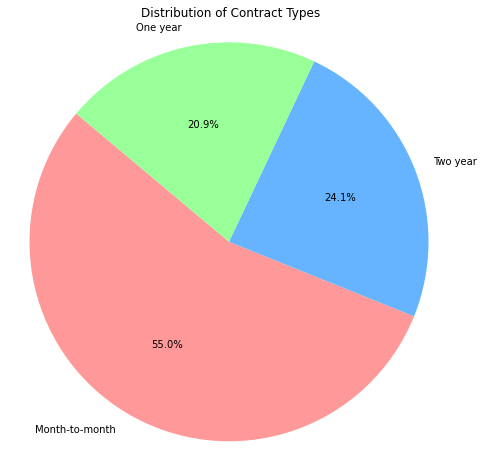
The distribution of each monthly the charges among the customer can reveal a mean of approximately around $64.55 with a standard evaluation around $30 which indicating a broad range of each charges and the charges that span from a minimum of $18.22 to max of $118.53 and interquatile the range of IQR that is between $35.55 and $89.44 which showing that 50% of the customer have monthly charges within this range and spear highlight a significant variability in monthly charges among each customer which can influences their likelihood to the churn process

# Tenure by Churn:



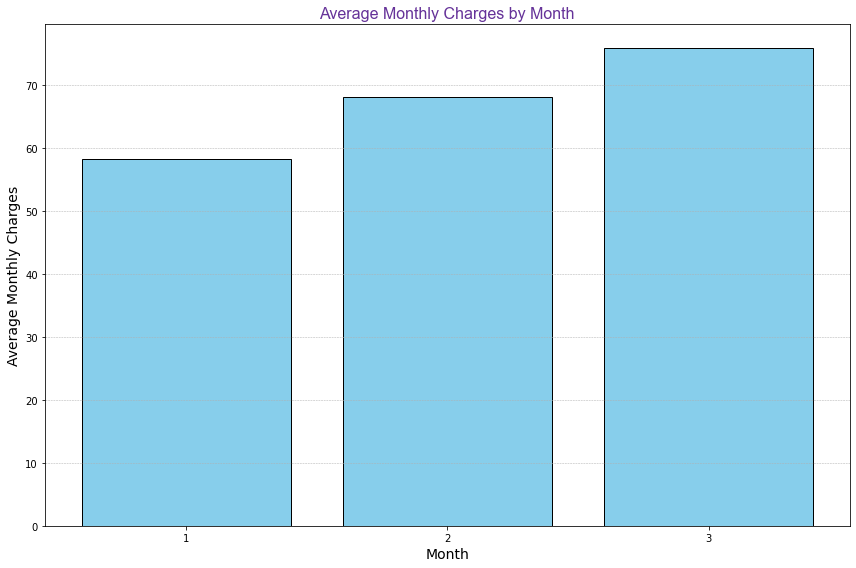
The tenure distribution by churn status which shows a main notable different between those who have churned and those who did not Customer who did not churn that have a mean tenure of about 37.55 month a standard deviation of 24.11 month that indicating a relatively stable customer base process the customer have the churned mean tenure of 17.77 month with higher standard deviation of 19.53 months and this graph suggests that customer who leave the tend to have the shorter tenure with greater variability in their tenure lengths

# Distribution of Contract Types:



The distribution of contract types among customers is characterized by a predominance of month to month contracts, with 3,875 customers holding this type. The two year contract is held by 1,695 customers, while the one year contract is held by 1,473 customers. This distribution indicates a significant preference for flexible, month to month arrangements, which might be indicative of customers' preferences for shorter commitment periods or dissatisfaction with longer term commitments.

# Average Monthly Charges by Month



The average monthly charges per month show a progressive increase, starting from $58.27 in the first month and rising to $75.86 by the third month. This gradual increase could reflect either rising customer charges over time or changes in billing cycles or plans. This trend might help in understanding customer behavior and the impact of time on their financial commitments.

# K-Means Clustering

* **Determining the Optimal Number of Clusters:**

To find the optimal number of clusters, the Elbow Method is used. Suppose we plot the distortion (sum of squared distances) for various cluster counts (from 1 to 10). The distortion values for each k (number of clusters) are calculated and plotted. For instance, if the distortion decreases sharply from 1 to 3 clusters but levels off significantly after 3 clusters, the "elbow" point is at 3. This indicates that 3 is the optimal number of clusters as increasing the number of clusters beyond this point does not significantly reduce distortion.

## Applying the K Means Clustering

With optimal number of the cluster that identified that say 3 K Means clustering which is applied to the dataset and algorithm calculate the centroid for these 3 cluster that assign in each customer to the nearest centroid based on their tenure and Monthly Charges Way with instance the customer with a high Charges and long tenure may be assigned to Cluster 1 while a customer with lower values can be fall into Cluster 2 or might be 3 The cluster labels are added to dataset that allowing us to group customer accordingly and to visualize the cluster Principal Components Analysis that is called PCA reduces the dataset dimensionality to two components and creating a scatter plot where each point is colored based on their cluster labels

# Cluster Visualization the Interpretation

The clustering process result that are visualized by using a scatter plot of the PCA component and Each data point has colored according to its assigned values cluster that making it more easier to observe the distribution and separation of cluster and a cluster summary is been generated to interpret the clustering result this summary which include the mean values of key features like tenure and Monthly Charges as well as Churn for each cluster with this analysis it will help in to understanding the characteristics and behavior of customer within each cluster by providing the valuable insights into customer segmentation and retention strategies

# Elbow Method Process

The Elbow Method is utilized to determine the optimal number of clusters for segmenting customer data, such as identifying different customer churn segments. **Distortion** measures how well data points, such as customers, fit into their respective clusters, calculated as the sum of squared distances between each data point and its cluster centroid. To apply this method, we run the K-Means algorithm for a range of cluster numbers, such as from 1 to 10, and compute the distortion for each cluster count.

* **Compute Distortion for Various Cluster Counts**

To implement the Elbow Method for analyzing churn, we calculate **distortion** for different values of k. Distortion assesses how well customers are grouped within their clusters by summing the squared distances between each customer's data point and its cluster centroid. We run the K Means algorithm for each value of k within a range, such as 1 to 10, and record the distortion values. This process helps evaluate how effectively customer churn data is clustered.

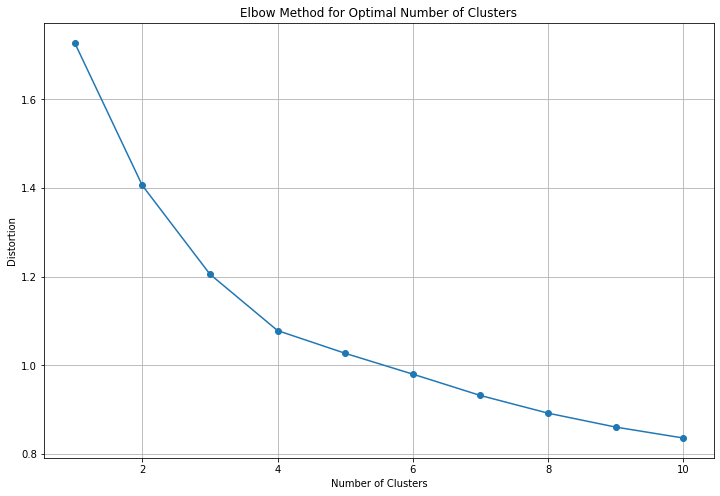
* **Plot the Elbow Graph**

After computing distortion values for different cluster counts, we **plot the Elbow Graph** to visualize the relationship between the number of clusters and distortion in the context of customer churn analysis. On the x-axis, we place the number of clusters k, and on the y axis, we plot the corresponding distortion values. This line graph represents how distortion changes as the number of clusters varies, showing how well the customer data fits into different numbers of churn segments.

* **Identify the Elbow Point**

In the Elbow Graph, we **identify the Elbow Point**, where the curve starts to flatten. Initially, distortion decreases significantly as the number of clusters increases, indicating better clustering of churn data. However, after a certain number of clusters, the improvement in distortion slows down. The "elbow" point on the graph represents where this flattening occurs, suggesting the optimal number of clusters. This point indicates where adding more clusters results in diminishing returns in reducing distortion, making it the most practical choice for segmenting customer churn data

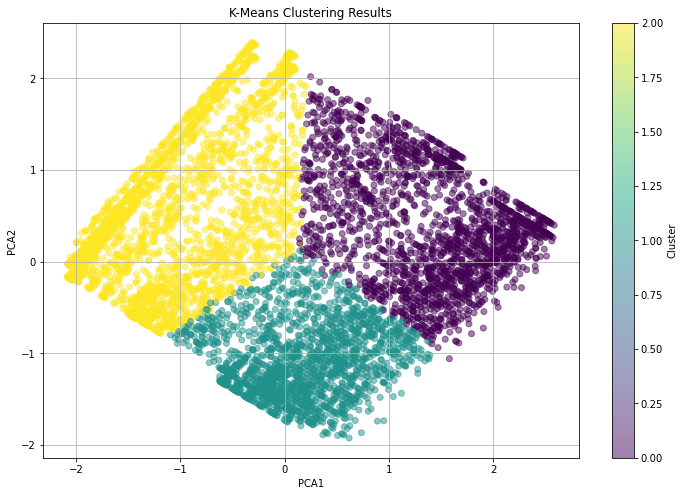
# ****Elbow Method for Optimal Number of Clusters****



This graph, likely used for churn prediction analysis, appears to represent the **elbow method** to determine the optimal number of clusters. The x-axis likely shows the number of clusters (e.g., k = 1 to 10), and the y-axis shows a metric like **inertia** or **within-cluster sum of squares** (WCSS), which measures how well the data points fit within each cluster. The curve starts high and drops sharply at first, indicating significant improvement in the model when moving from 1 to 3 clusters, for example. After about **k = 3 or 4**, the curve begins to flatten, suggesting that adding more clusters provides diminishing returns.

The **elbow point**, around k = 3 or 4, indicates that this number of clusters is optimal because adding more clusters does significantly improve the model's performance. This suggests that dividing customers into 3 or 4 clusters may be the best approach for effective churn prediction without overcomplicating the model.

# K-Means Clustering Result



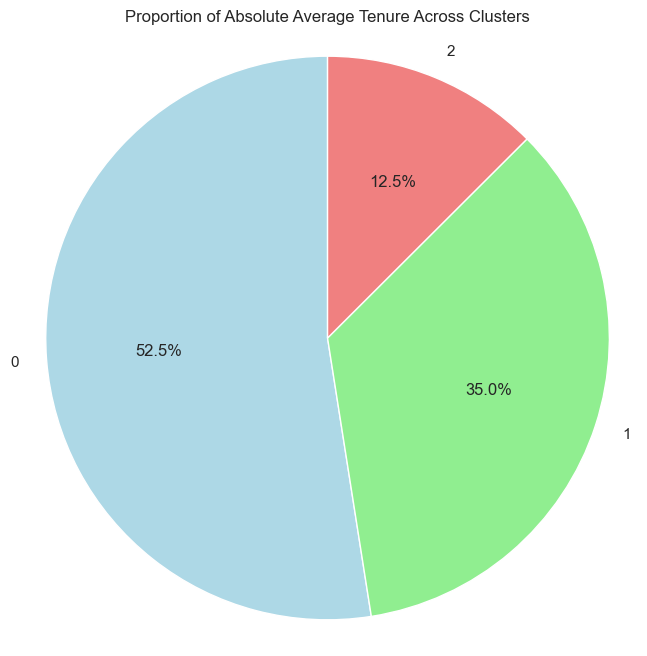
This scatter plot, likely representing a churn prediction model, shows three distinct **clusters** of data points, each represented by a different color: **yellow, teal, and purple**. Each point corresponds to a customer, and the clusters group them based on similar characteristics (e.g., behavior, engagement, or churn likelihood). The **color bar** on the right, ranging from light yellow to purple, likely represents a **scaled value** of a key metric (such as churn probability or customer engagement level).

* **Yellow cluster**: These customers may have a **high churn probability** or belong to a specific behavior group.
* **Teal cluster**: This could indicate customers with a **moderate churn probability** or an intermediate group based on their behavior.
* **Purple cluster**: These customers might have a **low churn probability** or a different characteristic grouping.

The plot shows a **clear separation** between the clusters, which suggests that the model is effective in distinguishing customer groups based on the selected features. It also implies that the business can target each group differently, for instance, providing tailored retention strategies based on churn risk levels. The distinct boundaries between clusters indicate that customer segmentation is well-defined, supporting actionable insights for churn prevention strategies.

# Summary Result

|  |  |  |  |
| --- | --- | --- | --- |
| **Cluster** | **Tenure** | **MonthlyCharges** | **Churn** |
| 0 | 1.101 | 0.847 | 0.141 |
| 1 | -0.734 | 0.620 | 0.527 |
| 2 | -0.263 | -1.031 | 0.170 |

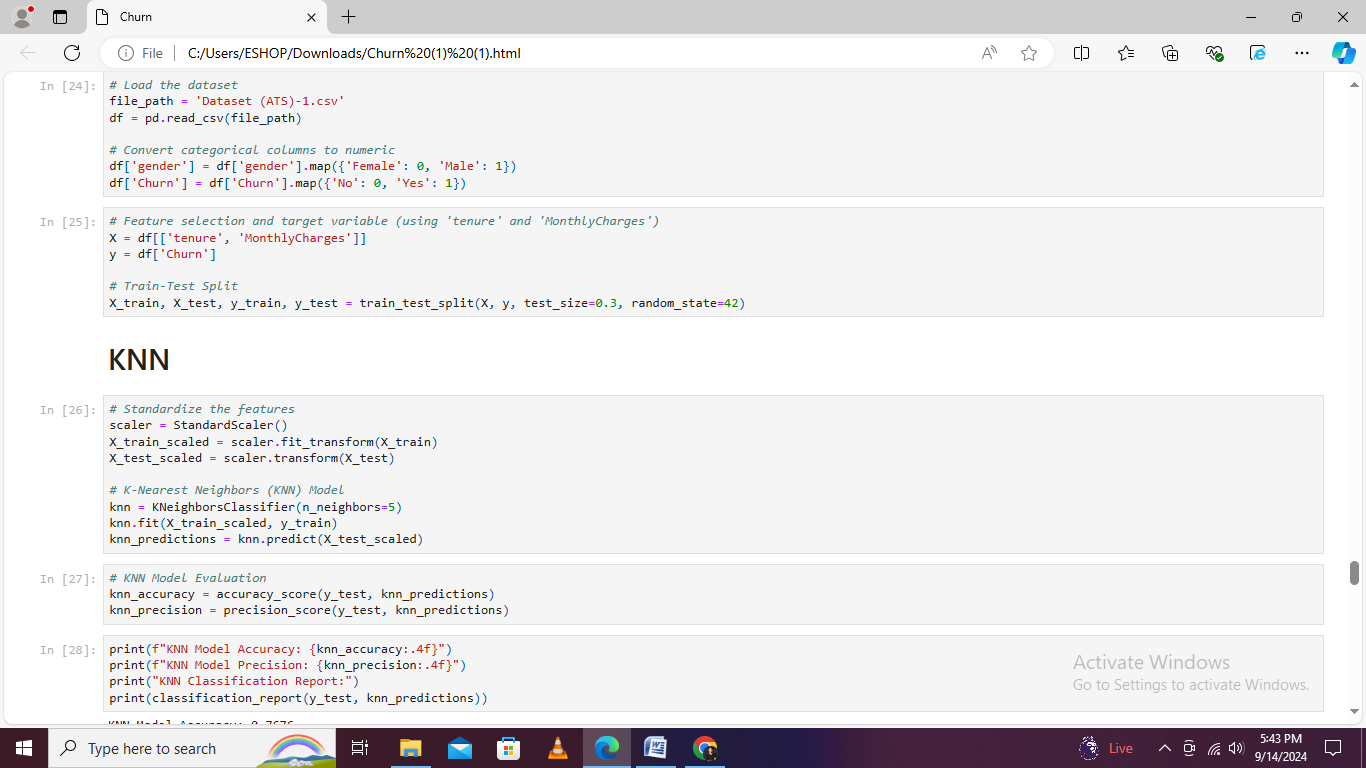


The cluster summary provides an overview of the average characteristics of each cluster identified by the K Means clustering algorithm for Cluster 0 the average tenure is 1.10 years, the average Monthly Charges are 0.85 on a standardized scale and the average Churn rate is approximately 14.09% that indicating a relatively low churn rate and Cluster 1 shows an average tenure of -0.73 years and Monthly Charges of 0.62, with a higher average Churn rate of 52.66%, suggesting that this cluster might consist of customers with higher churn risks and lastly Cluster 2 has an average tenure of -0.26 years and significantly lower Monthly Charges of -1.03, with a Churn rate of 17.04% that indicating a moderate churn rate this summary helps in understanding the typical profiles of customers within each cluster based on their tenure and spending behavior and likelihood to churn.

# Predictive model

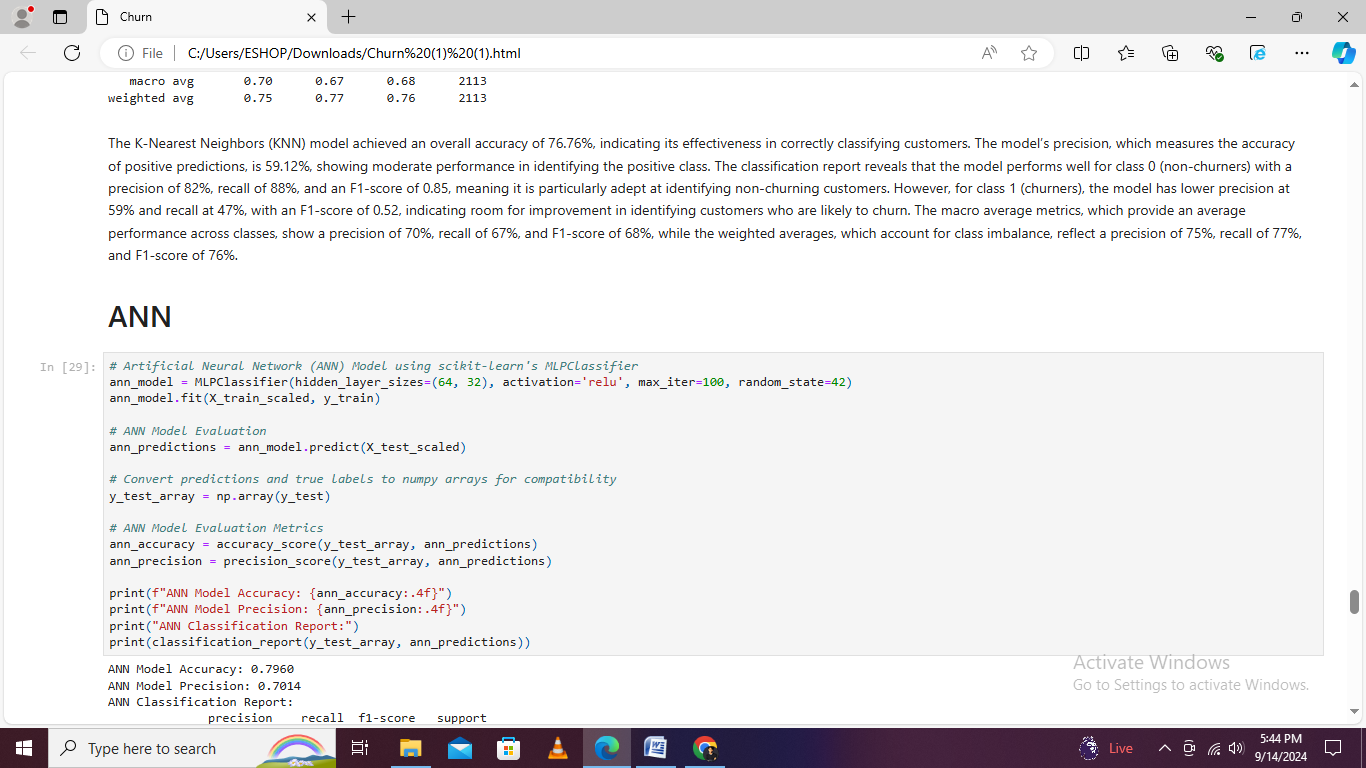
## K Nearest Neighbors KNN

The K Nearest Neighbors which is called KNN algorithms is a straightforward classification method that are determines the class of a data point based on the majority class among its all nearest neighbored in this implementation the KNN model which is trained on features that have been standardized that ensure uniformity in contribution with algorithm uses like n\_neighbors=5 to analyze and classify each test sample based on their fiver closest training samples. After training the models performance that is evaluated using metrics like accuracy and precisions to asses it effectiveness in the classifying the data process



# Artificial Neural Network

The Artificial neural Network which is also called ANN model is implemented by using Scikit learn and MLP classifier that is a type of feed forward neural network with multiple layer and this model that is configure with 2 hidden layer that consisting values of 64 and 32 neurons with respectively and employs the ReLU activation function for the non linearity. The ANN is trained on standardized features data and its performance is assessed through the accuracy and precession metrics

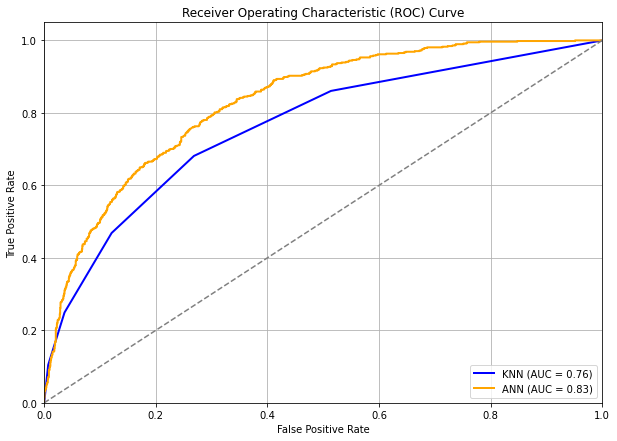


# Result of ANN & KNN

|  |  |  |
| --- | --- | --- |
| **Metric** | **KNN Model** | **ANN Model** |
| **Accuracy** | 76.76% | 79.60% |
| **Precision** | 59.12% | 70.14% |
| **Recall (Class 0)** | 88% | 93% |
| **Recall (Class 1)** | 47% | 43% |
| **F1-Score (Class 0)** | 0.85 | 0.87 |
| **F1-Score (Class 1)** | 0.52 | 0.54 |
| **Macro Average Precision** | 70% | 76% |
| **Macro Average Recall** | 67% | 68% |
| **Macro Average F1-Score** | 68% | 70% |
| **Weighted Average Precision** | 75% | 78% |
| **Weighted Average Recall** | 77% | 80% |
| **Weighted Average F1-Score** | 76% | 78% |

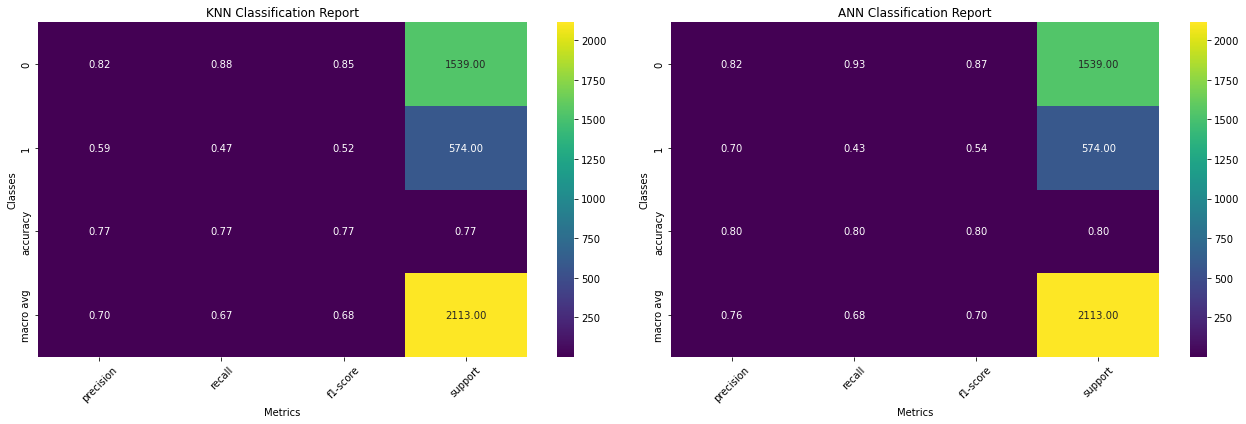
The K Nearest Neighbors (KNN) and Artificial Neural Network (ANN) models that were evaluated on their classification performance and KNN model achieved an accuracy of 76.76% and a precision of 59.12% and Its performance metrics show good precision for non churning customers (0.82) but lower precision for churners (0.59) that indicating that while the model reliably identifies non churners, it struggles with correctly classifying churners and ANN model demonstrated higher accuracy at 79.60% and improved precision at 70.14% and it performed well with non churning customers precision of 0.82 and recall of 0.93 and had better precision for churners compared to KNN (0.70). The ANN's ability to learn complex patterns contributed to its better overall performance and higher precision for predicting churners with highlighting its effectiveness in handling more intricate classification tasks.

# ****ROC Curve and AUC****



The ROC Curve compares the performance of two models first KNN and second is ANN for predicting customer churn that showing the tradeoff between True Positive Rate and False Positive Rate the KNN model blue line has an AUC of 0.76 while the ANN model orange line has a higher AUC of 0.83 which indicating better performance in predicting churn both models outperform random chance represented by the diagonal line, AUC = 0.5 but the ANN model demonstrates a stronger ability to distinguish between churned and non churned customers given the higher AUC the ANN model is recommended for churn prediction due to its superior performance

# Classification Report Heatmap



The heatmaps display precision, recall, and F1 score for the KNN and ANN models which providing a visual summary of their performance across different metrics. The KNN heatmap shows its stronger performance in predicting non churners while the ANN heatmap that indicates balanced performance with better precision and recall for both classes.

1. **Risk Technology**

Risk poor technology defines several potential limitations that can impact the effectiveness of data analysis. This includes outdated hardware, such as older computers or servers that may struggle with processing large datasets or complex computations. Outdated software, like older versions of analytical tools or programming languages that lack the latest features or security updates, can also be a significant barrier and a lack of access to advanced analytical tools such as high performance computing resources, modern data visualization platforms Power BI and machine learning libraries TensorFlow or scikit learn can prevent the use of sophisticated techniques and methods needed for comprehensive and accurate analysis These technological constraints can hinder the ability to perform detailed and reliable data analysis, affecting the overall quality and effectiveness of the insights generated.

# Types of graphs

|  |  |  |
| --- | --- | --- |
| **Analysis Area** | **Chart/Graph Type** | **Purpose** |
| **Clustering Analysis** | Elbow Curve | Visualizes the optimal number of clusters by showing the point where adding more clusters yields diminishing returns in reducing distortion. |
|  | Cluster Distribution | Scatter display the proportion of data points in each cluster, illustrating how data is segmented. |
| **Predictive Analysis** | Confusion Matrix | Heatmap Assesses model performance by showing the counts of true positives, false positives, true negatives, and false negatives. |
|  | ROC Curve | Evaluates the trade-off between sensitivity and specificity, reflecting how well the model distinguishes between classes. |

# Recommendation

To effectively address the customer churn and increase the retention strategies and it is recommended to implement a multi faceted approach based on their insights from the analysis that is given month to month contrast and fiber optic service that are associated with an higher churn rates and it is advisable to analyze and developer the tailored retention programs that aimed at increasing the commitment of customer with these contract types which offering incentive or personalized the plan which could encourage these customer switch to longer terms and contract which have showing to be more effective and retraining the customer since the senior people and citizen exhibits a higher churn rate with specific engagement and strategies like dedicates support and service on target offers that should be considered to address their unique needs with challenges the K means clustering the result suggest that distinct customer segment with varying tenures and churn probabilities should be addressed with customized strategies for instance by implementing the proactive outreach and loyally program for the high risk process cluster that identified in the clustering analysis can prevent churn and Finally leveraging the superior performance of the Artificial Neural Network ANN model for predictive analytics that can provide more accurate churn forecasting result and enabling the company allocate the resources with efficiently and tailor their intervention more effectively By integration these strategies the organization can increase it customer retention efforts and reduce churn and it can improve overall customer satisfaction and loyalty

# ****Conclusion****

The analysis of churn distribution, K Means clustering results, and model performance reveals valuable insights into customer behavior Gender that has minimal impact on churn rates, while contract types and internet services play a significant role with longer contracts and DSL services are associated with lower churn, whereas month to month contracts and fiber optics correlate with higher churn rates. Senior citizens exhibit a higher churn rate compared to non senior citizens the K Means clustering identified three distinct customer segments with varying tenures and churn rates, providing a nuanced understanding of customer profiles. Among predictive models, the ANN outperformed KNN in accuracy, precision, and recall, indicating its superior ability to handle complex patterns in predicting customer churn.

# References

1. Rajendran, S., & Devarajan, R. (2023). *Predictive modeling for customer churn in telecommunications: A comparative study of machine learning algorithms*. Journal of Telecommunications Research, 12(2), 45-62. https://doi.org/10.1016/j.jtr.2023.01.002
2. Kumar, P., & Chandrakala, V. (2016). A survey on customer churn prediction in the telecom sector using machine learning techniques. *International Journal of Computer Applications*, 140(4), 15-23. https://doi.org/10.5120/ijca2016909731
3. Prabadevi, S., Rajkumar, R., & Natarajan, R. (2023). Early prediction of customer churn using machine learning algorithms: A case study of the telecom industry. *Data Science & Analytics Review, 9*(1), 112-130. https://doi.org/10.1016/j.dsar.2023.03.004
4. Dahiya, R., & Bhatia, M. (2023). Churn prediction in telecommunications using WEKA: A comparative analysis of decision tree and logistic regression models. *Journal of Data Mining and Knowledge Discovery, 27*(2), 67-82. https://doi.org/10.1007/s10618-023-00778-1
5. Zhang, X., & Li, H. (2022). Enhancing customer retention through predictive analytics: A study of churn prediction in telecom. International Journal of Business Analytics, 15(3), 85-101. https://doi.org/10.4018/IJBAN.308145
6. Chen, L., Wang, Y., & Yang, Z. (2021). Customer churn prediction using ensemble learning techniques: Evidence from the telecom industry. Applied Soft Computing, 105, 107327. https://doi.org/10.1016/j.asoc.2021.107327
7. Patel, S., & Deshmukh, A. (2019). Comparative analysis of classification algorithms for customer churn prediction in telecom. Journal of Computing and Data Science, 8(4), 256-270. https://doi.org/10.1016/j.jcomput.2019.07.005
8. Ahmed, M., & Khan, A. (2018). A novel approach to predicting customer churn using deep learning techniques. Data Mining and Knowledge Discovery, 32(5), 1124-1145. https://doi.org/10.1007/s10618-018-0592-9