

## # Customer Churn Prediction - Final Report

### ## Project Overview

The objective of this project was to develop a predictive model to identify potential customer churn in the telecommunications sector. By analyzing customer data, the goal was to provide actionable insights to reduce churn and improve customer retention strategies.

### ## Data Overview

- **Dataset:** Telco Customer Churn Dataset from Kaggle
- **Size:** 7,043 entries, 21 columns
- **Target Variable:** Churn (Yes/No)
- **Key Features:** Tenure, MonthlyCharges, TotalCharges, Contract Type, Payment Method

### ## Data Preprocessing

1. **Data Cleaning:**
  - Removed null values in the 'TotalCharges' column by converting to numeric and imputing the median.
  - Dropped irrelevant columns such as 'customerID'.
2. **Feature Engineering:**
  - One-hot encoding for categorical variables (e.g., gender, contract type).
  - Standardization of numeric features (e.g., MonthlyCharges, TotalCharges).
3. **Data Splitting:**
  - 70% training set, 30% testing set

### ## Exploratory Data Analysis

- The distribution of tenure shows a significant spike at both low and high tenures, indicating potential contract completion or early termination.

![Tenure Distribution](attachment:file-UJtQeDBkKcPf58yom3baGF)

- MonthlyCharges exhibit a bimodal distribution, suggesting distinct pricing strategies or service packages.

![Monthly Charges Distribution](attachment:file-ECXMxg4Eb6kUG7UFgS3AZ1)

- Pairwise analysis highlights the relationship between tenure, MonthlyCharges, and TotalCharges, showing a clear separation between churn and non-churn customers.

![Pairwise Analysis](attachment:file-Nq1i1yCr9N4g53ySqqFEKh)

### ## Model Development

Three models were developed and evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1 Score
- ROC AUC

**\*\*Model Performance Summary:\*\***

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Logistic Regression	0.811	0.686	0.563	0.618	0.857
Random Forest	0.792	0.721	0.383	0.501	0.852
XGBoost	0.806	0.686	0.528	0.596	0.857

**## Best Model**

- **\*\*XGBoost\*\*** was identified as the best model with the highest ROC AUC score of 0.857.
- The model effectively balanced precision and recall, indicating its robustness in identifying churners without generating excessive false positives.

![[Model ROC AUC Comparison]](attachment:file-JE3fHFfBnAvPKicaspWSx1)

**## Recommendations**

1. **\*\*Target High-Risk Customers:\*\*** Focus marketing efforts on customers with high monthly charges and short tenure, as they are more likely to churn.
2. **\*\*Contract Optimization:\*\*** Promote long-term contracts to mitigate churn, as customers with longer tenures show lower churn rates.
3. **\*\*Billing Support Programs:\*\*** Implement targeted billing support for customers with high MonthlyCharges to prevent churn driven by financial strain.

**## Further Research**

- Explore time series analysis to detect churn patterns over time.
- Integrate additional customer touchpoints (e.g., customer support calls) to enhance model accuracy.
- Implement A/B testing to assess the effectiveness of targeted retention strategies.

**## Conclusion**

The project successfully developed a churn prediction model with actionable insights for retention strategies. The XGBoost model was the top-performing model, achieving an ROC AUC of 0.857, demonstrating its effectiveness in identifying potential churners. Future work will focus on refining the model through time-series analysis and integrating more granular customer data.