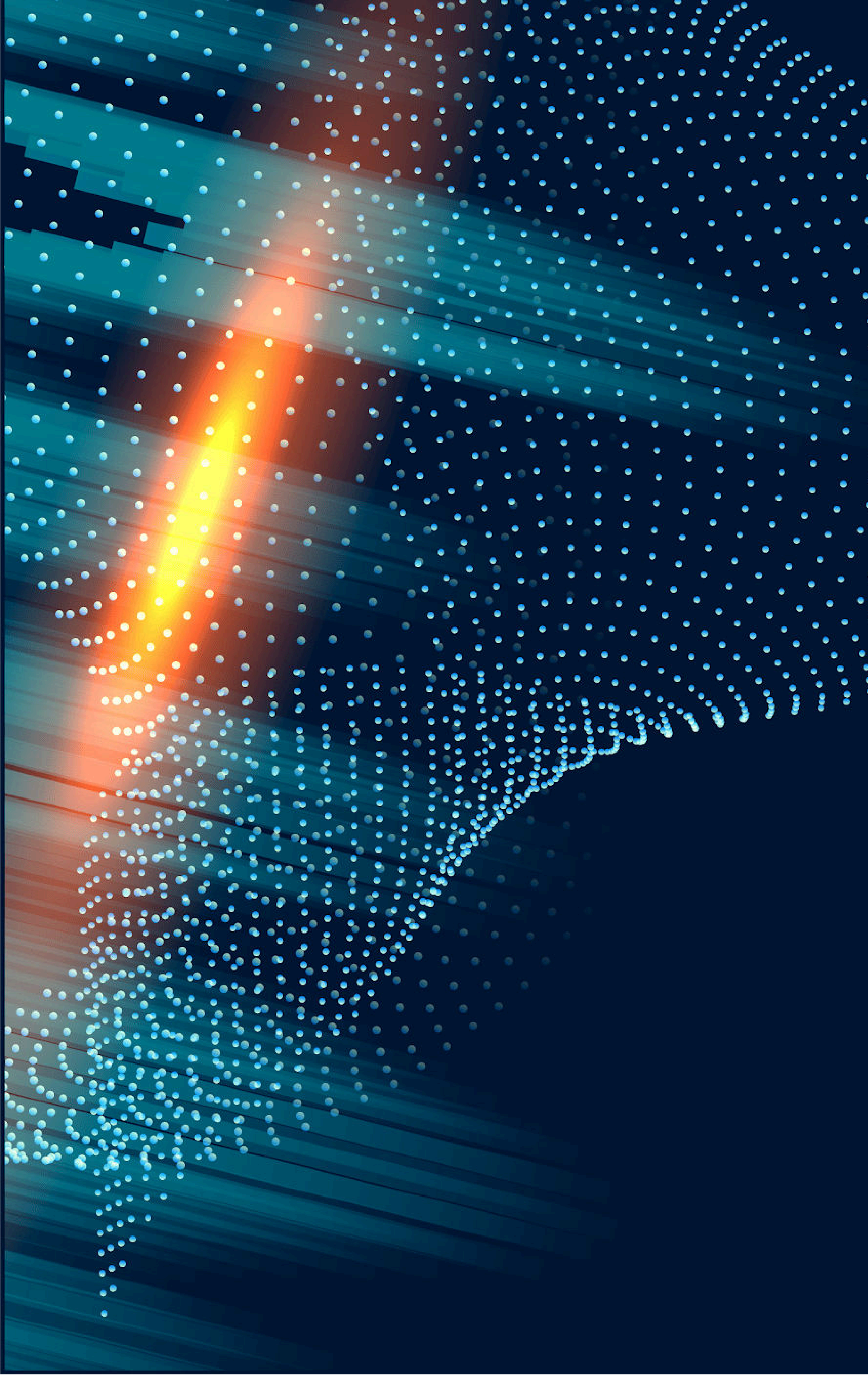
Title:-Telecom Churn Project



Introduction to Telecom Churn

The telecom industry plays a pivotal role in connecting people and businesses globally, fostering communication, innovation, and economic growth. Customer retention is critical in this competitive sector as acquiring new customers is often more expensive than retaining existing ones. Churn, defined as the rate at which customers discontinue their subscriptions, directly impacts revenue, leading to financial losses and reduced market share. High churn rates can undermine profitability, making it vital for telecom companies to implement strategies like personalized services, competitive pricing, and proactive customer engagement to ensure loyalty and sustain growth in this dynamic industry.

Objective

The primary goal of this project is to analyze customer churn to identify factors leading to attrition and develop strategies to improve customer retention.

Problem Statement

Churn analysis is critical as retaining existing customers is more cost-effective than acquiring new ones. High churn rates can harm revenue, brand reputation, and long-term growth, making it essential to understand and address underlying causes effectively.

Data Preprocessing

To ensure reliable analysis and modeling, the data preprocessing phase addresses several critical tasks:

1. Data Cleaning
2. Feature Engineering
3. Normalization and Scaling

Visualization

- Churn Rate Visualization: A bar plot showed churners as a smaller yet significant group among high-value customers.
- Usage Trends: Line plots highlighted sharp declines in usage for churners.
- Heatmap: A correlation heatmap identified key predictors of churn, such as reduced recharge activity.

Criteria for Selecting Features

Feature selection focuses on identifying variables that significantly influence customer churn. Criteria include:

- Predictive Power: Variables that strongly correlate with the target variable (churn).
- Data Completeness: Features with minimal missing or invalid values.
- Relevance: Columns relevant to customer usage, such as recharge amounts, call duration, and data usage.

The final model uses variables likeaverage recharge amount (two months),data usage, andcall minutes. These were identified as most predictive, enabling a focused and accurate churn prediction model.

Model Development

- The model development phase focuses on building and evaluating a machine learning pipeline to predict telecom customer churn. Initially, the dataset undergoes preprocessing, including handling missing values, scaling features, and splitting the data into training and testing sets. Customers are tagged as churners based on specific inactivity conditions, forming the target variable.

- The model is trained using a balanced class weighting strategy to address the class imbalance in churn labels. Performance is evaluated through metrics like confusion matrix, precision, recall, F1-score, and ROC-AUC score. Feature importance analysis is performed to identify the most influential predictors of churn, aiding interpretability.

Model Evaluation

Key metrics includeaccuracy,precision,recall,F1-score, andROC-AUC. Accuracy provides the proportion of correctly predicted instances, while precision and recall highlight the model's ability to correctly identify churners without excessive false positives or negatives. The F1-score combines precision and recall, offering a balanced evaluation metric. ROC-AUC measures the model's capability to distinguish between churners and non-churners.

Confusion Matrix Analysis

The confusion matrix highlights the model's strengths in correctly classifying churners (true positives) while minimizing false negatives. Insights from the confusion matrix aid in refining the model to focus on reducing misclassifications of high-value churners.

Important Indicators to the Target Variable

- Features such as average recharge amount during the good phase,total outgoing minutes of usage (OG MOU), anddata usage (2G/3G)were found to have a significant impact on churn prediction.
- The feature importance analysis highlights that high-value customers who exhibit a decline in these activities are more likely to churn, indicating that consistent engagement and recharge patterns are crucial for retention.

Based on the analysis, the following strategies are proposed to minimize churn:

- Targeted Retention Campaigns: Focus on customers showing declining recharge or usage trends by offering customized plans or incentives to encourage continued usage.
- Data Usage Optimization: Identify customers with reduced data activity and provide tailored data packs or better network service support.

Outcomes:

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

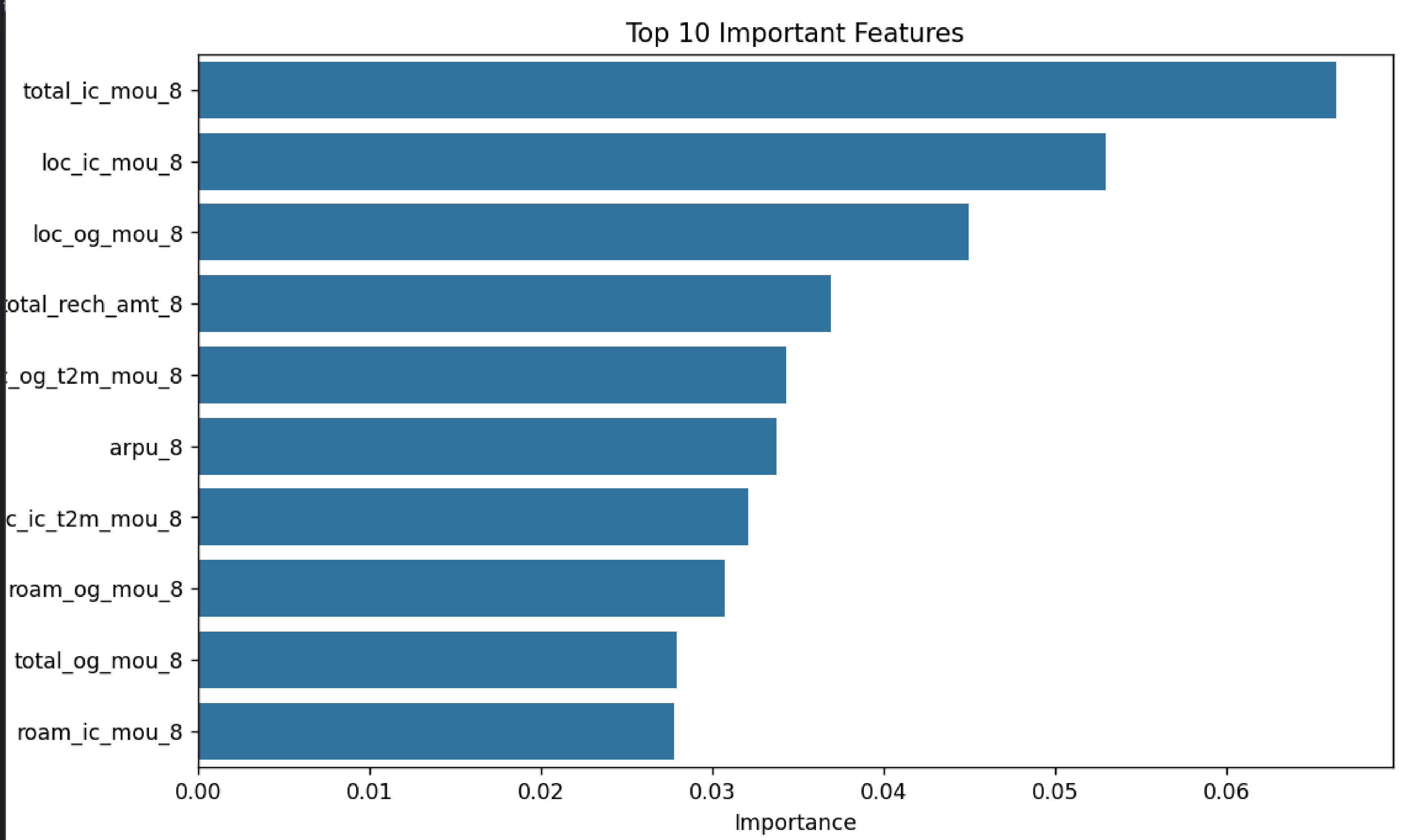
```
PS C:\Users\archa\Python> python -u "c:\Users\archa\Python\main.py"
Confusion Matrix:
[[8121  105]
 [ 428  350]]
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.99      0.97      8226
           1       0.77      0.45      0.57       778

    accuracy                           0.94      9004
   macro avg       0.86      0.72      0.77      9004
weighted avg       0.93      0.94      0.93      9004

ROC-AUC Score: 0.9261188738197339
```

Figure 1



Conclusion

- The Telecom Churn Project successfully demonstrates how machine learning models can be leveraged to identify high-value customers and predict churn with actionable insights. By analyzing key customer behaviors, such as recharge amounts and usage patterns, the project highlights critical factors contributing to customer retention and churn.

- This project emphasizes the importance of data preprocessing, feature engineering, and interpretability in building practical solutions.

- The visualizations, such as the feature importance bar plot, provide an intuitive understanding of the model's decisions, aiding business stakeholders in implementing targeted retention efforts.