Customer Churn Prediction Project

https://github.com/Berothely/phase3project.git

Machine Learning Models for Telecom Industry

Project Introduction

- This project addresses customer churn prediction for a telecom company. Churn, the phenomenon of customers discontinuing a service, directly impacts long-term profitability. The goal is to leverage customer behavioral and service data to predict churn and provide actionable insights.

Objectives

1

- Build models to predict whether a customer will churn

2

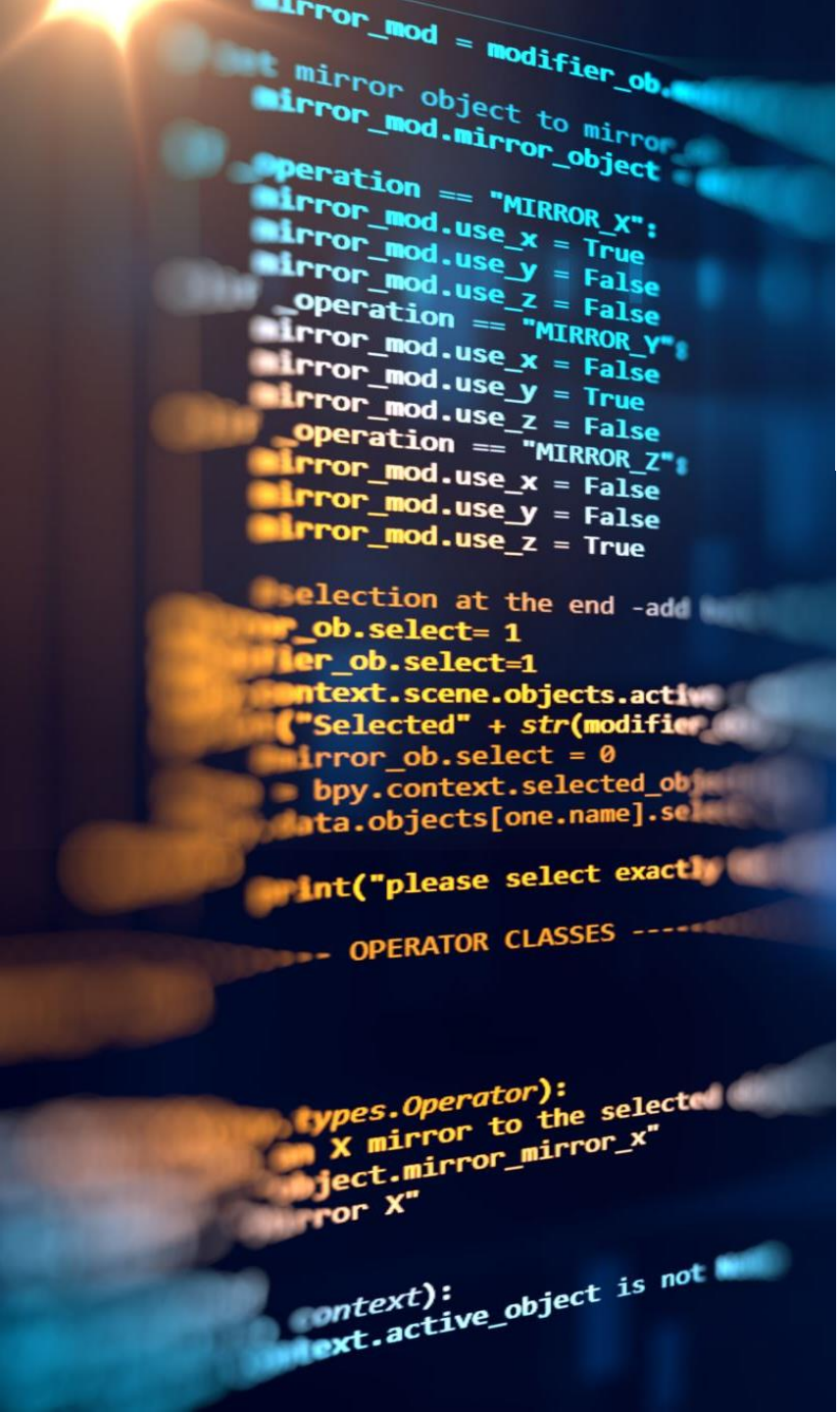
- Compare Logistic Regression and Random Forest performance

3

- Provide insights for retention strategies

4

- Ensure interpretability and business applicability

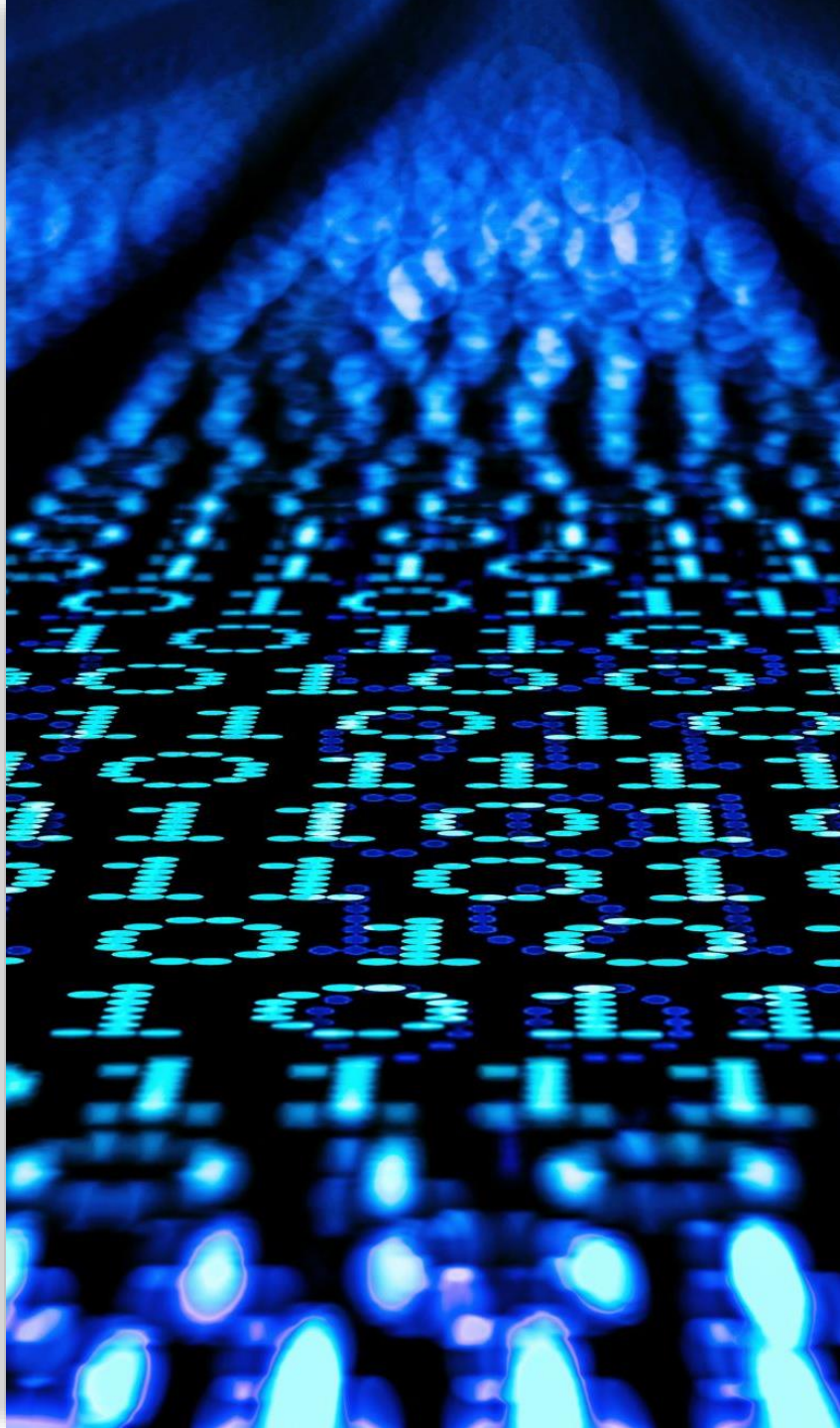


Data Understanding

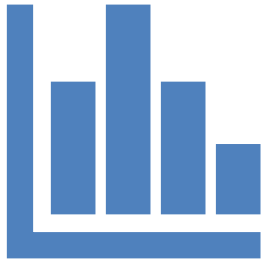
- Dataset includes customer demographics, account information, and service usage.
- - Total records: Telecom customer data
- - Features: Categorical and numerical variables
- - Target: Churn (Yes/No)

Data Preparation

- - Handled missing values and inconsistencies
- - Encoded categorical variables
- - Scaled numerical features
- - Addressed class imbalance



Models Applied



1. Logistic Regression: Interpretable baseline model

2. Random Forest: Ensemble method with higher predictive power

Model Training & Evaluation

- Hyperparameter tuning with RandomizedSearchCV

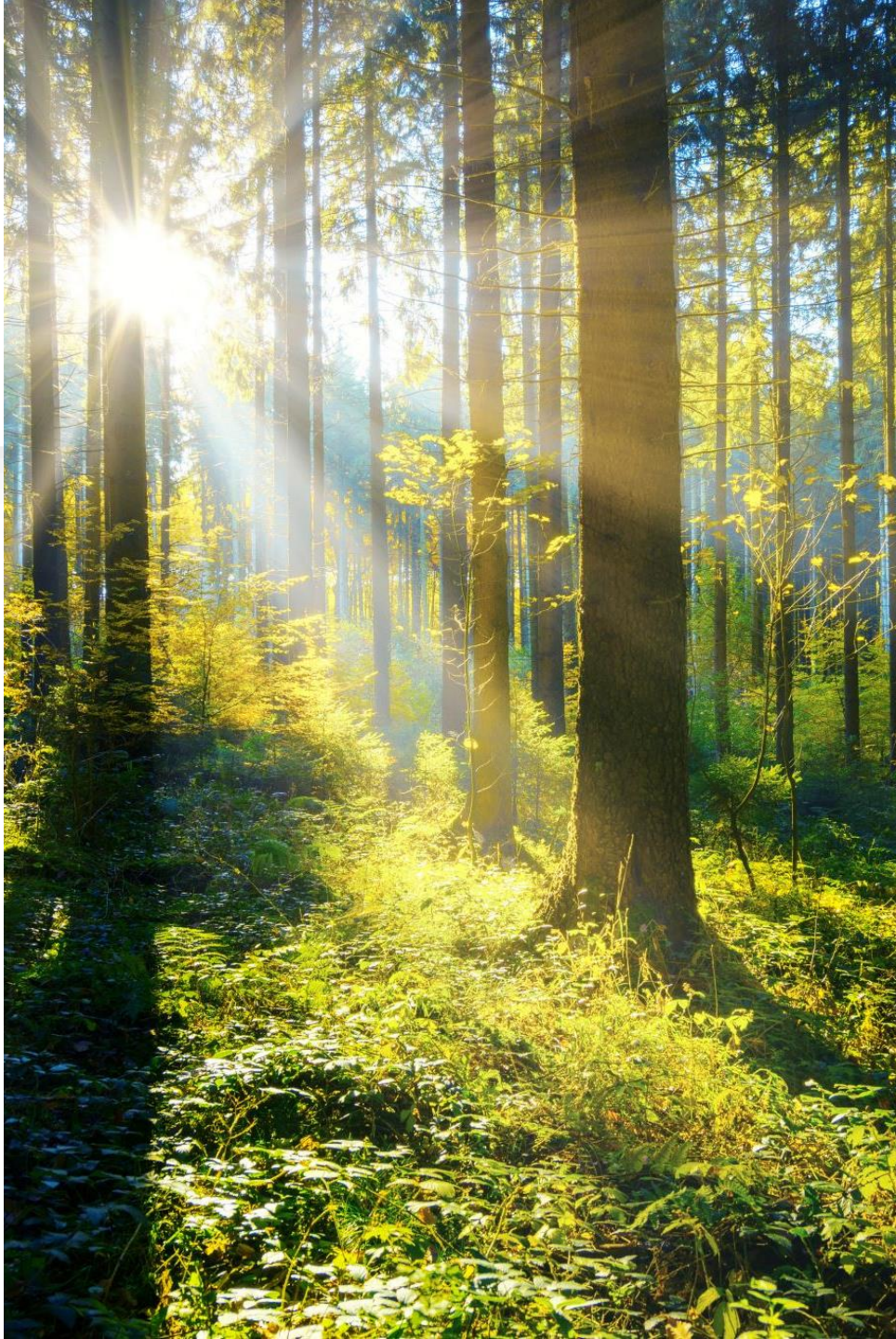
- Evaluated using Accuracy, Precision, Recall, F1-score, ROC-AUC

- Confusion matrix and classification report generated

Results & Comparison

- Logistic Regression:
- - Provided interpretability via coefficients

- Random Forest:
- - Higher performance
- - Feature importance insights
- Both models compared using standard metrics.



Business Insights

- Key drivers of churn identified

- Enables proactive retention strategies

- Practical recommendations for telecom management

Business Recommendations

Customer Retention Programs

Design loyalty programs, personalized offers, and discounts targeted at customers with high churn risk.

Offer bundle packages (internet + phone + TV) to increase switching costs and retain customers longer.

Improve Customer Service

Enhance call center responsiveness and problem resolution speed.

Provide proactive customer support for those showing signs of dissatisfaction.

Data-Driven Marketing

Use churn predictions to segment customers and focus marketing efforts on at-risk groups.

Implement personalized communication (emails, SMS, calls) to keep at-risk customers engaged.

Product & Service Optimization

Regularly review service quality (network reliability, pricing fairness).

Offer flexible contracts (month-to-month, upgrade/downgrade options) to reduce customer frustration.

Customer Feedback Integration

Collect and analyze feedback from churners to identify service gaps.

Adapt retention strategies continuously based on customer insights.

Recommendations

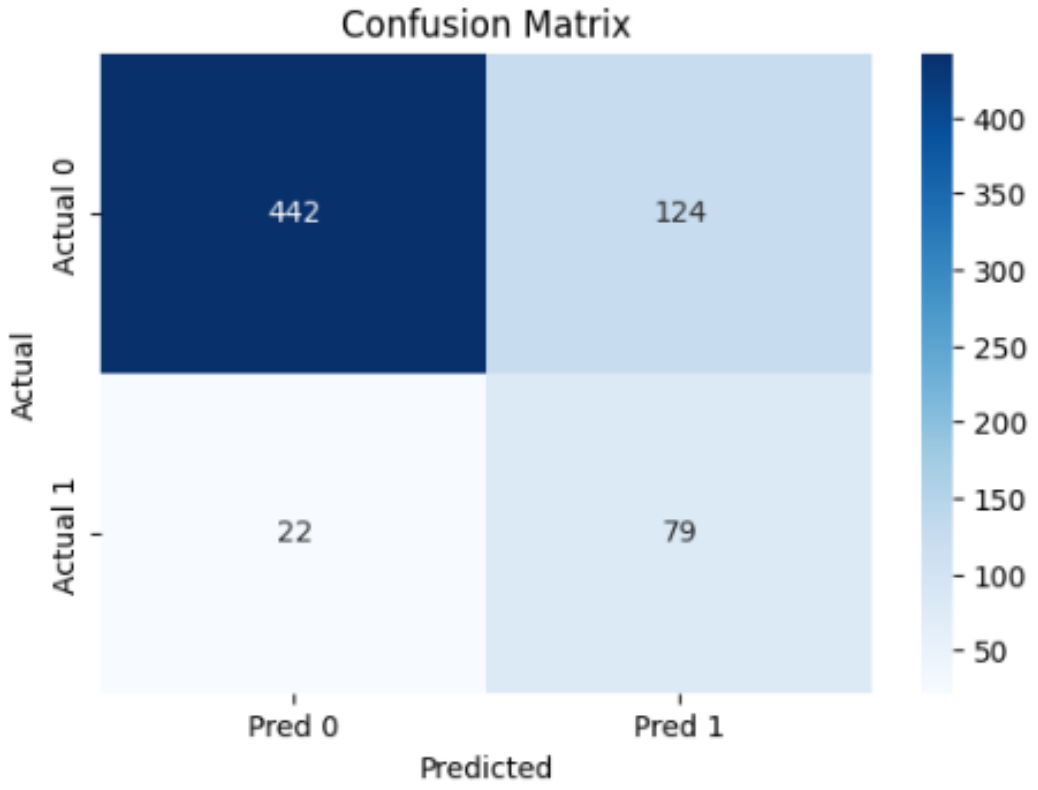
- Use the tuned **Logistic Regression with SMOTE pipeline** as a baseline predictive model for churn detection.
- Complement the analysis with more complex models (e.g., Random Forest, Gradient Boosting, XGBoost) to potentially improve predictive power.
- Continuously monitor model performance and **retrain periodically** as customer behaviors evolve.
- Beyond modeling, telecom management should:
  - Develop **targeted retention strategies** for at-risk customers flagged by the model.
  - Offer incentives (discounts, loyalty programs) to customers with high churn probability.
  - Use churn predictions to **prioritize customer support efforts** where they have the highest impact.

Graph1 *confusion matrix of the linear regression model*

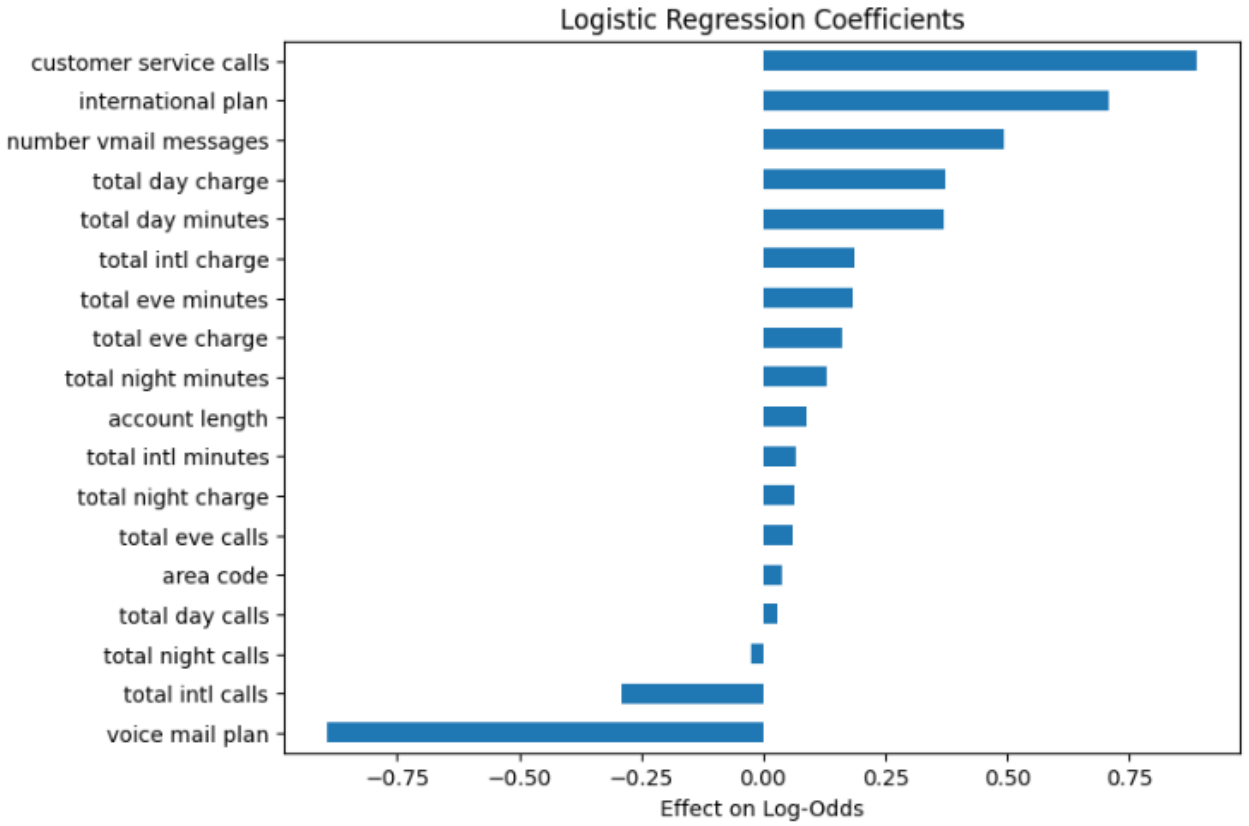
```
accuracy: 0.7811

roc_auc score: 0.8345

conf_matrix:
```



Graph2


Logistic Regression Coefficients
customer service calls
international plan
number vmail messages
total day charge
total day minutes
total intl charge
total eve minutes
total eve charge
total night minutes
account length
total intl minutes
total night charge
total eve calls
area code
total day calls
total night calls
total intl calls
voice mail plan
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
Effect on Log-Odds

Graph3
class_report of random forest

.9]:

	precision	recall	f1-score	support
0	0.973262	0.964664	0.968944	566.000000
1	0.811321	0.851485	0.830918	101.000000
accuracy	0.947526	0.947526	0.947526	0.947526
macro avg	0.892291	0.908075	0.899931	667.000000
weighted avg	0.948740	0.947526	0.948044	667.000000

Conclusion

- This project demonstrates how machine learning can effectively predict churn.
- Random Forest outperformed Logistic Regression in predictive accuracy.
- The insights generated can help reduce customer churn and improve long-term profitability.