Customer Churn Prediction - Classification Project Report

1. Problem Overview

Customer churn refers to customers discontinuing a service. The objective of this project is to predict whether a customer will churn (1) or not churn (0) using customer demographic, account, and service usage data from the Telco Customer Churn dataset.

2. Dataset and Preprocessing

The dataset contains approximately 7,000 customer records with features such as gender, tenure, contract type, payment method, and charges.

Preprocessing steps included:

- Removal of non-informative columns (customerID)

- Label encoding of categorical variables

- Feature scaling where required

- Handling class imbalance by focusing on F1-score and ROC-AUC metrics

3. Exploratory Data Analysis (EDA)

EDA revealed that customers on month-to-month contracts and those with short tenure are more likely to churn. Higher monthly charges also correlated with higher churn rates, while long-term contracts reduced churn probability.

4. Models Implemented

The following models were trained and evaluated:

- **Logistic Regression**: Used as a baseline with balanced and interpretable results.

- **Decision Tree Classifier**: Captured non-linear relationships but showed overfitting tendencies.

- **Random Forest Classifier**: Ensemble model that reduced overfitting and delivered the best overall performance.

5. Model Evaluation and Comparison

Evaluation metrics included **Accuracy, Precision, Recall, F1-score, and ROC-AUC.**

Random Forest consistently outperformed other models across most metrics, while Logistic Regression served as a strong baseline.

6. Cross-Validation

5-fold cross-validation was applied to ensure model stability and generalization. Random Forest maintained consistent performance across folds.

7. Final Results and Insights

Random Forest was selected as the final model. Key churn predictors included contract type, tenure, and monthly charges. Ensemble methods proved more effective than distance-based models.

8. Conclusion

This project demonstrates a complete machine learning workflow, from preprocessing and EDA to model evaluation and selection. The final model can be used to help businesses proactively identify customers at risk of churn.