# Customer Churn Prediction: Project Report

## 1. Introduction

Customer churn, the phenomenon where customers stop using a company's products or services, poses a significant challenge for businesses. Predicting churn allows companies to take proactive measures to retain customers, thereby improving revenue and customer satisfaction. This project focuses on building a predictive model to analyze customer churn using data analysis and machine learning techniques.

## 2. Objective

The primary objective of this project is to:

1. Analyze customer behavior and identify factors contributing to churn.
2. Build a predictive model to accurately classify customers as likely to churn or retain.
3. Provide actionable insights to reduce churn and improve customer retention strategies.

## 3. Dataset Description

The dataset used for this analysis consists of 7043 rows and 21 columns, with attributes related to customer demographics, services, and account information. Key features include:

- **Customer Information:** Gender, age, partner status, dependents.
- **Service Information:** Type of internet service, additional services (e.g., online security, streaming TV).
- **Contract Details:** Contract type, payment method, monthly charges.
- **Target Variable:** `Churn` (Yes/No).

### Data Preprocessing

1. Checked for missing values and found no significant missing data.
2. Converted relevant columns, such as `TotalCharges`, to appropriate numeric types.
3. Encoded categorical variables to prepare the data for modeling.

## 4. Methodology

### 4.1 Exploratory Data Analysis (EDA)

- Visualized churn distribution to understand class imbalance.
- Analyzed the impact of numerical features (e.g., `tenure`, `MonthlyCharges`) on churn.
- Examined correlations among features to identify key drivers of churn.

### 4.2 Feature Engineering

- Created dummy variables for categorical columns.
- Normalized numerical features to standardize scales.

### 4.3 Modeling

- Split the dataset into training and testing sets.
- Implemented and evaluated multiple machine learning models, including:
    - Logistic Regression
    - Decision Trees
    - Random Forest
    - Support Vector Machines (SVM)

### 4.4 Model Evaluation

- Used metrics such as accuracy, F1-score, and confusion matrix to evaluate model performance.
- Fine-tuned hyperparameters to optimize model results.

## 5. Key Findings

1. **Important Features:**
    - Customers with shorter tenure were more likely to churn.
    - Higher monthly charges correlated with increased churn rates.
    - Contract type (e.g., month-to-month contracts) significantly influenced churn likelihood.
2. **Model Performance:**
    - Random Forest achieved the best results, with an accuracy of ~85% and a balanced F1-score.
3. **Insights:**
    - Offering discounts or incentives for customers with month-to-month contracts can improve retention.
    - Enhancing customer experience for high-paying customers is crucial.

## 6. Benefits of Analysis

- **Customer Retention:** Enables targeted retention strategies, reducing churn.
- **Revenue Growth:** Retaining customers is more cost-effective than acquiring new ones.
- **Business Insights:** Provides actionable insights into customer behavior and preferences.

## 7. Conclusion

The project successfully built a predictive model to analyze and predict customer churn. By identifying key factors driving churn and implementing targeted retention strategies, businesses can enhance customer satisfaction and reduce revenue losses. Future work could involve:

- Integrating additional data, such as customer feedback, for deeper insights.
- Deploying the model into production for real-time churn prediction.

---

**Prepared By:** Priyanshu Mishra