Customer Churn Prediction Report

This report outlines the complete process of predicting customer churn using the Telco Customer Churn dataset. The objective is to identify customers who are likely to discontinue the service. The prediction model was built using Python, and the methodology includes data preprocessing, encoding, model training, evaluation, and feature importance analysis.

Dataset: Telco Customer Churn (publicly available via IBM) - Records: 7043 - Target: 'Churn' (Yes/No) - Features: Customer demographics, service usage, contract information, billing.

Step	Description
1. Load Data	Dataset was loaded from a public GitHub source.
2. Data Cleaning	Removed missing TotalCharges, dropped CustomerID.
3. Encoding	Binary columns label-encoded; categorical columns one-hot encoded.
4. Scaling	Used StandardScaler to normalize features.
5. Model	Trained a RandomForestClassifier with 100 trees.
6. Evaluation	Used confusion matrix and classification report for performance.
7. Feature Importance	Plotted top 10 important features influencing churn.

The churn prediction model built using Random Forests yielded good results and helped identify key factors like MonthlyCharges, tenure, Contract type, and OnlineSecurity in customer retention. The model can be further enhanced by incorporating time-based features or deploying it into a customer relationship management system for proactive intervention.