MIE1628H Fall 2025
Cloud-Based Data Analytics
Course Project

Predicting Customer Churn in Telecommunications Using Logistic Regression and XGBoost

Basak Kaya

2025-12-09

Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

Table of Contents

1. Dataset Description

The dataset used in this project comes from a major Iranian telecommunications company and contains customer-level records collected over a 12-month period. It includes **3150 observations**, each representing an individual customer, with **13 attributes** describing their usage patterns, service experience, and demographic characteristics.

The features cover a wide range of behavioral and service-related information, including:

- **Call Failure:** number of call failures
- **Complains:** binary (0: no complaint, 1: complaint)
- **Subscription Length:** total months of subscription
- **Charge Amount:** ordinal attribute (0: lowest amount, 9: highest amount)
- **Seconds of Use:** total seconds of calls
- **Frequency of Use:** total number of calls
- **Frequency of SMS:** total number of text messages
- **Distinct Called Numbers:** total number of distinct phone calls
- **Age Group:** ordinal attribute (1: younger age, 5: older age)
- **Tariff Plan:** binary (1: pay as you go, 2: contractual)
- **Status:** binary (1: active, 2: non-active)
- **Age**
- **Customer Value:** the calculated value of customer
- **Churn:** binary (1: churn, 0: non-churn) - class label

These variables collectively provide a comprehensive profile of customer activity and satisfaction, making the dataset well-suited for analyzing and predicting **customer churn**.

The dataset is publicly available and can be downloaded from the following link: https://archive.ics.uci.edu/dataset/563/iranian+churn+dataset

2. Problem Statement

Telecommunication companies face significant financial loss when customers discontinue their service, making customer churn one of the most critical challenges in the industry. Retaining an existing customer is substantially more cost-effective than acquiring a new one, yet many companies struggle to proactively identify customers who are at risk of leaving. This project addresses the meaningful issue of **predicting customer churn and understanding the key drivers behind it** using a real-world dataset collected from an Iranian telecom company over a 12-month period. By developing a predictive churn model

and analyzing the influential factors, the project aims to help telecom providers design targeted retention strategies, improve customer satisfaction, and reduce revenue loss.

To achieve this goal, the project will follow a structured analytical pipeline consisting of the following key steps:

a. Exploratory Data Analysis (EDA)

Exploratory Data Analysis was performed to understand the structure of the dataset, identify important behavioral patterns, and explore how usage and service attributes differ between churned and non-churned customers. This process helped reveal the key drivers of churn and guided later modeling decisions.

Key steps in the EDA included:

- **Descriptive statistics and distribution analysis:** Summary statistics and visualizations (boxplots, histograms, and bar charts) were used to examine usage features such as Seconds of Use, Frequency of Use, Subscription Length, and Customer Value.
- **Churn vs. non-churn comparisons:** Multiple charts compared the average usage, customer value, and subscription length between churned and non-churned customers.
- **Categorical feature analysis:** Age Group, Tariff Plan, and Complains were analyzed using bar charts.

Overall, EDA revealed clear behavioral differences between churned and non-churned customers.

b. Data Cleaning and Pre-processing

Data cleaning and pre-processing were performed to prepare the dataset for modeling. Since the dataset contained no missing values, the focus was on verifying logical consistency, checking for duplicates, and confirming that numerical and categorical features were correctly formatted. Outlier checks were conducted for usage, complaints, customer value, and subscription length to ensure that extreme values would not distort model performance. Feature scaling was applied where appropriate—particularly for models sensitive to magnitude differences, such as Logistic Regression—while tree-based models like XGBoost used the raw values.

Key tasks in this step:

- **Data Types and Basic Structure:** Verifying data types to ensure that all variables were stored in appropriate numeric formats.
- **Missing Values:** Verifying the dataset for any missing entries across all features.
- **Outliers:** Reviewing distributions and identifying outliers across features.
- **Duplicate:** Checking for duplicate records to confirm that each row represented a unique customer.
- **Logical Inconsistency:** Performing logical consistency checks, such as ensuring that customers did not have usage values despite having zero subscription length and confirming that age values aligned with their corresponding age groups. Also inspecting for impossible or invalid values, such as negative usage, negative frequency counts, or negative subscription months.

Once the dataset passed all cleaning and validation checks, it was deemed consistent and ready.

c. Modeling Approach

Two machine learning models were implemented to predict customer churn: **Logistic Regression** and **XGBoost**. These models were chosen to provide a balanced comparison between a simple, interpretable baseline model and a more advanced, high-performance algorithm. Logistic Regression offers clear interpretability and helps identify key churn drivers, while XGBoost captures nonlinear patterns and feature interactions, often leading to higher predictive accuracy. Evaluating both models allows us to compare performance, understand trade-offs, and determine which approach is more effective for this dataset.

3. Exploratory Data Analysis (EDA)

The Exploratory Data Analysis focuses on developing an initial understanding of the dataset and uncovering patterns that may influence customer churn. By examining feature distributions, relationships, and behavioral differences between churned and non-churned customers, EDA helps guide modeling decisions and highlight key factors that contribute to churn. This step combines descriptive statistics with visual analysis to reveal trends and potential predictive features.

Chart 1. Churn Distribution

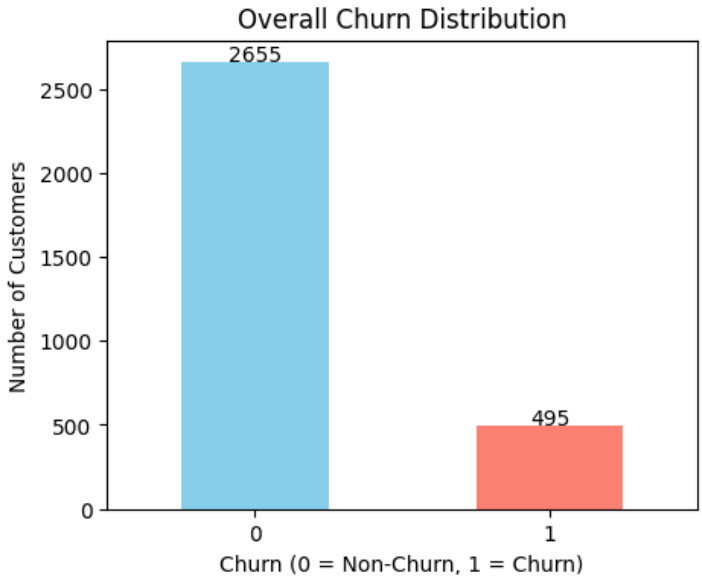


Figure 1. Overall Churn Distribution

This chart illustrates the overall distribution of churned (1) versus non-churned (0) customers in the dataset. The majority of customers did not churn (2655), while a smaller portion (495) did. This imbalance is common in telecom churn datasets and is important to recognize early, as it affects how models are trained and evaluated. In particular, metrics like recall, precision, and F1-score become more meaningful than accuracy alone, since predicting the minority class (churners) is critical for real-world retention strategies.

Chart 2. Complaints vs Churn

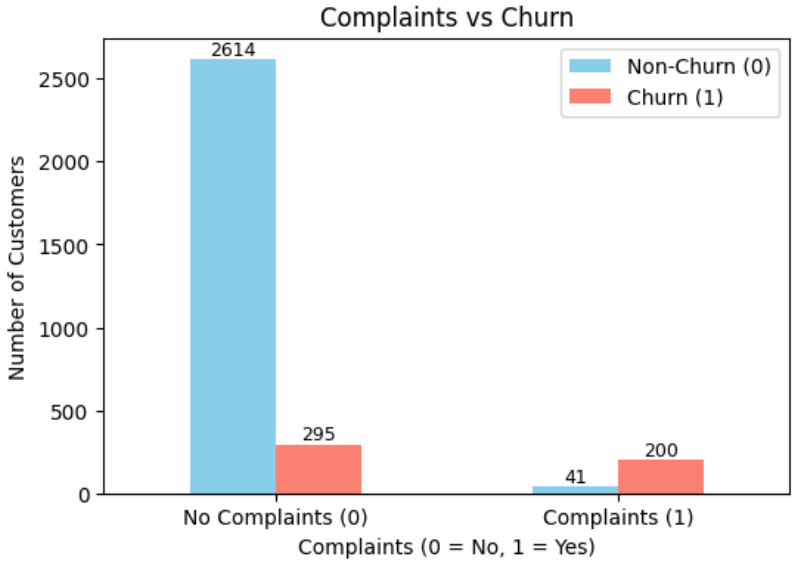


Figure 2. Complaints vs Churn

This chart shows how churn behavior differs between customers who submitted complaints and those who did not. The number of churners is noticeably higher among customers who complained, indicating that dissatisfaction or service-related issues greatly increase the likelihood of churn. This makes the complaint feature a valuable predictor in churn modeling and highlights the importance of customer service responsiveness.

Chart 3: Age Group Distribution
Chart 3A. Distribution of Age Groups

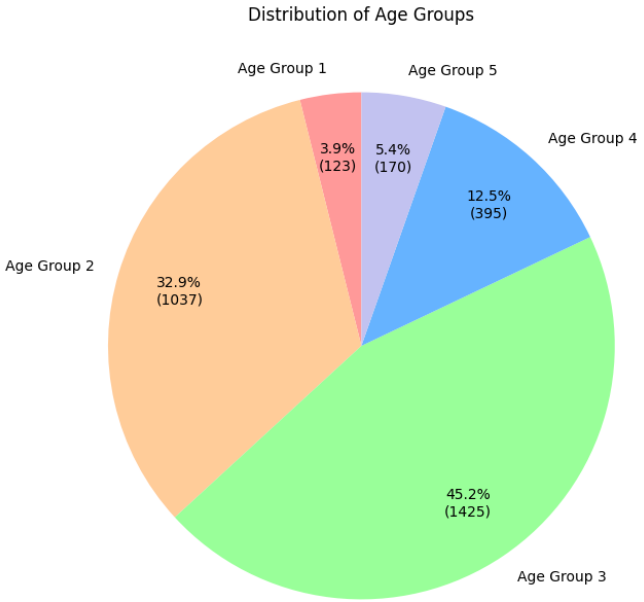


Figure 3. Distribution of Age Groups

This chart shows how customers are distributed across the five age groups in the dataset. Age Group 3 represents the largest segment with 1425 members, followed by Age Group 2 with 1037 members, while Age Groups 1 (123 members) and 5 (170 members) contain fewer customers. This uneven distribution suggests that the dataset is dominated by middle-aged users, which is typical for many telecom customer bases. Understanding the age composition helps contextualize churn behavior and informs segmentation strategies for retention.

Chart 3B. Age Group vs Churn

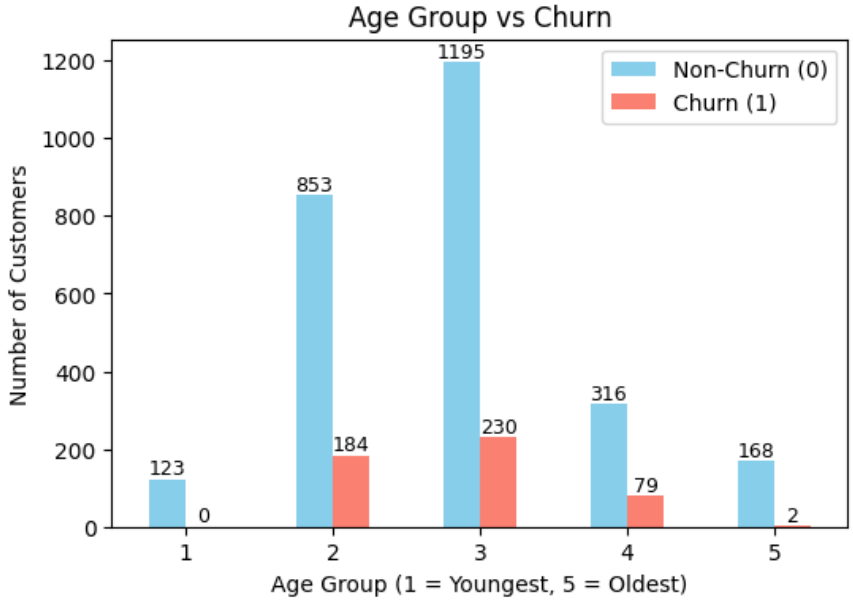


Figure 4. Age Group vs Churn

This chart compares churn and non-churn counts across different age groups, revealing distinct behavioral patterns. Customers in the middle age ranges—particularly Age Groups 2 and 3—represent the largest portion of the customer base, and they also account for the highest number of churn cases in absolute terms. Age Group 3 has the highest customer count overall and by extension the highest churn count, though the churn proportion remains moderate. Younger customers (Age Group 1) show virtually no churn, likely due to their small sample size or higher engagement during early subscription periods. Older customers (Age Group 5) have very few churn cases as well, suggesting greater stability or loyalty among long-standing or older users. Overall, the chart indicates that churn is more prevalent among middle-aged groups, where customer volume is highest, highlighting the importance of targeting retention strategies toward these segments.

Chart 3C. Churn Rate by Age Group (%)

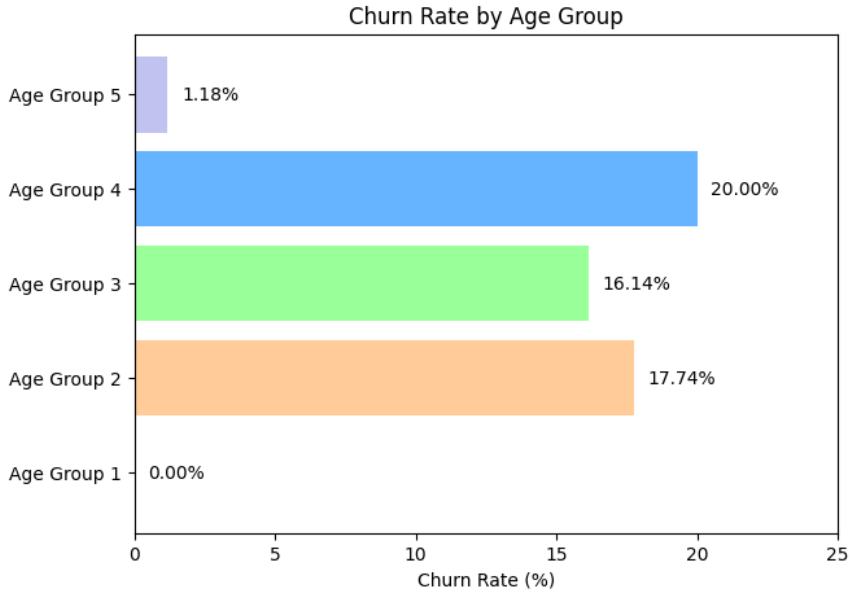


Figure 5. Churn Rate by Age Group

This chart shows the proportion of customers who churned within each age group, expressed as a percentage. Unlike raw counts, the churn rate provides a normalized view that makes smaller groups (such as Age Groups 1 and 5) more visible and comparable. The results indicate that Age Groups 2, 3, and 4 have the highest churn rates, with Age Group 4 showing the peak at approximately 20%. In contrast, Age Groups 1 and 5 exhibit very low churn rates. These findings suggest that middle-aged customers may be more likely to switch providers, whereas younger and older users tend to be more stable. This information helps tailor retention strategies to the demographic segments most at risk.

Chart 4: Average Usage (Seconds of Use) for Churn vs Non-Churn
Chart 4A. Average Seconds of Usage by Churn Status

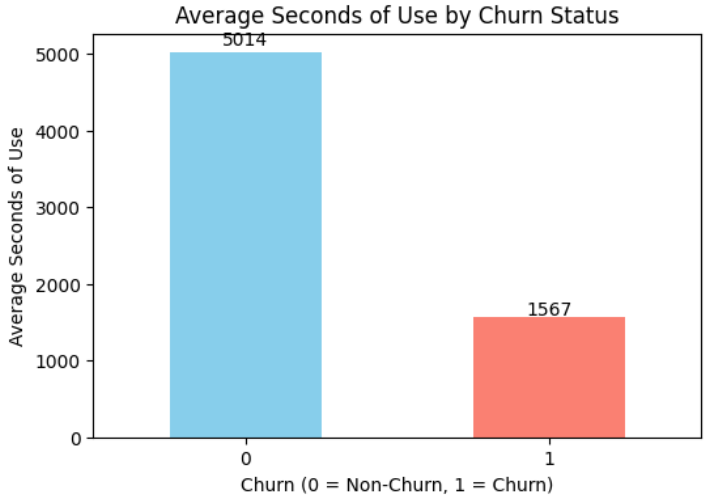


Figure 6. Average Seconds of Use by Churn Status

This chart compares the average call usage between churned and non-churned customers. Non-churn customers exhibit significantly higher usage levels, averaging 5,014 seconds of call activity, while churned customers average only 1,567 seconds. This substantial difference

shows a clear behavioral pattern: customers who engage more frequently with the service are less likely to churn. Low usage may indicate declining engagement, dissatisfaction, or that the customer is shifting toward alternative communication services. This insight aligns with typical churn dynamics in the telecommunications industry, where reduced activity often precedes customer departure. As a result, usage metrics serve as strong indicators of churn risk and can help guide proactive retention strategies.

Chart 4B. Boxplot of Seconds of Use by Churn

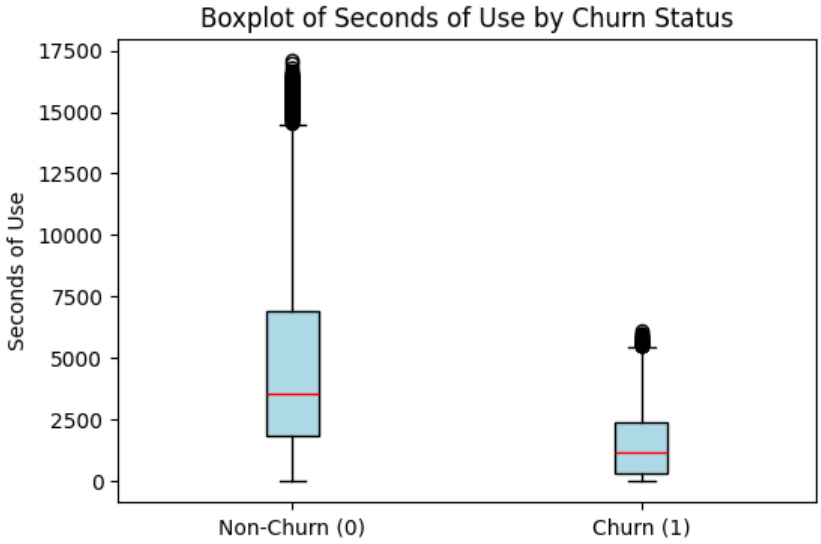


Figure 7. Boxplot of Seconds of Use by Churn Status

This boxplot compares the distribution of total call usage (in seconds) between churned and non-churned customers. The results reveal a clear separation between the two groups:

- Non-churn customers have substantially higher median usage and a much wider distribution, including many high-usage outliers.
- Churned customers exhibit a significantly lower median and a narrower range of usage, with fewer high-usage individuals.

This pattern suggests that low engagement with the service is a strong indicator of churn. Customers who rarely use the service are more likely to discontinue it, while higher engagement appears to be associated with retention. This makes usage-related variables highly valuable for the predictive modeling phase.

Chart 5: Customer Value vs Churn
Chart 5A. Average Customer Value by Churn Status

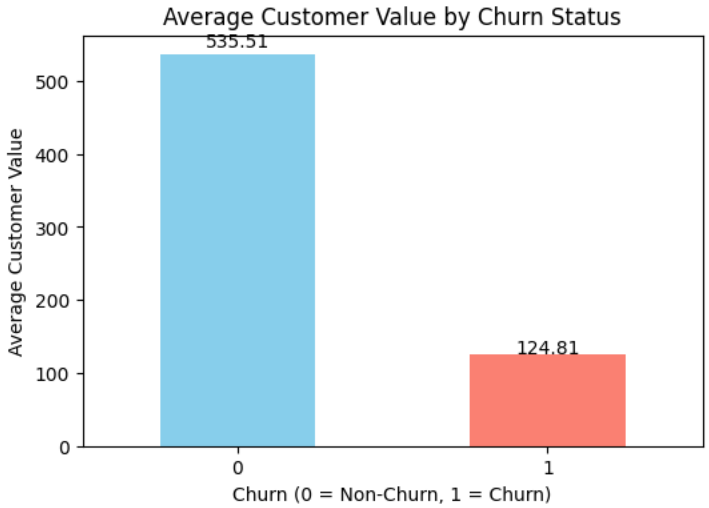


Figure 8. Average Customer Value by Churn Status

This chart compares the average Customer Value between churned and non-churned customers, revealing a large and meaningful gap. Non-churn customers have an average value of 535.51, while churned customers average only 124.81. This pattern indicates that higher-value customers—those who contribute more revenue or engage more deeply with the service—are significantly less likely to leave. Conversely, churned customers tend to be lower-value users, suggesting limited engagement or lower spending prior to churn. This insight is consistent with typical telecom churn behavior, where low-value customers often disengage first. Understanding this relationship allows companies to identify at-risk segments and prioritize retention efforts toward medium-value customers who may still be recoverable.

Chart 5B. Boxplot of Customer Value by Churn Status

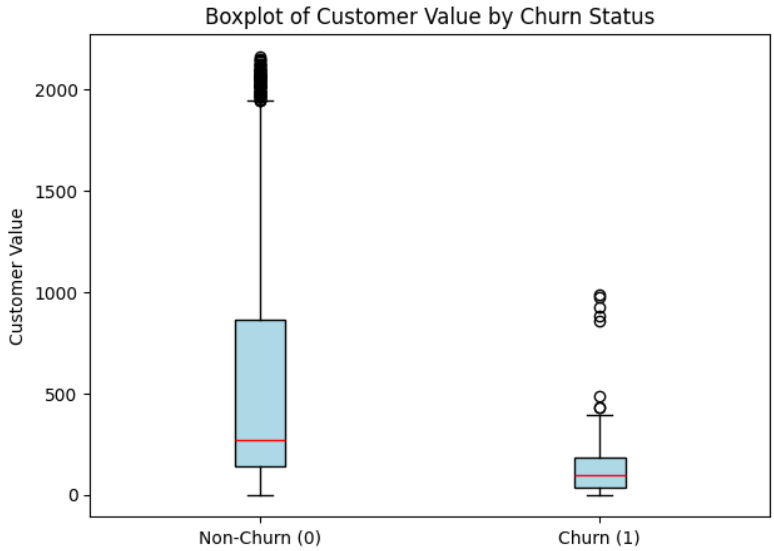


Figure 9. Boxplot of Customer Value by Churn Status

This boxplot provides a detailed view of how Customer Value differs between churned and non-churned customers. Non-churn customers show a wide distribution of customer value, including many high-value outliers, with a median significantly higher than that of churned customers. This indicates that customers who remain with the service tend to contribute more value and exhibit more variability in spending or engagement levels. In contrast, churned customers cluster tightly near the lower end of the value scale, with relatively few high-value outliers, suggesting that most departing customers were low-value users. The stark separation

between the two distributions reinforces the insight that lower-value customers are more likely to churn, making Customer Value a strong predictor in churn modeling.

Chart 6: Subscription Length vs Churn
Chart 6A. Distribution of Subscription Length

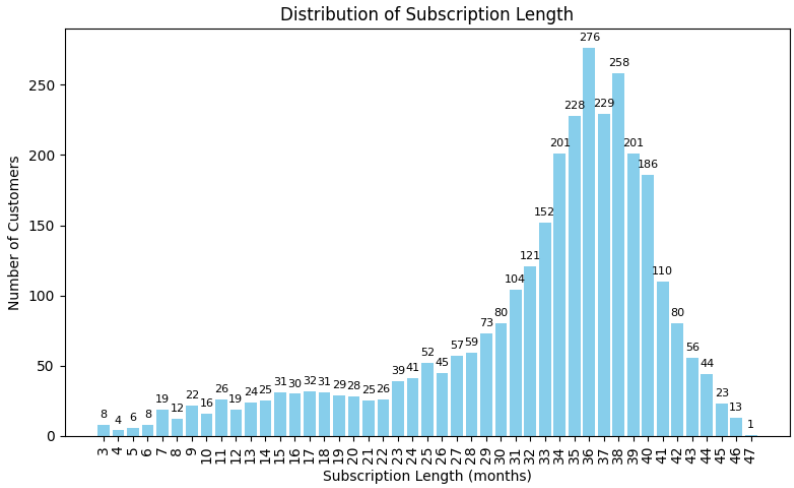


Figure 10. Distribution of Subscription Length

This chart shows how long customers have been subscribed to the service, measured in months. The distribution reveals a clear pattern: most customers fall within the 30–40 month range, with a peak around 36 months, indicating that the majority have been with the company for roughly three years. Early subscription lengths (under 12 months) have relatively few customers, suggesting limited recent onboarding or higher early-term churn. The long right tail extending to 47 months shows that a portion of customers have been with the company for four years or more, although these long-tenure customers become increasingly rare. Overall, the distribution is right-skewed and highlights that the company's customer base is dominated by medium- to long-term subscribers, an important factor when examining churn behavior and retention strategies.

Chart 6B. Boxplot of Subscription Length by Churn

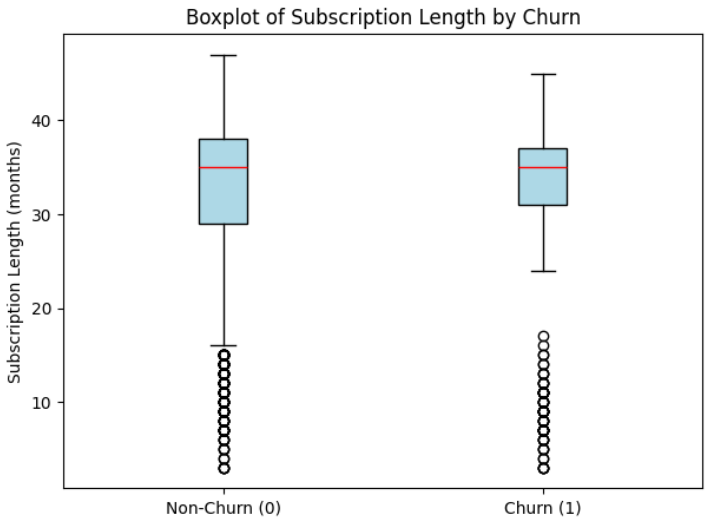


Figure 11. Boxplot of Subscription Length by Churn Status

This boxplot compares subscription length between churned and non-churned customers. The two groups appear broadly similar, with both distributions centered around roughly 33–36 months of subscription. However, churned customers tend to show a slightly lower median subscription length, suggesting that customers are somewhat more likely to churn earlier in their lifecycle. Both groups contain a number of short-tenure outliers (customers subscribed for fewer than 10 months), though these are more pronounced among churners, reinforcing the idea that newer customers are at greater risk of leaving. While subscription length alone does not sharply distinguish churn behavior, the subtle downward shift in median and the concentration of short-tenure churners indicate that tenure still plays a meaningful role in churn prediction.

4. Data Cleaning and Pre-processing

Data cleaning and pre-processing were performed to prepare the dataset for modeling. Since the dataset contained no missing values, the focus was on verifying logical consistency, checking for duplicates, and confirming that numerical and categorical features were correctly formatted. Outlier checks were conducted for usage, complaints, customer value, and subscription length to ensure that extreme values would not distort model performance. Feature scaling was applied where appropriate—particularly for models sensitive to magnitude differences, such as Logistic Regression—while tree-based models like XGBoost used the raw values.

a. Data Types & Basic Structure: The dataset consists of 3,150 customer records with 13 features capturing usage behavior, demographics, and service characteristics. All variables were inspected to ensure correct data types, with numerical fields stored as integers or floats and categorical variables properly encoded. This verification ensures that each feature is interpreted correctly by downstream analytical and machine learning processes.

b. Missing Values: The dataset was examined for missing values across all features, and no missing entries were found. Since the dataset was complete, no imputation or removal of records was required, allowing the analysis and modeling stages to proceed without additional handling of missing data.

c. Outliers:

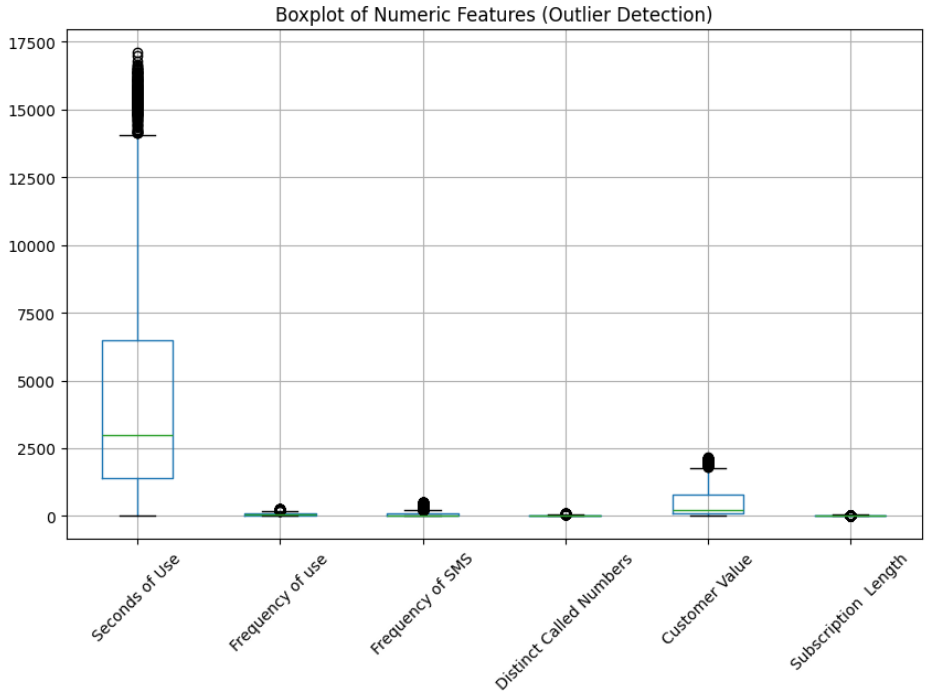


Figure 12. Boxplot of Numeric Features (Outlier Detection)

A boxplot of the numerical features was generated to identify potential outliers. Several variables—particularly Seconds of Use and Customer Value—contain a number of high-end outliers, reflecting customers with unusually high activity or value. These outliers appear to represent genuine customer behavior rather than data errors, as telecom usage can vary widely across individuals. Since tree-based models like XGBoost are robust to extreme values, and because these points may hold important behavioral information, no outlier removal was performed.

d. Duplicates: During the cleaning phase, 300 duplicate rows were identified. Duplicate records can bias the model by overrepresenting certain customers or behaviors. Therefore, all duplicates were removed, ensuring that each observation represents a unique customer. This step improves data integrity and prevents distorted model learning.

e. Logical Inconsistency: To ensure the dataset reflects realistic customer behavior, several logical consistency checks were performed. First, records with zero months of subscription but non-zero usage were identified, as these cases likely indicate data entry issues. Additionally, customers with unusually high Customer Value but very low usage were examined, since such patterns may be inconsistent with how customer value is calculated. Age and Age Group alignment was also inspected to verify that age categories corresponded logically to actual ages. Finally, the dataset was checked for impossible values, such as negative usage or negative subscription length. These checks help validate the integrity of the dataset and ensure that the modeling process is based on internally consistent data. These validations ensure that the dataset is internally consistent and reliable for modeling. No logical inconsistencies were detected in the data.

5. Implementation of 2 ML Models (Logistic Regression + XGBoost)

To predict customer churn, two machine learning models were implemented: Logistic Regression and XGBoost. Logistic Regression serves as a simple and interpretable baseline, while XGBoost provides a more powerful, non-linear approach capable of capturing complex relationships within the data. Both models were trained on the processed dataset and evaluated using standard classification metrics to compare their predictive performance.

a. Setup – Features, Target, Train/Test Split

The modeling process began by defining the input features (customer usage, demographics, and service attributes) and the target variable, Churn. The dataset was then split into training and testing sets using a 70/30 stratified split to preserve the original churn distribution. This setup ensures that the models learn underlying patterns from the training data and are evaluated fairly on unseen data.

b. Model 1 – Logistic Regression (with scaling)

Logistic Regression was used as the baseline model for churn prediction due to its simplicity, interpretability, and effectiveness in binary classification tasks. Because Logistic Regression is sensitive to differences in feature magnitude, numerical features were standardized before training to ensure balanced model behavior. Despite its simplicity, the model achieved strong recall, correctly identifying most churned customers, making it a useful reference point for evaluating more advanced models.

=== Logistic Regression ===
Accuracy : 0.8444
Precision: 0.5021
Recall : 0.8881
F1-score : 0.6415
ROC-AUC : 0.9347

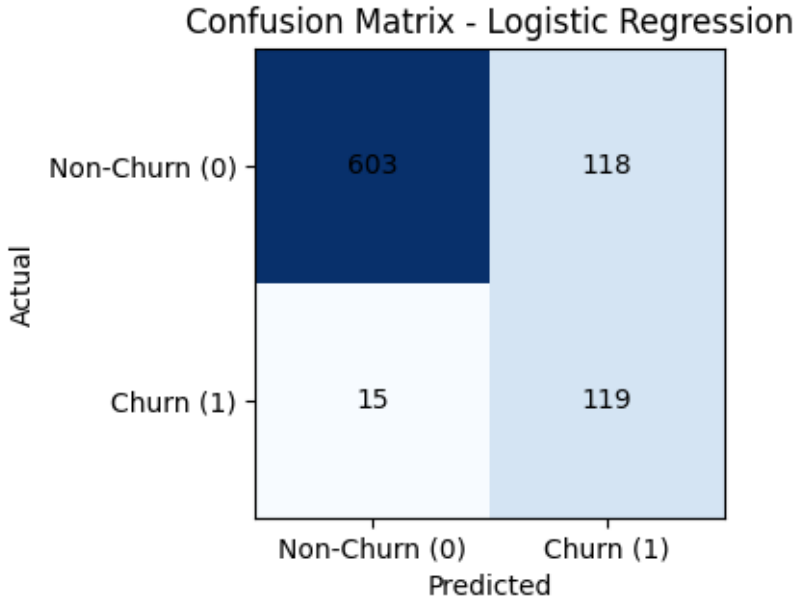


Figure 13. Confusion Matrix of Logistic Regression Model

Logistic Regression achieved an accuracy of 84.4% and a recall of 88.8%, meaning it correctly identified most churners. However, precision was relatively low (50.2%), indicating many false positives—customers incorrectly predicted as churners. This is visible in the confusion matrix, where the model misclassified 118 non-churn customers as churners. Logistic Regression is useful for understanding feature importance but is limited in predictive power.

c. Model 2 – XGBoost Classifier

XGBoost was selected as the second model due to its strong performance on structured tabular data and its ability to capture nonlinear relationships and complex interactions between features. Unlike Logistic Regression, XGBoost does not require feature scaling and is highly robust to outliers and varying feature magnitudes. The model achieved significantly higher accuracy, precision, and ROC–AUC compared to the baseline, making it the strongest predictor of churn in this project and the final model used for interpretation and insights.

=== XGBoost ===
Accuracy : 0.9497
Precision: 0.8760
Recall : 0.7910
F1-score : 0.8314
ROC-AUC : 0.9803

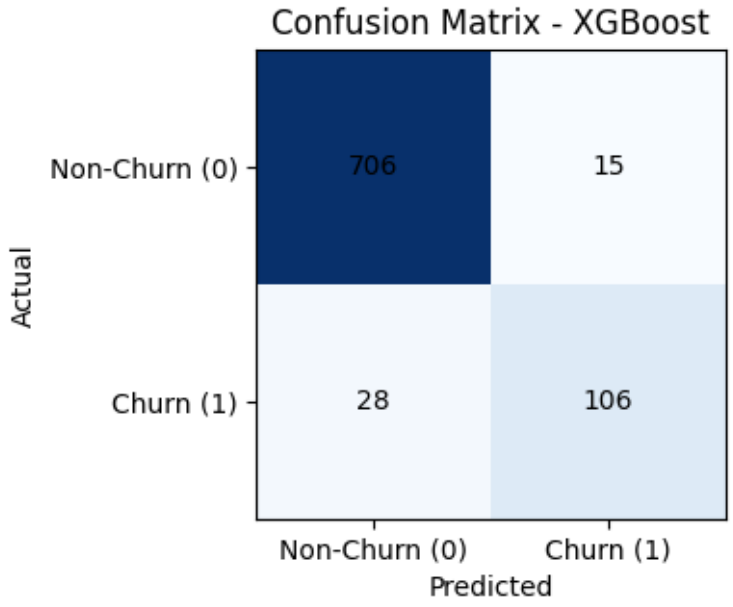


Figure 14. Confusion Matrix of XGBoost Model

XGBoost significantly outperformed the baseline model, achieving 94.97% accuracy and an ROC-AUC of 0.9803. Precision (87.6%) and F1-score (83.1%) were much higher than Logistic Regression, meaning XGBoost produced far fewer false positives while still maintaining strong recall (79.1%). The confusion matrix shows that XGBoost correctly classified most non-churn customers (706) and substantially reduced false positives (15).

Although it missed slightly more churners than Logistic Regression (28 vs. 15), it achieved a much better balance between precision and recall.

d. Comparison of Two Models

	ACCURACY	PRECISION	RECALL	F1-SCORE	ROC-AUC
LOGISTIC REGRESSION	0.844444	0.502110	0.888060	0.641509	0.934704
XGBOOST	0.949708	0.876033	0.791045	0.831373	0.980329

Table 1. Comparison of Two Models

Overall, XGBoost is the superior model for this dataset, delivering the strongest performance across nearly all metrics and achieving the highest ability to distinguish between churners and non-churners (ROC-AUC = 0.9803). Logistic Regression remains valuable for interpretation but is less effective for production-level churn prediction. The comparison demonstrates how more advanced tree-based models can capture complex customer behaviors that linear models fail to detect.

6. Hyperparameter Tuning

To further improve model performance, hyperparameter tuning was performed on the XGBoost classifier using RandomizedSearchCV. The search space included parameters such as the number of trees (n_estimators), tree depth (max_depth), learning rate, subsample ratio, and column sampling rate. The tuning procedure used 3-fold cross-validation on the training set and optimized the F1-score, which balances precision and recall for the churn class.

Although hyperparameter tuning identified a new combination of parameters, the tuned model did not outperform the original XGBoost model on the held-out test set. In particular, the tuned model achieved slightly lower F1-score and ROC–AUC compared to the default XGBoost configuration, suggesting that the baseline model was already close to optimal for this dataset. As a result, the original XGBoost model was retained as the final model for interpretation (feature importance and SHAP analysis) and for drawing business conclusions.

=== Tuned XGBoost ===
Accuracy : 0.9450
Precision: 0.8537
Recall : 0.7836
F1-score : 0.8171
ROC-AUC : 0.9798

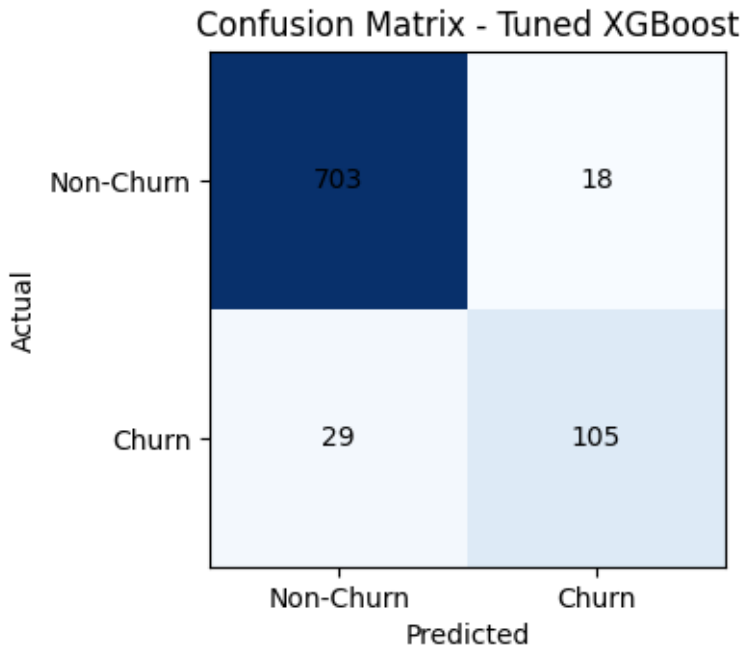


Figure 15. Confusion Matrix of Tuned XGBoost Model

7. Feature Contribution Analysis with SHAP

SHAP (SHapley Additive exPlanations) was used to interpret the final XGBoost model and understand how each feature contributes to churn predictions. SHAP provides both global and local explanations by showing how individual features push a prediction toward either churn or non-churn. This method is particularly valuable for churn analysis because it highlights which customer behaviors—such as low usage, short subscription length, or low customer value—have the strongest impact on churn risk. Using SHAP ensures that the model's decisions are transparent, interpretable, and aligned with meaningful business insights.

a. SHAP Summary Plot

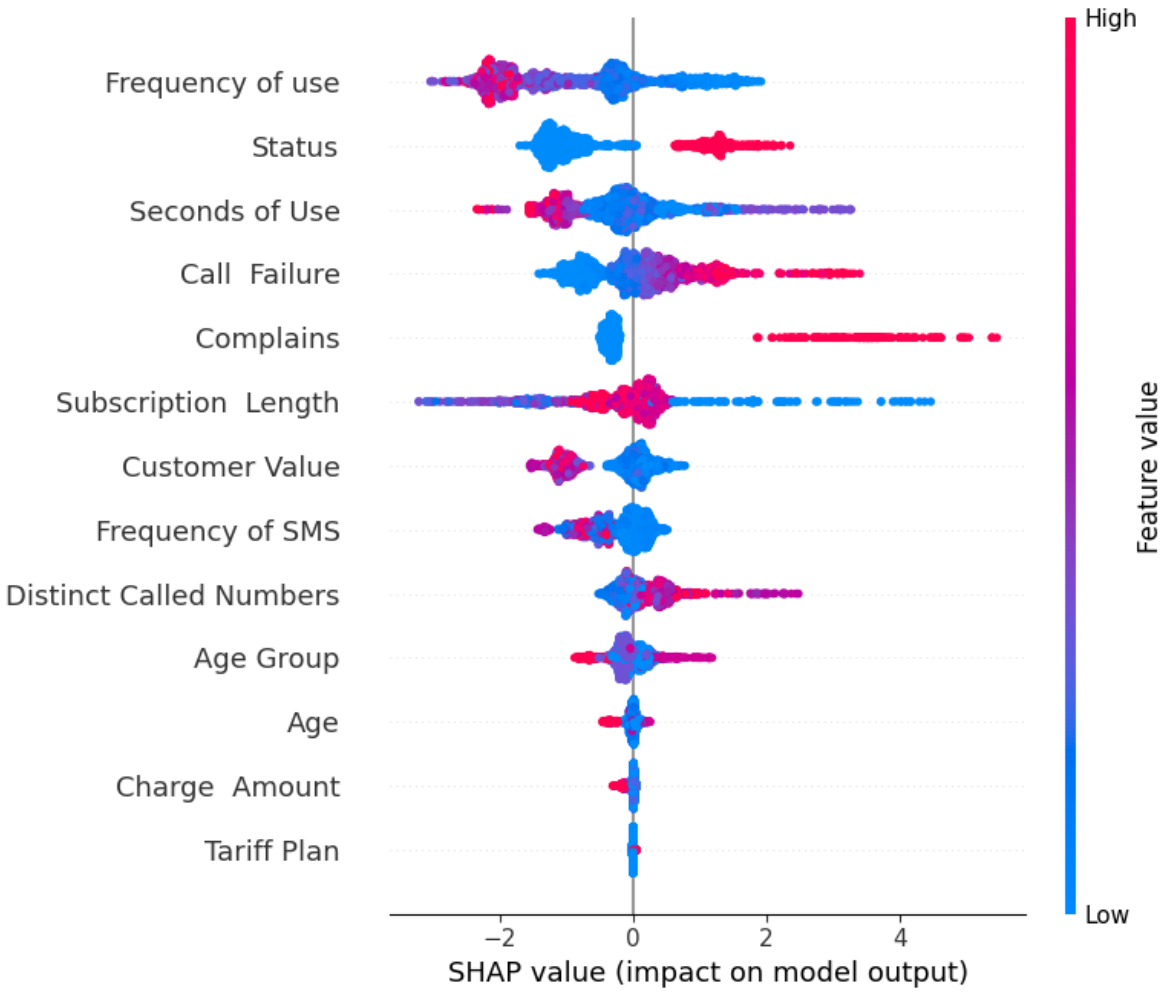


Figure 16. SHAP Summary Plot

The SHAP summary plot illustrates how each feature influences the XGBoost model's predictions for churn. Features at the top have the greatest overall impact, while the color gradient indicates whether high (pink) or low (blue) feature values push the prediction toward churn. The plot shows that shorter subscription length, high complain history, low customer value, and lower usage metrics (such as seconds of use and frequency of use) are major drivers of churn risk. In contrast, customers with longer tenure, higher usage, and greater customer value are more likely to remain active. This visualization provides a clear, interpretable view of the model's behavior and confirms that the most influential predictors align with known patterns in telecom churn.

b. SHAP Feature Importance

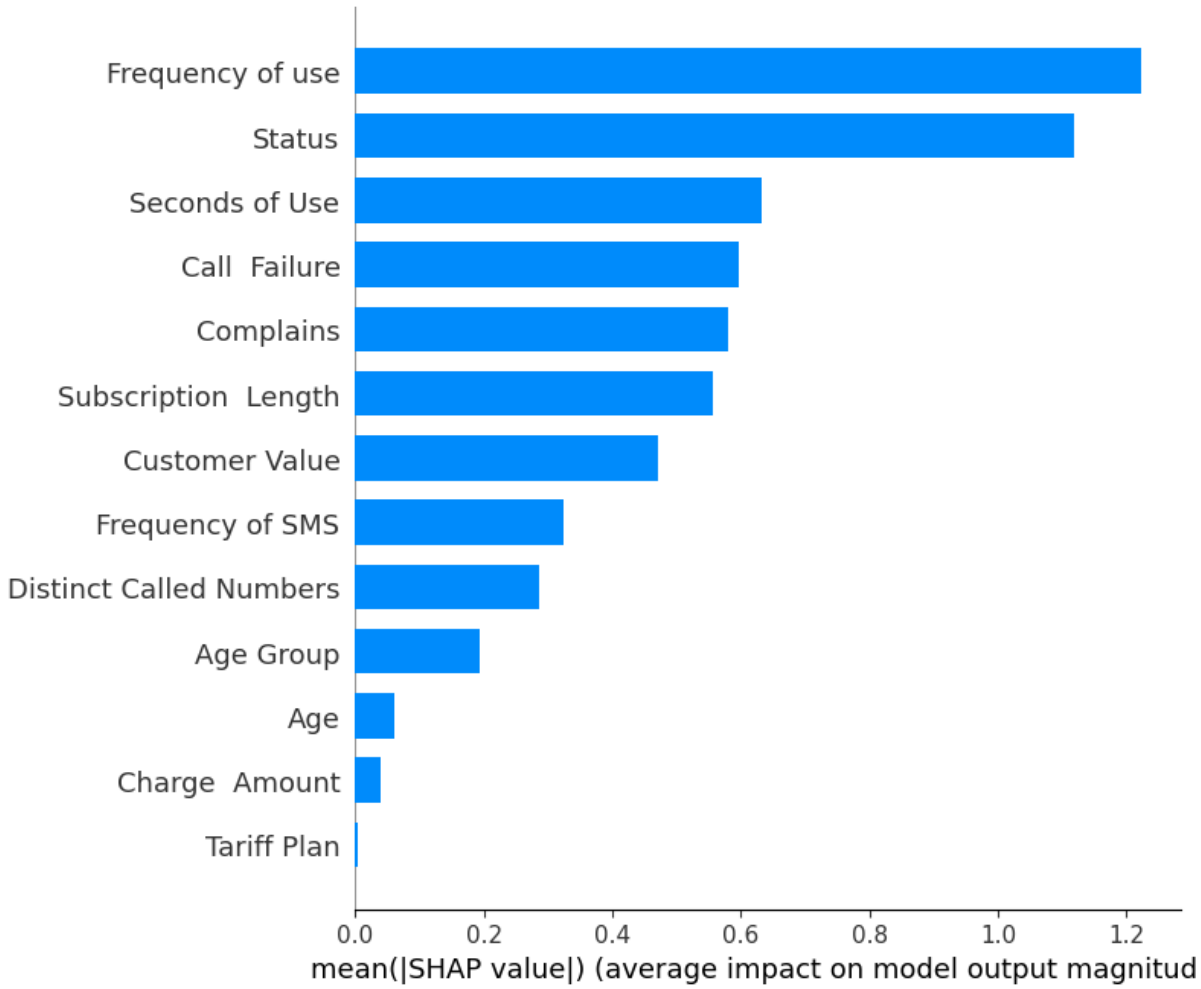


Figure 17. SHAP Feature Importance

This SHAP bar plot shows the average impact of each feature on the model's churn predictions. The longer the bar, the more important the feature is in influencing the model's decisions across all customers. The chart reveals that Frequency of Use, Status, and Seconds of Use are the most important predictors, meaning that how often a customer uses the service and their activity level strongly influence churn risk. Features like Call Failure, Complains, Subscription Length, and Customer Value also play meaningful roles, indicating that service issues, engagement, and customer worth affect the likelihood of churn. Lower-ranked features such as Age, Charge Amount, and Tariff Plan have minimal impact on the model. Overall, the bar plot provides a clear ranking of which factors matter most for churn prediction, supporting targeted business actions based on importance.

c. SHAP Force Plot (Decision Plot)

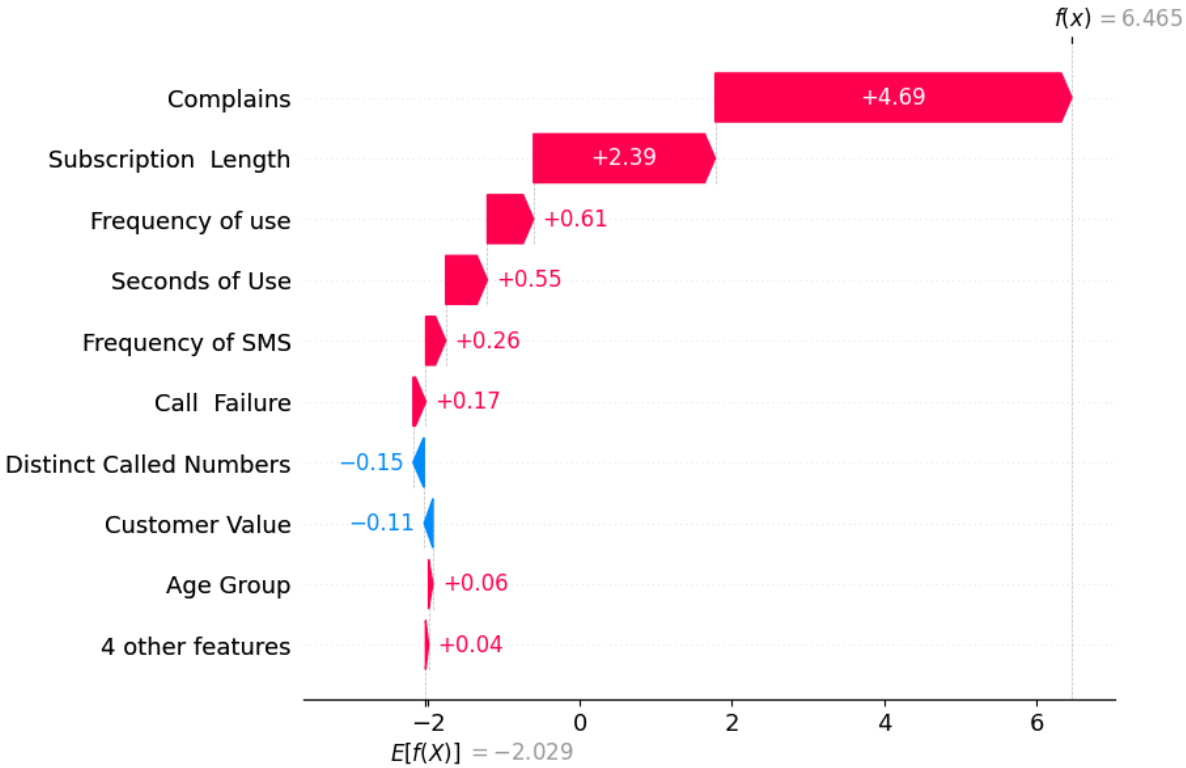


Figure 18. SHAP Force Plot (Decision Plot)

SHAP decision plot shows why the model predicted a specific customer (in this example 315) is likely to churn by breaking down how each feature pushes the prediction higher or lower.

8. Customer Churn Rate Prediction

To make the model usable for real-world decision-making, a prediction function was implemented that allows inputting the characteristics of any new customer and generating an estimated churn probability. The function collects feature values through user input, constructs a properly formatted DataFrame matching the model's training structure, and then applies the trained XGBoost model to compute both the predicted churn class (0 = non-churn, 1 = churn) and the associated probability. This enables telecom analysts or customer service teams to quickly assess the risk level of individual customers and proactively target retention strategies where they are most needed.

New Customer Churn Prediction
Customer info:
Call Failure: 3 --> A few failed calls
Complains: 1 --> Customer has complained
Subscription Length: 8 --> Only 8 months subscribed
Charge Amount: 4 --> Medium–low charge tier
Seconds of Use: 1200 --> Low call duration
Frequency of use: 15 --> Very few calls
Frequency of SMS: 5 --> Very low SMS usage

Distinct Called Numbers: 4 --> Limited social engagement
Age Group: 2 --> Younger customer
Tariff Plan: 1 --> Pay-as-you-go
Status: 1 --> Active customer
Age: 27 --> Age in years
Customer Value: 95

Prediction:
Predicted class: 1
Churn probability: 0.999

Using the trained XGBoost model, a churn prediction was generated for a sample customer based on their usage behavior, subscription history, and service experience. The model predicted a high likelihood of churn, assigning a probability of 0.999, primarily due to low usage levels, short subscription length, and the presence of a complaint. This demonstrates how the model can be applied to assess individual customer risk and support proactive retention strategies.

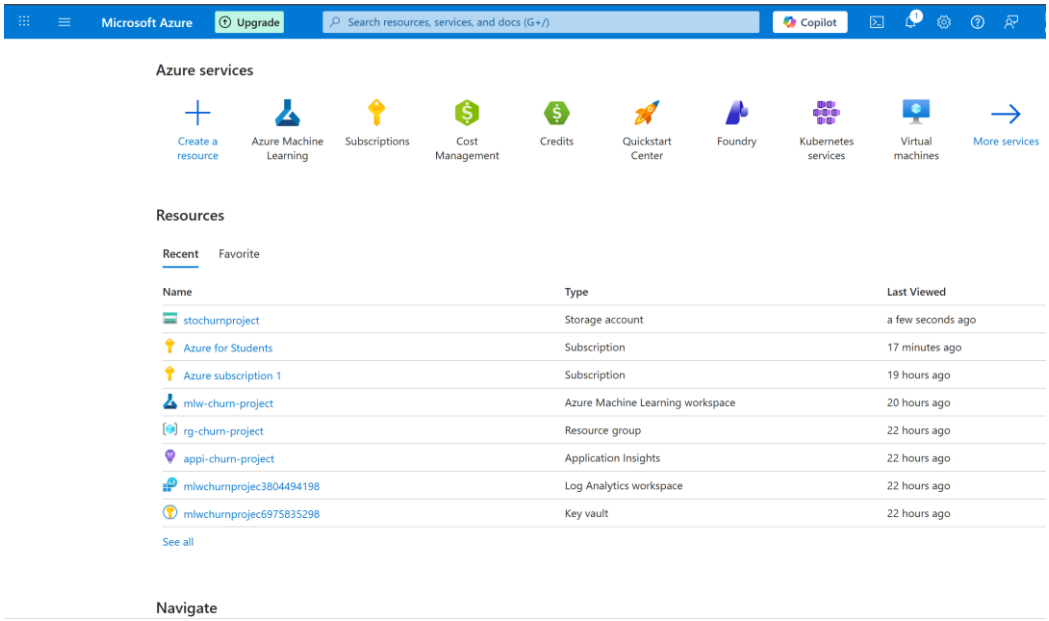
9. Conclusion

The analysis of the Iranian telecom churn dataset showed clear behavioral and value-based differences between customers who stayed and those who left. Churned customers consistently exhibited **lower usage (seconds and frequency of use)**, **shorter subscription length**, **lower customer value**, and were more likely to have **filed complaints**, confirming these factors as key drivers of churn. Logistic Regression provided a strong, interpretable baseline with high recall, but the **XGBoost model** performed substantially better overall, achieving higher accuracy, precision, F1-score, and ROC–AUC, and thus served as the final model for insight generation. Feature importance and SHAP analysis reinforced the finding that engagement intensity, tenure, and complaint history are the most influential predictors of churn. These results suggest that telecom providers should prioritize proactive retention strategies for low-usage, short-tenure, low-value customers—especially those who have recently complained. Future work could extend this study by incorporating time-based features, testing additional ensemble or deep learning models, or evaluating the financial impact of targeted interventions based on the model's churn risk scores.

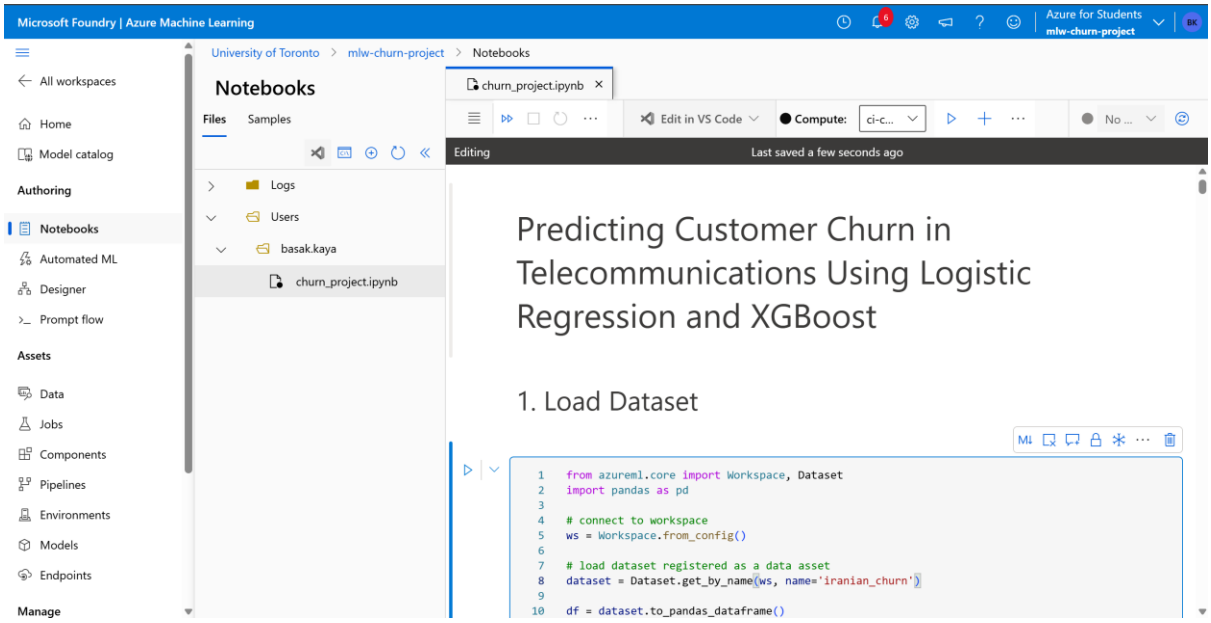
10. Microsoft Azure Machine Learning

To complete this project, Microsoft Azure Machine Learning was used as the primary platform for data preparation, model development, hyperparameter tuning, and interpretability analysis. Azure ML provided a cloud-based workspace for running notebooks, managing compute resources, and carrying out the full machine learning workflow efficiently.

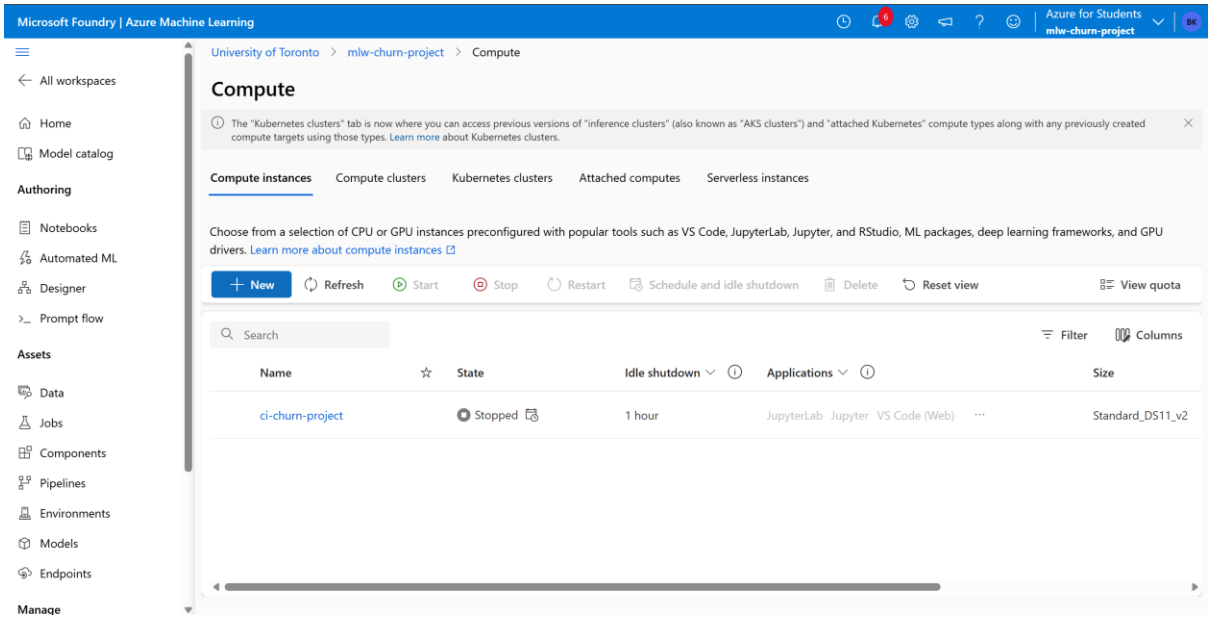
Resources:



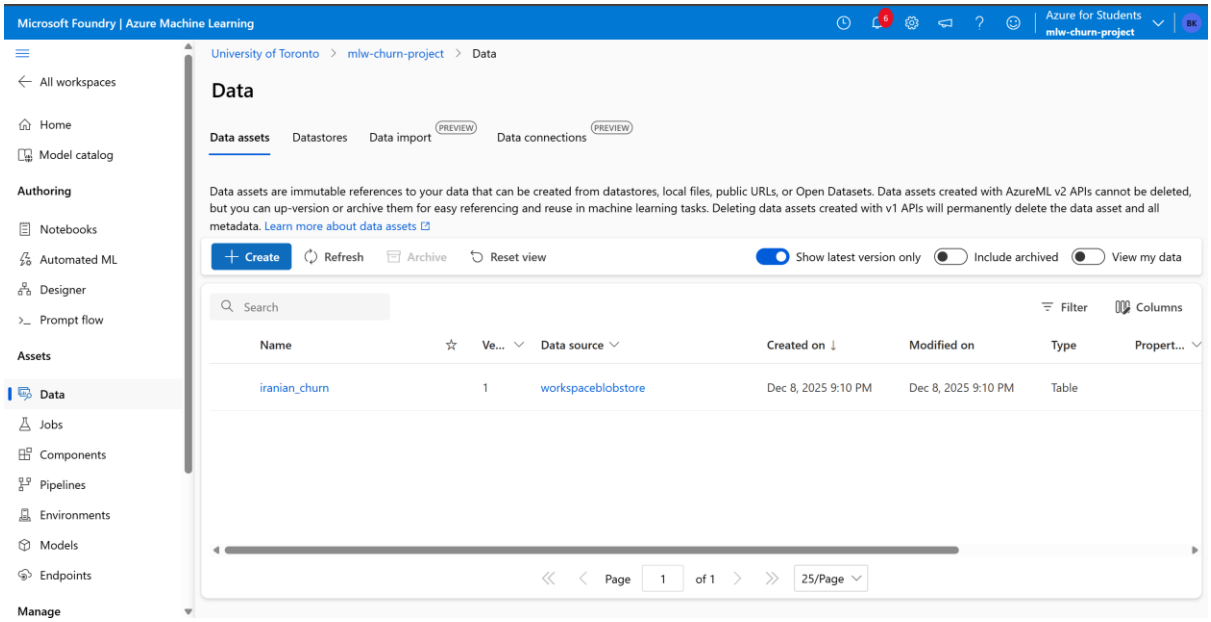
Azure ML Studio:

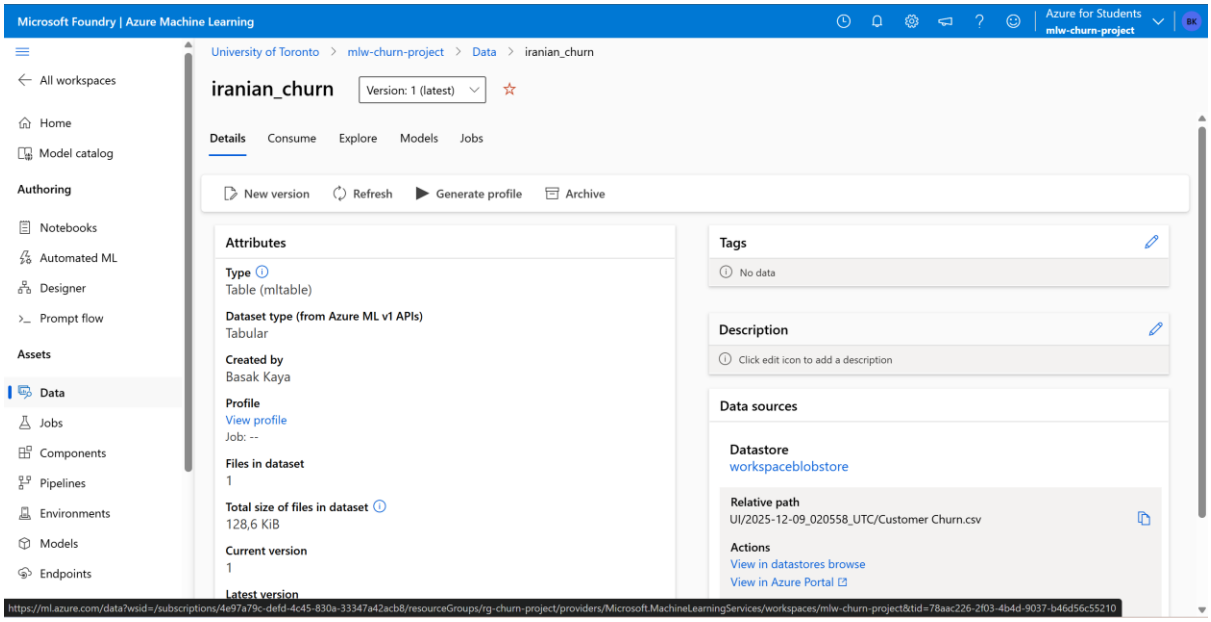


Compute Instance:



Dataset:




Microsoft Foundry | Azure Machine Learning
Azure for Students
mlw-churn-project
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
University of Toronto > mlw-churn-project > Data > iranian_churn
iranian_churn
Version: 1 (latest)
Details
Consume
Explore
Models
Jobs
New version
Refresh
Generate profile
Archive
Attributes
Type
Table (mltable)
Dataset type (from Azure ML v1 APIs)
Tabular
Created by
Basak Kaya
Profile
View profile
Job: --
Files in dataset
1
Total size of files in dataset
128,6 KiB
Current version
1
Latest version
Tags
No data
Description
Click edit icon to add a description
Data sources
Datastore
workspaceblobstore
Relative path
UI/2025-12-09_020558_UTC/Customer Churn.csv
Actions
View in datastores browse
View in Azure Portal
https://ml.azure.com/data?wsid=/subscriptions/4e97a79c-defd-4c45-830a-33347a42acb8/resourceGroups/rg-churn-project/providers/Microsoft.MachineLearningServices/workspaces/mlw-churn-project&tid=78aac226-2f03-4b4d-9037-b46d56c55210