Detail project report

NoMoreChurn – Telco Risk Intelligence

Date: 25-May-2025

Submitted By :-

Soumen Baidya

Gourisankar Maity

Disha Jaiswal

Contain

Introduction

The telecommunication industry is highly competitive, with companies continuously seeking ways to retain their customers. One of the significant challenges faced by telecom providers is customer churn, where customers discontinue their service subscriptions. Identifying customers likely to churn, understanding the reasons behind their decisions, and taking strategic actions can reduce revenue loss significantly.

This project, titled NoMoreChurn – Telco Risk Intelligence,aims to predict customer churn using machine learning techniques and visualize insights through a comprehensive Power BI dashboard. The insights derived from the analysis help business stakeholders develop proactive strategies for customer retention.

Objectives

To analyze telecom customer data to identify factors influencing churn.

To build a machine learning-based prediction model to forecast customer churn probability.

To evaluate and compare various ML algorithms for optimal performance.

To present key insights through an interactive and user-friendly Power BI dashboard.

To support data-driven decisions for reducing customer churn and improving business strategies.

Problem Statement

Telecom service providers often lose customers due to various reasons such as service dissatisfaction, better offers from competitors, or high billing. Predicting such churn events in advance allows companies to intervene and offer personalized retention solutions. However, manual analysis of such large-scale data is impractical, which calls for an automated and intelligent system to predict churn and provide actionable insights.

Dataset Overview

The dataset used is a publicly available Telco customer churn dataset containing various customer attributes such as:

Demographics: Gender, Senior Citizen, Partner, Dependents

Services: Phone Service, Internet Service, StreamingTV, etc.

Billing: Monthly Charges, Total Charges, Payment Method, Contract

Tenure and Churn status

The dataset contains approximately 7,000 records.

Data Preprocessing

Missing Value Handling: Handled null and empty strings, especially in Total Charges.

Feature Encoding: Used Label Encoder and One Hot Encoding to convert categorical data.

Feature Scaling: Applied Standard Scaler to normalize numerical features.

Binning: Created Tenure bins to group customers by tenure duration.

EDA: Explored variable relationships using pair plots, histograms, and correlation heatmaps.

Model Building

Train-Test Split: Dataset split into 80% training and 20% testing.

Random Forest Classifier: Implemented with K-Fold Cross Validation (K=5). Hyperparameters optimized via GridSearchCV (n\estimators, max\depth).

XGBoost Classifier: Applied similar K-Fold strategy.

Grid Search CV used to tune learning_rate, max\depth, and n\estimators.

Voting Classifier: Combined Random Forest and XGBoost (soft voting) to improve performance.

Evaluation Metrics: ccuracy, Precision, Recall, F1-score, AUC-ROC Curve, Confusion Matrix.

Final Dataset Export for Power BI

After model prediction, the test set was enriched with:

Predicted Churn Label

Churn Probability

Key features: Tenure, Monthly Charges, etc.

This dataset was exported to CSV format for Power BI integration.

Final Dataset Export for Power BI

The dashboard is structured into three interactive sheets:

Sheet 1: Churn Overview & Model Performance

KPI Cards:

Total Customers (test set)

Actual Churned Customers

Predicted Churned Customers

Model Accuracy

Bar Chart: Actual vs Predicted churn comparison

Confusion Matrix: TP, FP, TN, FN breakdown

Cards: Precision, Recall

Gauge/Pie Charts: Actual Churn Rate and Predicted Churn Rate

Sheet 2: Churn Analysis by Customer Attributes

Churn Rate by Contract Type (Bar Chart)

Churn by Payment Method

Churn by Tenure Bins (0–12, 13–24, etc.)

Monthly Charges vs Total Charges (Scatter with churn color-code)

Demographic Comparison: Gender (Pie Chart)

Partner (100% stacked bar)

Dependents (100% stacked bar)

Churn Persona Card:

Interactive slicer showing Tenure, MonthlyCharges, and Churn Probability

Sheet 3: Churn Risk & Ticket Impact

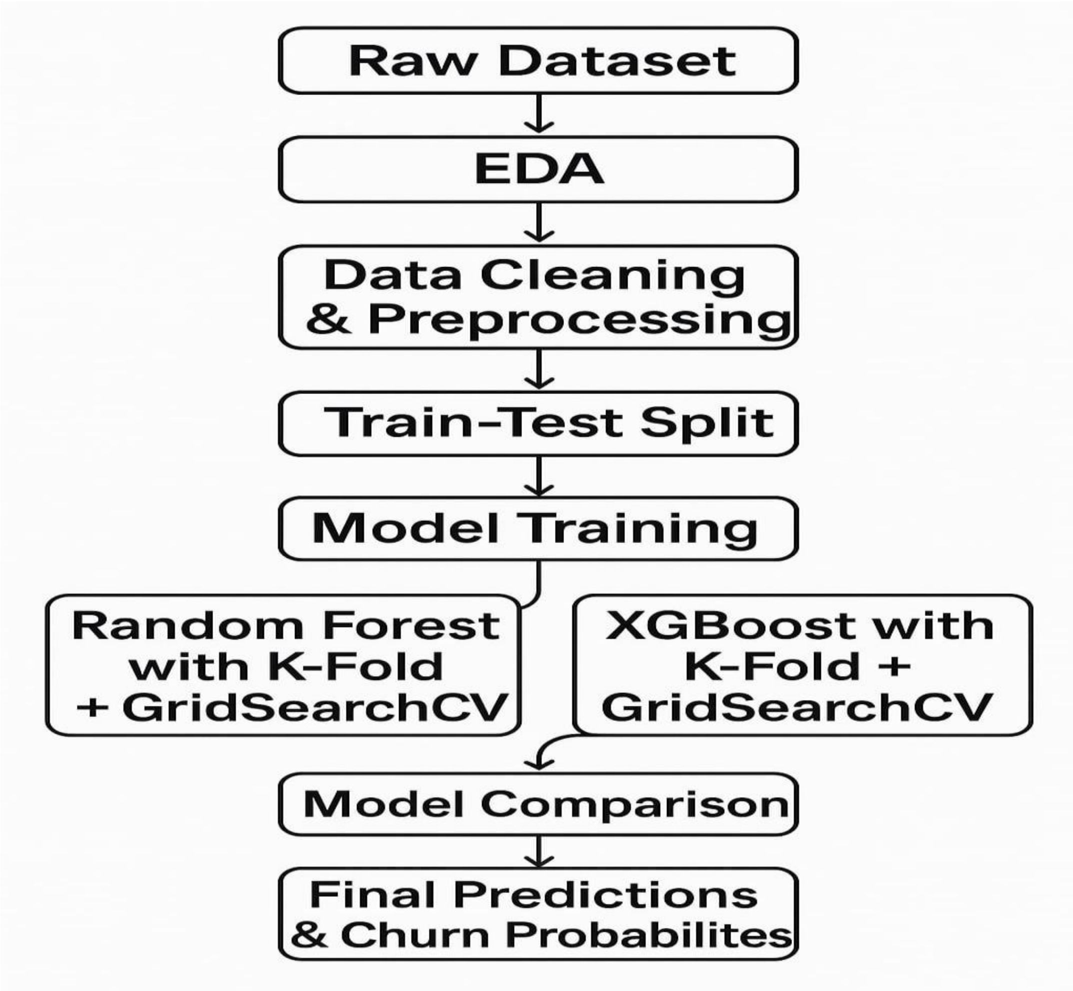
Top 10 High-Risk Customers Table

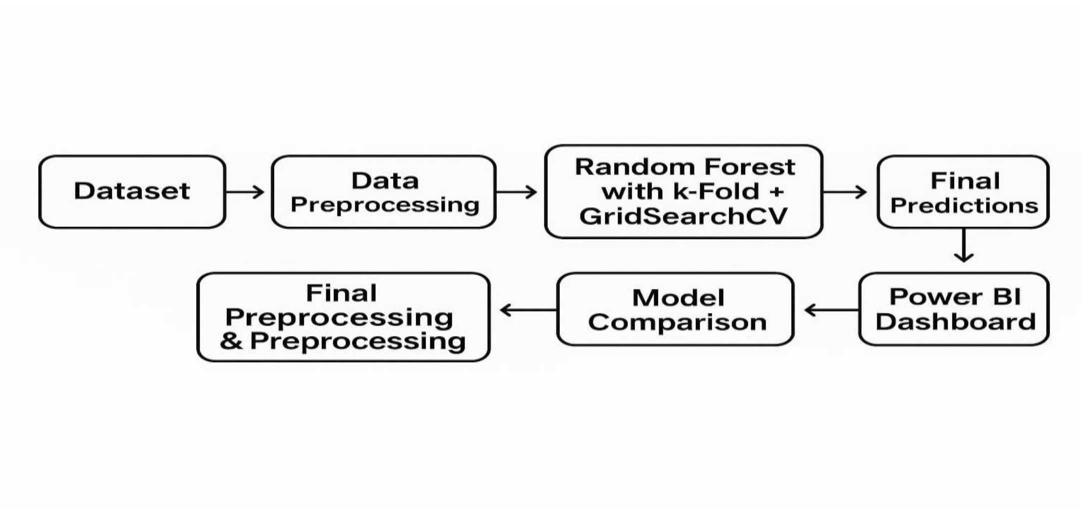
Includes churn probability, customer ID, tenure, charges

Churn Probability Banding Distribution

Feature-wise Histograms (for visual exploration)

Architecture and Functional Flow





Tools & Technologies Used

Python (pandas, scikit-learn, seaborn, matplotlib, xgboost)

Jupyter Notebook for development

Power BI for dashboard and visualization

Mermaid.js for architectural diagram

Challenges Faced

Handling null values in billing data

Data imbalance: Addressed using proper evaluation metrics

Hyperparameter tuning complexity

Creating interactive visuals that align with model output

Applications

Telecom churn prediction systems

Marketing campaign targeting

Customer segmentation and personalized offers

Business reporting and executive decision-making tools

Future Scope

Integrate real-time data pipelines for continuous churn monitoring

Incorporate NLP analysis from customer support data

Deploy predictive model on cloud platforms

Connect the dashboard with live databases using Power BI Gateway

References

https://www.kaggle.com/datasets/datazng/telecom-company-churn-rate-call-center-data/data

Scikit-learn Documentation XGBoost Documentation

Power BI Official Documentation