

PREDICTING CUSTOMER CHURN

Regression-Based and Tree-Based Methods

Group: Abad, Monfero, Yang | Course: DSC1107

PROBLEM OVERVIEW

Objective: Predict customer churn using Orange Telecom dataset.



```
graph TD; A[Objective: Predict customer churn using Orange Telecom dataset.] --> B[Why it matters:]; B --> C["- Customer retention is more cost-effective than acquisition."]; C --> D["- Enables proactive service and marketing actions."];
```

The diagram is a vertical flowchart with four rectangular boxes connected by downward-pointing arrows. The boxes are colored in a gradient from light green at the top to light blue at the bottom. The first box contains the objective, the second box asks 'Why it matters:', the third box lists a point about customer retention, and the fourth box lists a point about proactive service and marketing actions.

Why it matters:

- Customer retention is more cost-effective than acquisition.

- Enables proactive service and marketing actions.

DATA PREPARATION & EDA



Dropped identifiers (e.g., phone number, state).



Converted 'Churn' to binary; applied one-hot encoding.



Class imbalance detected: majority are non-churners.



Service call frequency and international plans linked to churn.



Correlation analysis informed feature selection.

MODEL COMPARISON (SUMMARY)

- Gradient Boosting: Best F1 (0.802), ROC AUC (0.925).
- Random Forest: Highest precision (0.969), strong overall.
- Decision Tree: Interpretable with good balance.
- Logistic/Lasso/Ridge: High recall but low precision.
- Interpretability vs. predictive strength trade-off.

TOP FEATURES (RANDOM FOREST)

- 1. Customer service calls (0.1444)
- 2. Total day charge (0.1255)
- 3. Total day minutes (0.1245)
- 4. International plan (Yes) (0.0928)
- Behavioral patterns like usage & service dissatisfaction drive churn.

BUSINESS STRATEGY & IMPACT



Identify high-risk churners using usage behavior.



Focus retention campaigns on target groups (e.g., intl. plan users).



Use interpretability for internal reporting (e.g., Logistic Regression).



Deploy Gradient Boosting for accurate churn detection.

CONCLUSION



GRADIENT BOOSTING
= TOP-PERFORMING
MODEL.



LOGISTIC REGRESSION
= BEST FOR
EXPLAINABILITY.



RECOMMENDATION:
USE BOTH IN TANDEM
FOR INSIGHT +
ACTION.



CHURN MODELING
SUPPORTS PROACTIVE
CUSTOMER RETENTION
STRATEGIES.