Short Project Report: Telecom Churn Prediction

1. Introduction

Customer churn means when telecom customers leave causes revenue loss. Predictive modelling identifies likely churners to retain them. This project applies machine learning (ML) models to predict churn and evaluates them using standard metrics.

2. Objectives

- **Predict customer churn** – *SDG 9: Innovation*.
- **Evaluate models (Accuracy, F1, ROC-AUC)** – *SDG 8: Efficiency*.
- **Visualize & deploy inference pipeline** – *SDG 12: Reduce waste*.
- **Select the best ML model** – *SDG 9 & 8: Innovation & efficiency*.

3. Dataset

- **Instances:** 7032 customers
- **Features:** 20+ (tenure, charges, contract, payment method, etc.)
- **Target:** Churn (0 = stay, 1 = churn)
- **Preprocessing:** Missing values handled, categorical features encoded, feature vector created for inference.

4. Methodology

- **Models:** Logistic Regression, Decision Tree, Random Forest, KNN, SVM
- **Metrics:** Accuracy, Precision, Recall, F1-score, Confusion Matrix, ROC-AUC
- **Cross-validation:** 5-fold for stability
- **Model rationale:**
    - Random Forest & Decision Tree: capture non-linear patterns
    - Logistic Regression: baseline linear model
    - KNN & SVM: alternative classifiers for comparison

5. Key Results

Model	Accuracy	F1 Score
Logistic Regression	0.78	0.78
Decision Tree	0.80	0.80
Random Forest	0.83	0.82
KNN	0.75	0.74
SVM	0.68	0.70

- **Best model:** Random Forest
- Confusion matrices & ROC curves visualized
- Inference on new customer: predicted to stay (probability: 0.57)

6. Literature Review (2024–2025)

1. **Alotaibi & Haq (2024):** Ensemble methods (RF, XGBoost, LightGBM) improve predictive performance.
2. **Chang et al. (2024):** Random Forest + XAI (LIME/SHAP) enhance interpretability and metrics.
3. **Hybrid Framework (2024):** XGBoost + SMOTE-ENN handles class imbalance, improves ROC/F1.
4. **Wakhidah et al. (2025):** SMOTE + RF/XGBoost increases F1 under imbalance.
5. **Ozkurt (2025):** SHAP vs. Partial Dependence improves feature importance interpretation.
6. **Nature Scientific Reports (2025):** RF identifies key churn predictors, strong accuracy/precision.
7. **Chen et al. (2025):** Deep neural networks improve churn prediction accuracy.
8. **Shaikhsurab & Magadum (2024):** Stacked ensembles achieve 99%+ accuracy.
9. **Fuzzy Churn Patterns (2024):** Fuzzy logic improves interpretability of churn predictions.
10. **Faraz et al. (2024):** Classical ML approaches highlight feature extraction and imbalance handling.

7. Challenges & Future Work

- **Challenges:** Imbalanced datasets, SVM underperformance, feature importance interpretation
- **Future:** Deep learning, survival analysis for time-based churn, real-time CRM deployment

8. Conclusion

- Random Forest is the most effective churn predictor.
- Visualization & inference pipeline allow practical deployment.
- Results align with recent research emphasizing ensembles, evaluation metrics, and explainability.