

Customer Churn Prediction Using XGBoost

1. Introduction

Customer churn prediction is an essential task in industries like telecommunications, banking, and e-commerce. Churn refers to the loss of customers over a period, and predicting it allows businesses to take proactive steps to retain customers. This project aims to build a predictive model using XGBoost to classify customers into churn and non-churn categories based on historical data.

2. Dataset Overview

The dataset used in this project contains customer information, such as tenure, monthly charges, contract type, and various service subscriptions. The target variable is 'Churn', which indicates whether a customer has left the service.

Key Dataset Details:

- Total Records: [Insert Number]
- Number of Features: [Insert Number]
- Target Variable: Churn (0: No, 1: Yes)
- Missing Values Handling:
 - `TotalCharges`: Converted to numeric and missing values replaced with the median.
 - Other categorical features: Encoded using Label Encoding.

3. Data Preprocessing

Steps Taken:

- Dropped unnecessary columns (e.g., `customerID`).
- Converted categorical variables using Label Encoding.
- Scaled numerical features using StandardScaler.
- Split dataset into training (80%) and testing (20%).

4. Model Selection & Training

The XGBoost classifier was chosen for its efficiency in handling structured data. The model was trained using hyperparameter tuning with the following settings:

Model Hyperparameters:

- n_estimators: 200
- max_depth: 6
- learning_rate: 0.1
- subsample: 0.8
- colsample_bytree: 0.8
- random_state: 42

5. Performance Evaluation

The model's effectiveness was assessed using various metrics:

Evaluation Metrics:

Metric	Score
--------	-------

----- -----	
Accuracy	[Insert Value]
Precision	[Insert Value]
Recall	[Insert Value]
F1 Score	[Insert Value]

Confusion Matrix:

A confusion matrix was used to analyze true positives, false positives, true negatives, and false negatives.

6. Key Findings & Visualizations

Several visualizations were created to analyze the dataset and model performance:

- Confusion Matrix Heatmap
- Feature Importance Bar Chart
- Churn Distribution Pie Chart
- Performance Metrics Comparison (Bar Graph)

7. Conclusion

The XGBoost model achieved an accuracy of approximately [Insert Accuracy], demonstrating its effectiveness in predicting customer churn. Key factors influencing churn include [Insert Key Features]. Businesses can use these insights to implement retention strategies and improve customer satisfaction.

8. Future Work

To further improve the model, future enhancements can include:

- Experimenting with Deep Learning models.
- Adding customer feedback sentiment analysis.
- Implementing real-time churn prediction using streaming data.

This project provides a strong foundation for proactive customer retention strategies based on predictive analytics.