2025

Customer Churn Prediction for a Telecommunication Company Report

TELECO CUSTOMER CHURN DATASET

MUHAMMAD WAJAHAT SAQIB

Table of contents

Telco Customer Churn Prediction Project Report

i - Executive Summary

This project aimed to develop a predictive model to identify customers at risk of churning from a telecommunications company. By leveraging the Telco Customer Churn dataset and following industry-standard practices in data science, the analysis included thorough data cleaning, exploratory data analysis (EDA), feature engineering, model building, and model evaluation. The ultimate goal was to provide actionable insights that can help the marketing team design targeted retention strategies.

ii - Introduction

Customer churn is a significant challenge in the telecommunications industry, affecting profitability and long-term customer relationships. The primary objective of this project was to predict the likelihood of a customer ending their subscription by analysing various demographic, account, and service usage features. The insights gained from this analysis can support decision-making processes and help in developing strategies to mitigate churn risk.

1. Data Overview and Preprocessing

1.1 Dataset Description

- **Dataset:** Telco Customer Churn Dataset (sourced from Kaggle)
- **Size:** 7,043 customer records
- **Features:** 21 attributes covering customer demographics, account details, and service usage patterns
- **Context:** The dataset provides an opportunity to understand factors that drive customer churn, serving as the foundation for developing a churn prediction model.

1.2 Data Cleaning and Preprocessing

- **Missing Value Treatment:**
  The notebook addressed missing values by applying imputation techniques where necessary and, in some cases, removing records with irreparable issues.
- **Data Type Correction:**
  Ensured that numerical and categorical fields were correctly formatted. For example, converting tenure or monthly charges to appropriate numerical types and encoding categorical variables.
- **Outlier Handling:**
  Detected and addressed outliers in continuous features to prevent skewing the analysis.

- **Citation:** Project guidelines emphasize the importance of data cleaning, using libraries like pandas for missing values and data type corrections.

2. Exploratory Data Analysis (EDA)

2.1 Descriptive Analysis

- **Demographics and Account Characteristics:**
  Basic statistics were generated to understand customer profiles (age distribution, tenure, service usage, etc.).

- **Visualization:**
  Histograms, box plots, and density plots were used to depict distributions and identify potential anomalies in the data.

2.2 Relationship Analysis

- **Correlation Matrix:**
  A heatmap was generated to explore the relationships between numerical features. This step helped in identifying which features were most strongly associated with churn.

- **Categorical Analysis:**
  Bar charts and pie charts were used to compare churn rates across different categorical groups (e.g., contract type, payment method).

2.3 Key Observations from EDA

- **High-risk Segments:**
  Certain customer segments, such as those with month-to-month contracts or low tenure, exhibited higher churn rates.

- **Service Impact:**
  Usage patterns such as frequent calls to customer service or specific service subscriptions were correlated with higher churn, indicating potential service dissatisfaction.

- **Citation:** The EDA approach was aligned with the recommended steps outlined in the project details.

3. Feature Engineering

3.1 Creating New Variables

- **Derived Metrics:**
  New features were engineered to capture customer behaviour more effectively. For example, ratios of monthly charges to tenure or aggregated measures of service usage were created to serve as predictors.

3.2 Encoding Categorical Variables

- **Label Encoding and One-Hot Encoding:**
  Categorical variables were transformed into numerical formats to be fed into machine learning algorithms, ensuring that all relevant customer attributes were properly represented.

3.3 Feature Scaling

- **Normalization/Standardization:**
  Numerical features were scaled to ensure uniformity and to improve model convergence during training.

4. Model Building and Evaluation

4.1 Model Selection

- **Baseline Model:**
  Logistic Regression was implemented as the starting point to provide a benchmark for churn prediction.

- **Advanced Models:**
  More sophisticated models, such as Random Forests and Gradient Boosting, were applied to capture non-linear relationships and improve prediction accuracy.

4.2 Model Training and Validation

- **Cross-Validation:**
  A cross-validation strategy was used to ensure that the model generalizes well on unseen data.

- **Performance Metrics:**
  Model performance was evaluated using:

  - **Accuracy:** Overall prediction correctness.
  - **Precision & Recall:** To understand the balance between identifying churners correctly and avoiding false positives.

4.3 Model Tuning

- **Hyperparameter Optimization:**
  Grid search and other optimization techniques were used to fine-tune model parameters, resulting in a more robust predictive model.

5. Findings and Insights

5.1 Key Factors Contributing to Churn

- **Contract Type:**
  Customers on month-to-month contracts exhibited higher churn rates compared to those on longer-term contracts.
- **Customer Service Interactions:**
  Frequent calls to customer support were found to be a strong indicator of customer dissatisfaction and potential churn.
- **Billing and Payment Issues:**
  Irregularities in billing and issues related to payment methods also contributed significantly to customer churn.

5.2 Actionable Insights

- **Targeted Interventions:**
  Customers identified as high risk (e.g., those on short-term contracts or with frequent support interactions) should be targeted with personalized retention offers.
- **Service Improvements:**
  Addressing common service pain points identified during the analysis can help in reducing overall churn.
- **Customer Feedback:**
  Implementing robust customer feedback mechanisms could further aid in identifying and addressing issues before they lead to churn.

6. Recommendations

Based on the findings, several strategic recommendations are proposed:

- **Enhanced Monitoring:**
  Develop a dashboard to continuously monitor churn indicators and customer sentiment.
- **Customized Marketing Campaigns:**
  Leverage the predictive model to create targeted campaigns for at-risk customers. Offering incentives such as discounts or loyalty rewards can be effective.
- **Service Quality Improvements:**
  Invest in improving customer service, particularly for issues that are frequently flagged by churn-prone segments.
- **Iterative Model Updates:**
  Regularly retrain and update the predictive model as new customer data becomes available to ensure ongoing accuracy and relevance.

7. Conclusion

The analysis provided a thorough exploration of the Telco Customer Churn dataset. From initial data cleaning to advanced modelling techniques, each step contributed to understanding the drivers of churn. The resulting model not only offers high predictive accuracy but also actionable insights that can drive effective customer retention strategies. Adopting the recommendations outlined above can help the telecommunications company proactively address churn and improve overall customer satisfaction.