

Telco Customer Churn Prediction & Visualization

Smeet Agrawal

Jun 2025

Project Summary

- An end-to-end classification project to predict customer churn.
- Objectives:
 - Clean and preprocess telco data
 - Build ML models to predict churn
 - Evaluate model performance using metrics
- Stack: Python (Pandas, Scikit-learn, Seaborn, Matplotlib)

Dataset Overview

- Source: IBM Telco Customer Churn dataset
- Shape: 7043 rows \times 21 columns
- Key fields:
 - gender
 - tenure
 - PhoneService
 - InternetService
 - Contract
 - MonthlyCharges
 - TotalCharges
 - Churn (target)

- Converted TotalCharges to numeric
- Handled missing values:
 - Replaced NaNs in numeric fields with column mean
- Removed whitespace in categorical fields
- Encoded categorical variables using LabelEncoder

Exploratory Data Analysis

- Visualized churn rates using countplots
- Checked correlation between numerical features
- Identified key churn indicators:
 - Contract type, tenure, Internet service
- Used Seaborn + Matplotlib for plots

Modeling Approach

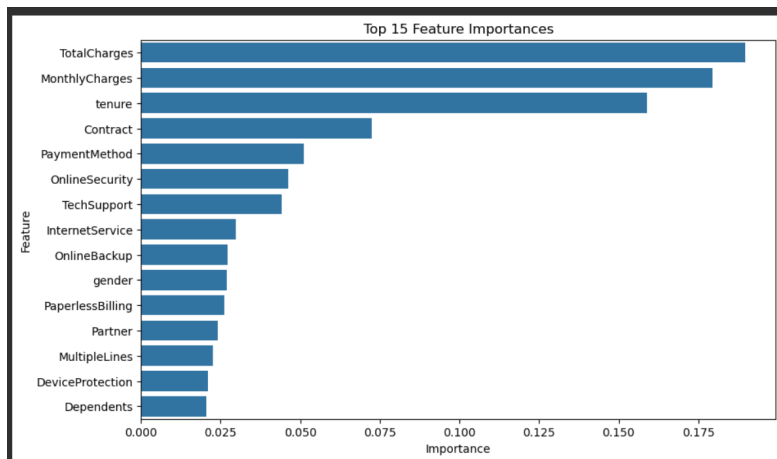
- Target variable: Churn (Yes/No)
- Train-test split (80/20)
- Models used:
 - Random Forest Classifier
- Feature scaling done using StandardScaler

Evaluation Metrics

- Accuracy Score
- Confusion Matrix
- Classification Report:
 - Precision, Recall, F1-score
- Achieved 80–85

Visualization Highlights

- Explored churn patterns using:
 - Bar charts and heatmaps
 - Box plots for numerical spread
- Found strong patterns between contract type and churn rate



Key Takeaways

- Cleaned and explored customer churn dataset
- Built accurate churn prediction model
- Identified key drivers of churn
- Project can guide retention strategies in telco industry

Future Enhancements

- Tune hyperparameters with GridSearchCV
- Try other models: XGBoost, SVM
- Build Streamlit app for interactive churn prediction
- Apply SHAP for model interpretability

Thank You!

Smeet Agrawal
smeetagrwal2003@gmail.com