CUSTOMER CHURN PREDICTION – DETAILED PROJECT REPORT

Customer churn occurs when a customer discontinues using a service. For telecom companies, predicting churn is critical because retaining an existing customer is far cheaper than acquiring a new one. This project uses machine learning techniques to predict customer churn based on demographic, account, and service-related features from the Telco Customer Churn dataset.
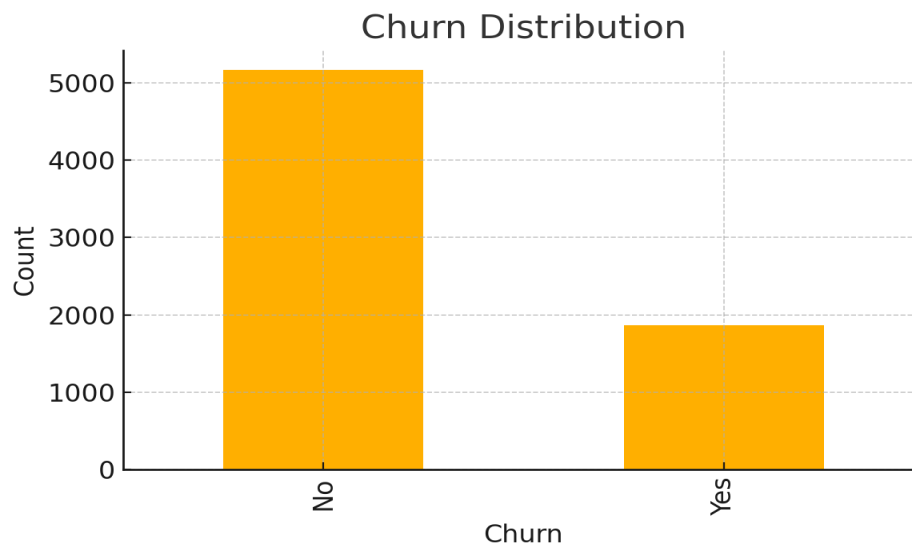
Dataset Overview
The dataset contains 7,043 rows and 21 features, including:
• Customer Account Information
• Demographic Information
• Service Subscription Details
• Billing Information

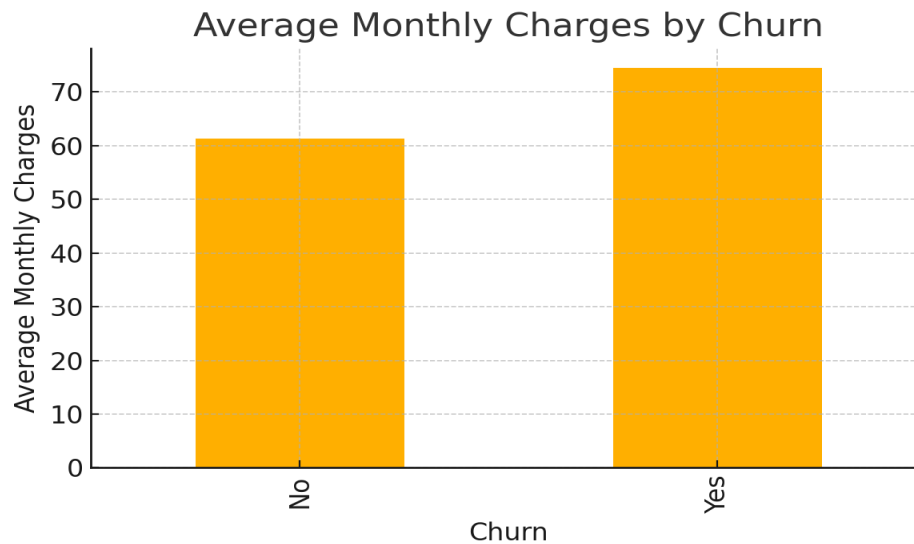**Target Variable:** Churn (Yes/No)

Churn Distribution

Data Cleaning
• Converted TotalCharges to numeric type
• Removed rows with missing values
• Dropped irrelevant columns such as CustomerID

Handling Missing Values:
Missing values were mainly in the TotalCharges column due to blank entries. These were converted or dropped to ensure clean training data.

Monthly Charges and Churn Relation

Average Monthly Charges by Churn

Data Preprocessing
Preprocessing included:
• Converting Yes/No values to binary (1/0)
• Encoding categorical variables using one-hot encoding
• Standardizing numerical features (tenure, MonthlyCharges, TotalCharges)
• Splitting the data into training and test sets (80/20 ratio)

Model Development
Two machine learning models were developed:

1. Logistic Regression
• Baseline model
• Accuracy ≈ 75%

2. Random Forest Classifier
• Performed significantly better
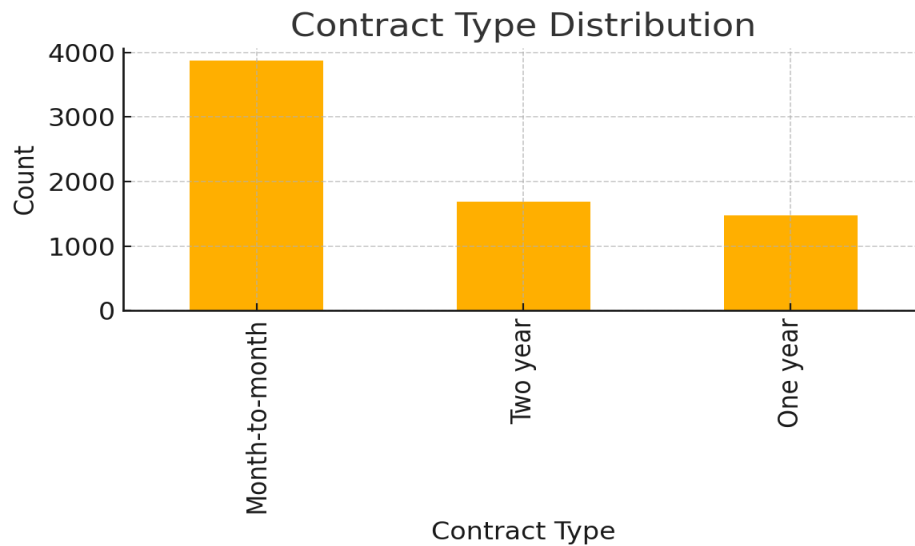• Accuracy ≈ 78%
• Captures nonlinear patterns

Hyperparameter Tuning (GridSearchCV)
Fine-tuned model performance using different combinations of depth, estimators, and leaf size.
Achieved:
• Best ROC-AUC ≈ 0.83
• Best Estimators: 400 trees, depth=6

Contract Type Distribution

Business Insights
• Customers with month-to-month contracts churn more.
• Higher churn observed in Fiber Optic internet users.
• Customers without Online Security and Tech Support are at higher risk.
• Low tenure customers are most likely to churn.

Final Model Selected: Tuned Random Forest Classifier

Conclusion
This churn prediction project successfully applied machine learning methods to identify customers likely to leave the telecom service. The tuned Random Forest model delivered strong predictive performance and provided meaningful insights to reduce churn rates. These insights can be used for targeted retention strategies, improving customer satisfaction, and reducing company losses.