Telecom Churn Prediction Project Report

# 1. Project Overview

This project analyzes customer behavior data from a telecom company to identify the factors influencing churn, predict churn risk scores, and classify customers using a CHURN\_FLAG. The ultimate goal is to support targeted marketing campaigns and improve customer retention.

# 2. Data Source and Setup

The data is stored in a MySQL database and accessed using Python:  
- Host: 18.136.157.135  
- Database: project\_telecom  
- Table: telecom\_churn\_data  
- User: dm\_team3  
Connection and retrieval were done using mysql-connector-python.

# 3. Data Preparation

The original dataset contained unnamed or default column names. These were renamed for better understanding:  
State, Account Length, Area Code, Phone, International Plan, VMail Plan, VMail Message, Day Mins, Day Calls, Day Charge, Eve Mins, Eve Calls, Eve Charge, Night Mins, Night Calls, Night Charge, International Mins, International Calls, International Charge, CustServ Calls, Churn.

# 4. Data Cleaning & Encoding

String columns such as 'International Plan', 'VMail Plan', and 'Churn' were cleaned by removing whitespace, converting to lowercase, and removing hidden characters. These were then mapped to binary values:  
- Yes → 1  
- No → 0

# 5. Exploratory Data Analysis (EDA)

EDA was performed to understand relationships between variables and churn. This included:  
- Correlation heatmap  
- Churn distribution  
- Feature impact visualizations (e.g. Day Mins vs Churn)

# 6. Model Building

A Random Forest Classifier was trained to predict churn using customer features. The dataset was split into training and testing sets. The model performance was evaluated using a classification report.

# 7. Risk Scores and CHURN\_FLAG

The trained model was used to predict churn risk probabilities. A new column 'Churn\_Risk\_Score' was created to store these probabilities. Another binary column 'CHURN\_FLAG' was derived using a threshold of 0.5:  
- CHURN\_FLAG = 1 if risk > 0.5 (Likely to churn)  
- CHURN\_FLAG = 0 otherwise

# 8. Exporting Results

The final dataset, including churn scores and flags, was exported as a CSV file for use in email retention campaigns.