Customer Churn Prediction - Project Report

Introduction:

Customer churn, which refers to customers discontinuing a service, is a significant challenge for subscription-based businesses, particularly in the telecom sector where competition is high and switching costs are low. High churn rates can lead to substantial revenue loss and increased costs in acquiring new customers. By accurately predicting which customers are likely to leave, companies can take proactive measures such as offering personalized incentives, improving service quality, or addressing customer concerns in advance. This helps in enhancing customer satisfaction, building long-term loyalty, and ultimately improving business profitability. Predictive analytics thus becomes a valuable tool in churn management strategies.

Abstract:

This project leverages machine learning techniques to predict whether a customer is likely to churn, using historical data from the telecom industry. The Telco Customer Churn dataset serves as the foundation for analysing various customer attributes and behaviours, such as contract type, tenure, and payment method. By training predictive models on this data, the project aims to uncover patterns associated with churn. These insights can support data-driven decision-making and help businesses implement targeted strategies to retain customers and reduce churn rates.

Tools Used:

- Python
- Jupyter Notebook
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn (RandomForest, train/test split, metrics)
- LabelEncoder for preprocessing

Steps Involved in Building the Project:

1. Data Loading:

- Import the Telco Customer Churn dataset using Pandas.

2. Data Cleaning:

- Dropped customer ID.
- Converted TotalCharges to numeric and handled missing values.
- Encoded the 'Churn' column as binary (Yes=1, No=0).

3. Exploratory Data Analysis (EDA):

- Plotted churn distribution using countplots.
- Visualized correlation among numeric features with a heatmap.

4. Feature Engineering:

- Encoded categorical variables using LabelEncoder.

5. Model Building:

- Split data into training and testing sets.
- Trained a Random Forest classifier.
- Evaluated performance using confusion matrix, classification report, and accuracy score.

6. Feature Importance:

- Identified key factors influencing churn using model-based feature importance.

Conclusion:

Our Random Forest model demonstrated high accuracy in predicting customer churn, making it a reliable tool for early churn detection. Key features that influenced churn included contract type, customer tenure, and payment method, indicating strong behavioral patterns. With additional hyperparameter tuning and potential real-time deployment, the model can be further optimized. This predictive approach can enable telecom providers to implement proactive retention strategies, ultimately reducing churn and improving long-term customer satisfaction.