



# Customer Churn Analysis – Project Report

---



## Project Overview

- The objective of this project is to analyze customer behavior and predict whether a customer is likely to **churn (leave the service)**.
  - Dataset used: **Telecom Customer Churn Dataset** containing demographic, account, and service-related customer details.
  - Machine Learning models were applied to classify customers into **Churn = Yes/No**.
  - Helps telecom companies to **retain customers**, optimize service plans, and improve user satisfaction.
- 



## Project Goals

- Identify reasons contributing to customer churn.
  - Perform Exploratory Data Analysis (EDA) to discover patterns.
  - Build machine learning models to predict churn accurately.
  - Determine key features influencing churn decisions.
  - Improve business strategy using predictive insights.
- 



## Dataset Information

- Total Records: **1000 customers**
- Total Features: **10**
- Target Variable: **Churn**
- Major Columns:
  - **CustomerID**
  - **Age**
  - **Gender**
  - **Tenure**
  - **MonthlyCharges**
  - **InternetService**
  - **TotalCharges**
  - **TechSupport**

- **ContractType**
  - **Churn (Label)**
- 

## Data Preprocessing

- Removed unnecessary column: **CustomerID**
  - Converted TotalCharges to numeric and handled missing values.
  - Filled missing values in **InternetService** using **mode()**.
  - Label encoded categorical columns (Gender, TechSupport, etc.)
  - Standardized numerical values using **StandardScaler**.
  - Split dataset into **Training (75%)** and **Testing (25%)** partitions.
- 

## Exploratory Data Analysis (Key Insights)

- Customers with higher **Monthly Charges** tend to churn more.
  - **Short-tenure** customers have a higher churn probability.
  - Customers without **Tech Support** churn more frequently.
  - **Contract Type** strongly influences churn:
    - Month-to-Month contract has highest churn rate, while 1- or 2-year contract customers are more stable.
- 

## Machine Learning Models Used

Model	Performance
Random Forest Classifier	Best accuracy
Logistic Regression	Baseline model
SVM / XGBoost (optional testing)	Comparable results

---

## Model Evaluation

- Best Model: **Random Forest Classifier**
- Performance Metrics:
  - **Accuracy:** 0.80+ (Depending on dataset variations)
  - **Precision, Recall, F1-score evaluated**

- Confusion Matrix used to analyze classification results.
- 

## ❖ Final Results

- Model successfully predicts customer churn with high reliability.
  - Important features affecting churn:
    - **MonthlyCharges**
    - **Tenure**
    - **ContractType**
    - **TechSupport**
    - **InternetService**
- 

## 📦 Business Impact

- Helps businesses **proactively target high-risk customers**.
  - Supports strategic decisions like:
    - Personalized offers for high-churn-risk customers
    - Improvement in service performance (Internet, Tech Support)
    - Contract upgrade campaigns
- 

## 🚀 Future Enhancements

- Deploy the model using **Streamlit Web App**
  - Integrate with real-time CRM system
  - Add additional customer behavior features
  - Try advanced models like **XGBoost, ANN, or Deep Learning**
- 

## 💡 Conclusion

- Customer churn prediction using machine learning provides powerful insights for telecom operators.
  - Predictive analytics helps companies save revenue by improving retention strategies.
  - The project demonstrates the complete pipeline from **data cleaning to model prediction** and **business decision recommendations**.
- 

## ⭐ Thank You

