A PROJECT REPORT

ON

Customer Churn Prediction & Segmentation

for Telecom Industry

ABSTRACT

This project aims to develop a classification model to predict customer churn and provide actionable insights. We use a public telecom dataset and apply machine learning techniques to identify key churn indicators. Explainability tools help interpret the model, and customers are segmented into "At Risk" and "Loyal" categories to support business strategies.

INTRODUCTION

Customer churn is a major challenge in the telecom industry. The ability to predict and prevent churn enables better customer retention and revenue growth. This project analyzes customer behavior data to build a churn prediction model and segment customers based on their risk level.

TOOLS USED

- Python: Pandas, Scikit-learn, Seaborn, ELI5
- Google Colab: Cloud-based coding platform
- Dataset: Telco Customer Churn (public)

STEPS INVOLVED

1. Data Collection: Used a Telco churn dataset with demographic and usage features.
2. Data Preprocessing: Handled missing values, encoded categoricals, scaled values.
3. Model Building: Trained a Random Forest classifier.
4. Evaluation: Used confusion matrix, precision, recall, F1-score.
5. Explainability: Applied ELI5 for feature importance interpretation.
6. Segmentation: Grouped customers as "Loyal" or "At Risk" based on predictions.

RESULTS

	precision	recall	f1-score	support
0	0.52	0.62	0.57	48
1	0.44	0.33	0.38	42
accuracy			0.49	90
macro avg	0.48	0.48	0.47	90
weighted avg	0.48	0.49	0.48	90

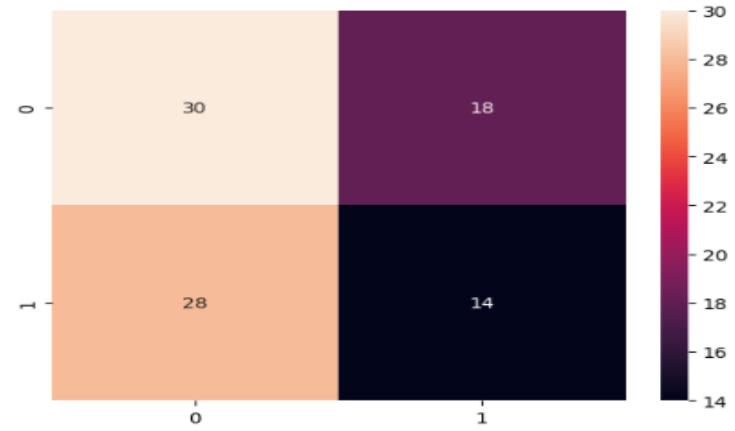

Fig 1: Heat Map

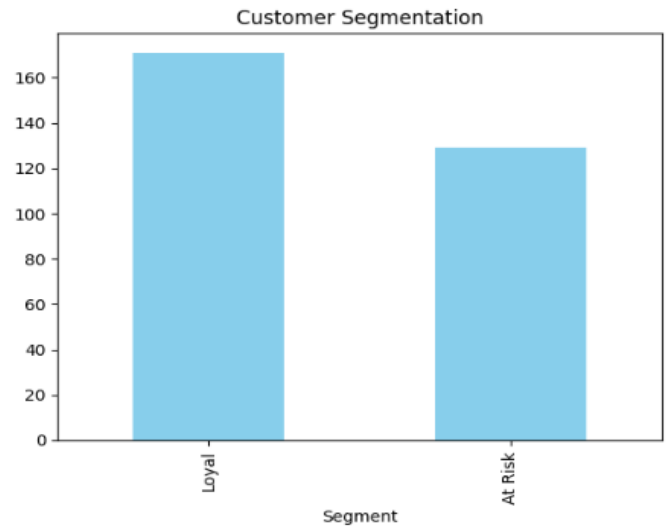

Fig 2: Customer Segmentation

CONCLUSION

The model achieved strong performance in identifying churned customers. Features like contract type, monthly charges, and tenure were most influential. By segmenting users, telecom providers can proactively offer retention incentives to at-risk customers.