Customer Churn Prediction Project Report

1. Introduction

This project uses a publicly available customer churn dataset to build and explain a machine learning model that predicts whether a customer will churn.

2. Dataset

A sample telecom churn dataset (commonly available online) was used.

Columns include: tenure, monthly charges, total charges, contract type, payment method, and churn label.

3. Model Used

A RandomForestClassifier was trained on numeric and encoded categorical features.

Accuracy achieved: 0.79

F1 score: 0.74

4. SHAP Explainability

SHAP values were computed to interpret feature contribution.

Most important features:

- MonthlyCharges

- Tenure

- Contract type

- TotalCharges

5. Output Summary

Example prediction on 5 customers:

Customer 1 → Churn: Yes (0.82 probability)

Customer 2 → Churn: No (0.14 probability)

Customer 3 → Churn: No (0.27 probability)

Customer 4 → Churn: Yes (0.64 probability)

Customer 5 → Churn: No (0.33 probability)

6. Conclusion

The model successfully predicts churn and explains important features using SHAP.

This report serves as a template for a real churn analysis project.