Browsing or Buying?

Predicting Online Purchasing Behavior

Varsha Garla

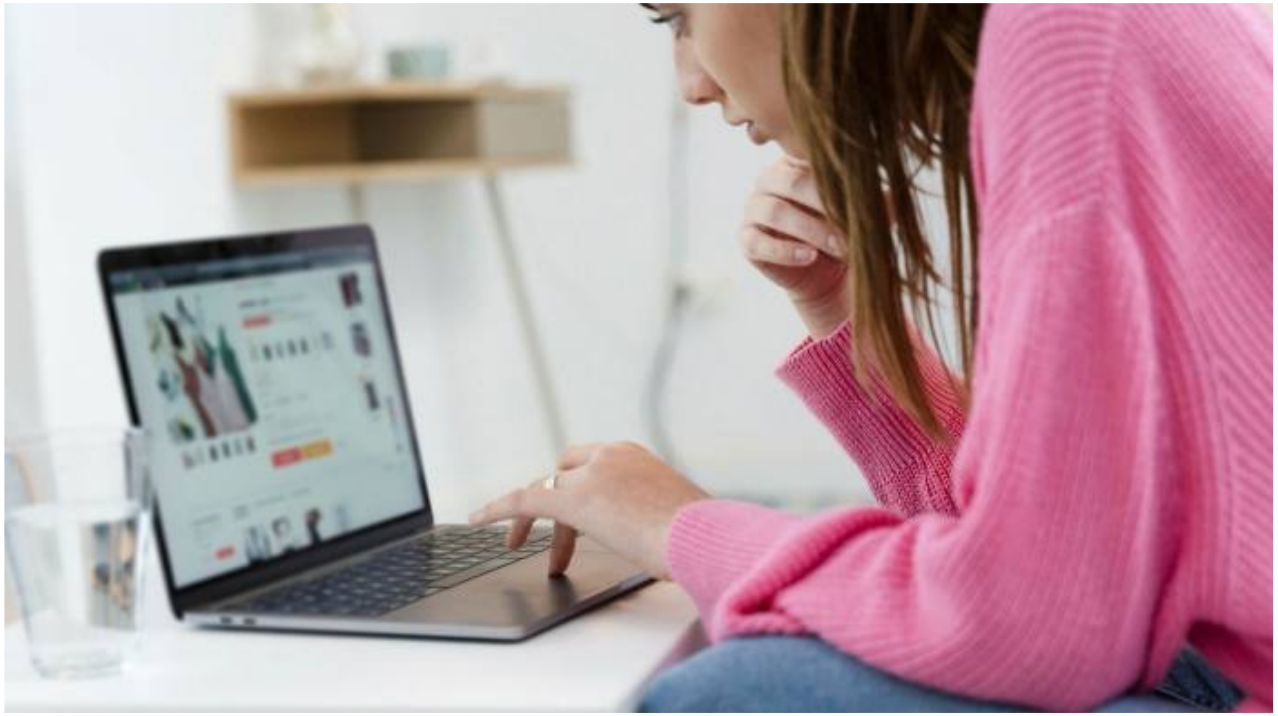
Metis | September 2021 | Classification

Problem

E-COMMERCE CLIENT

85%

Visits to website DON’T end in transaction

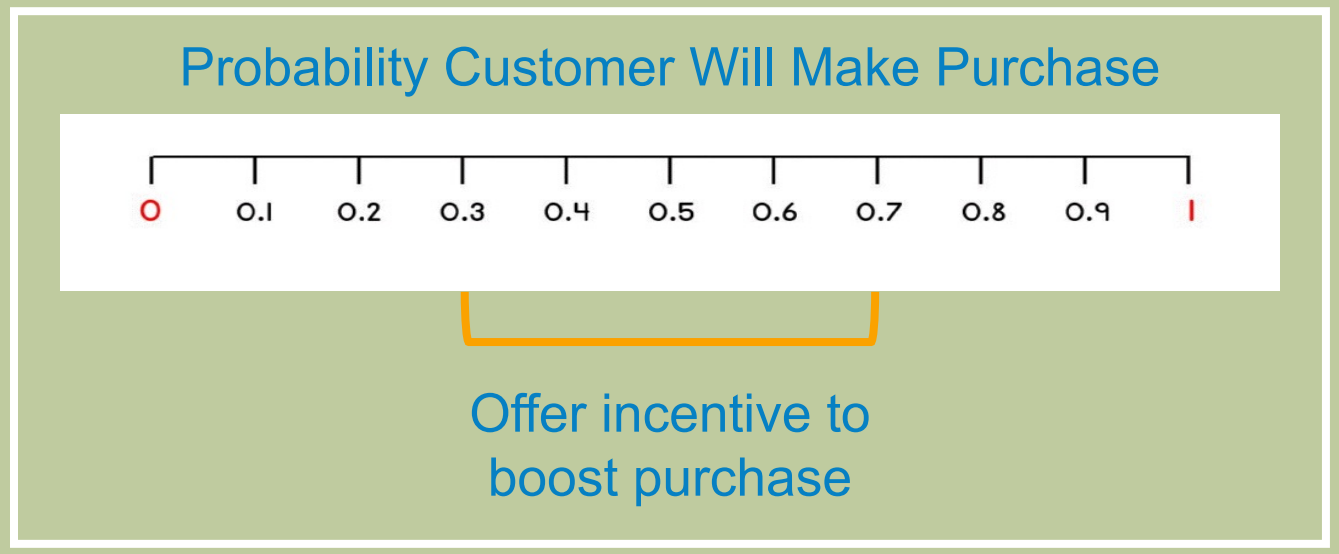


Objective

1. Increase purchase conversion rate
2. Gain insights to guide promotional strategy

Solution

- Build classification model → soft predictions
- Offer recommendations based on feature importance



Methodology

Data

Online Shoppers Purchasing Intention Dataset
(Sakar, C.O., Polat, S.O., Katircioglu, M. et al.)

- Clickstream and user info
- 12,330 observations

Cleaning/ EDA

pandas

NumPy

Baseline Models

scikits learn

Logistic Regression, KNN, Decision Tree, Random Forest

XGBoost

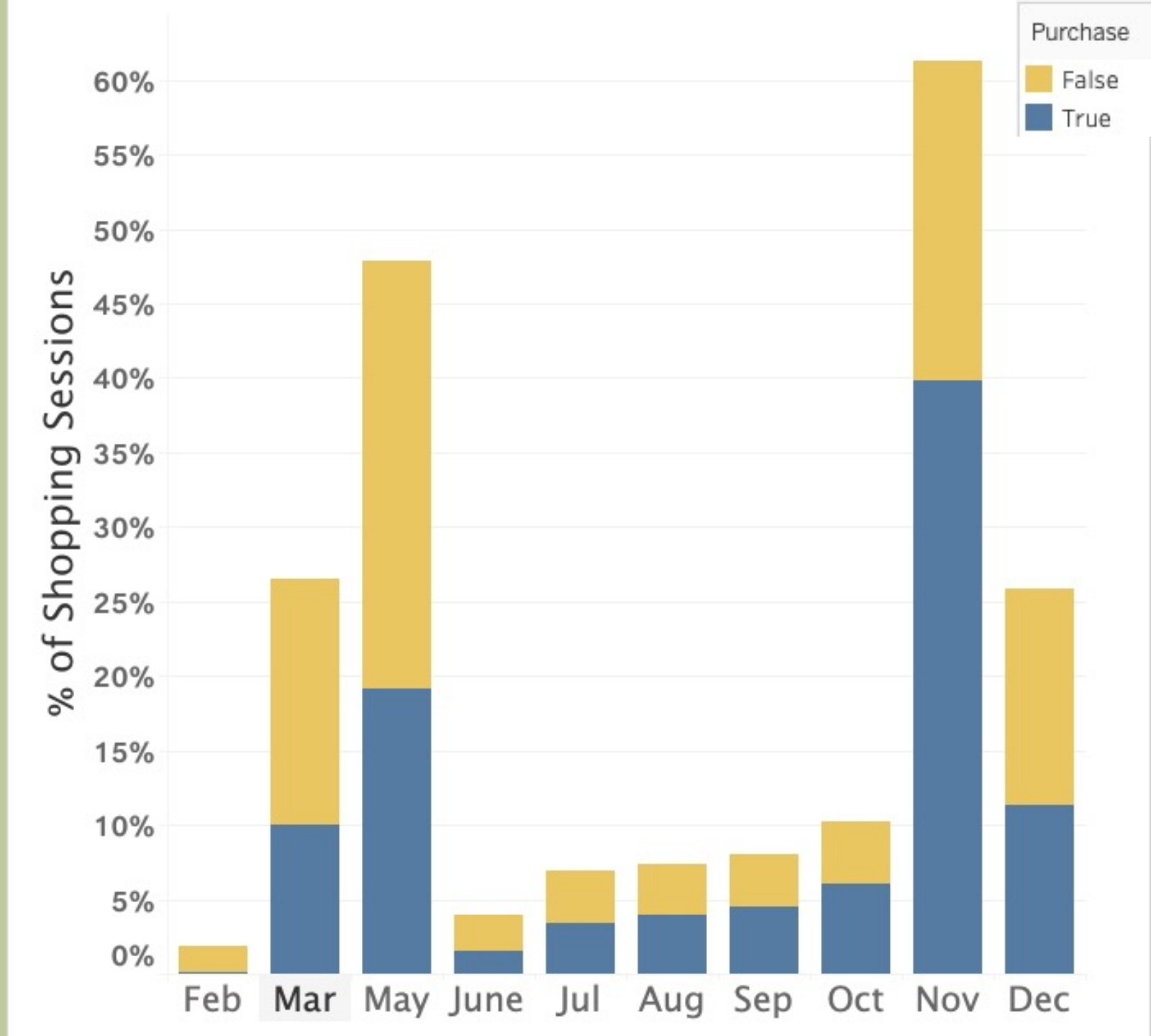
Model Selection/ Tuning

Logistic Regression
- Soft predictions
- Interpretable
- Scalable

- Log-Loss
- ROC AUC

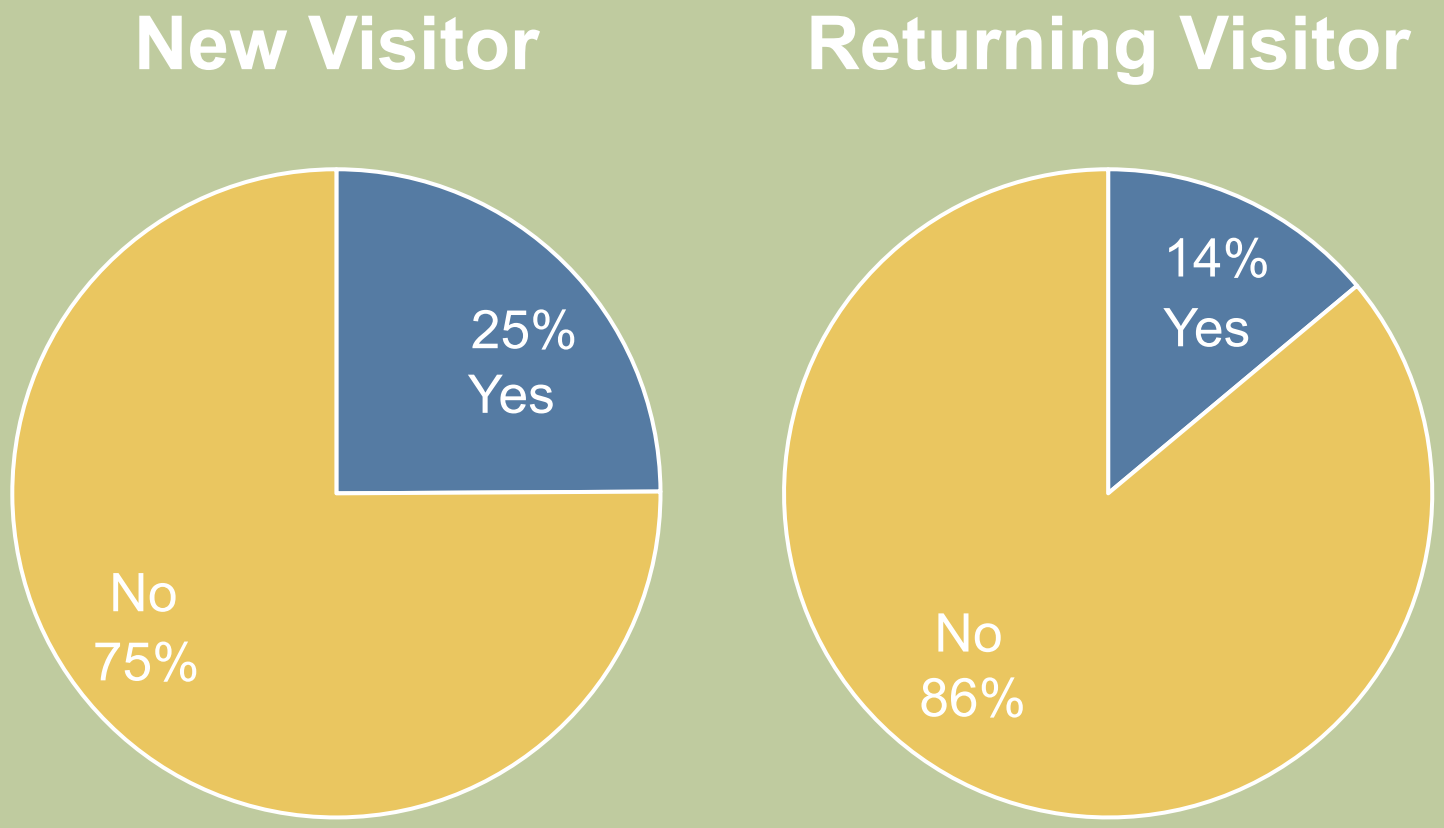
Insights

- High-traffic months
- Customers shopping in **November** are more likely to purchase



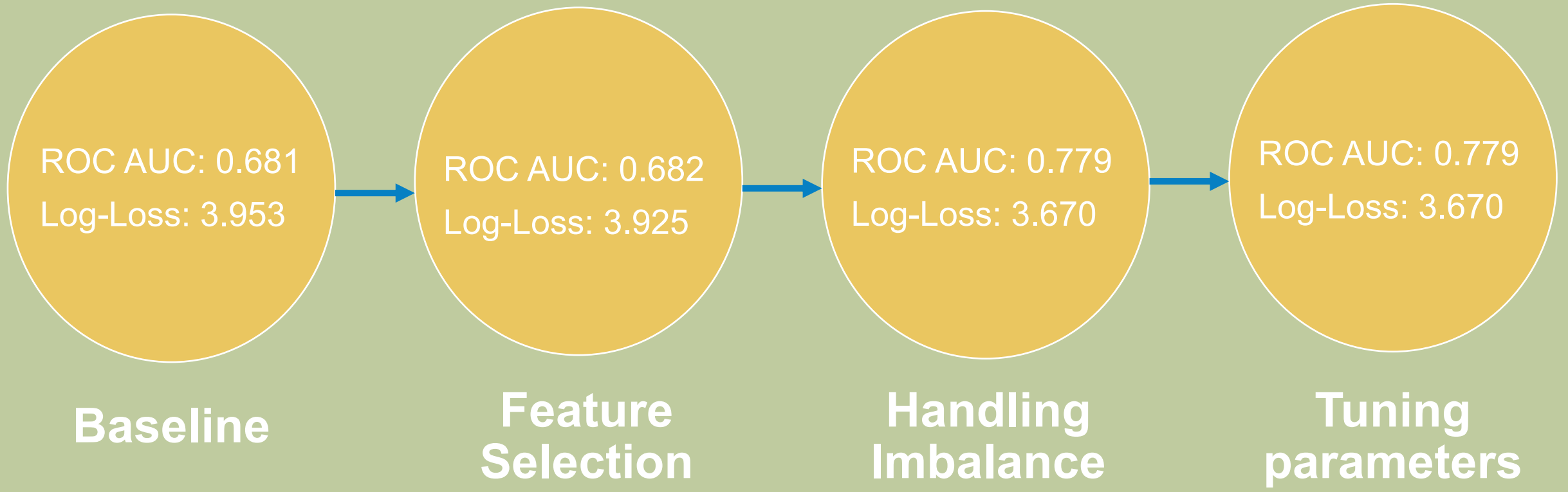
Insights

New visitors are more likely to purchase than returning visitors



Results

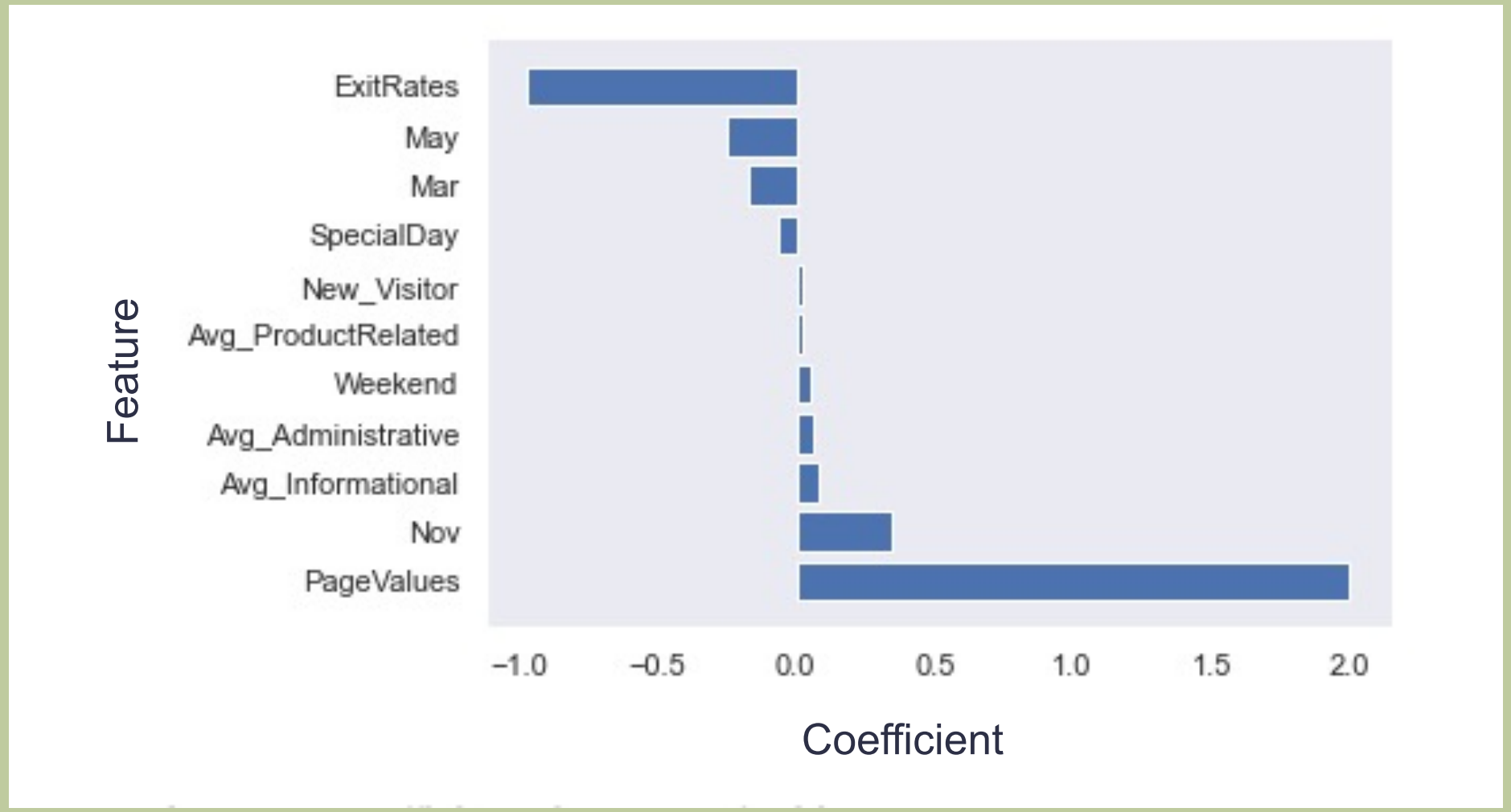
CV Scores



Test Scores

ROC AUC: 0.865
Log-Loss: 3.896

Results: Feature Importance



Recommendations

Feature	Insight	Suggestion
ExitRates	If a customer visits pages with high exit rates, they are less likely to purchase	Remove or improve the pages that have high exit rates
May, March	Customers who visit in these high-traffic months are less likely to purchase	Push promotions
November	Customers who visit the site in November are more likely to purchase	Capitalize on this
PageValues	If a customer visits pages that have high page value, they are more likely to purchase	Remove or improve pages that don’t have high page values

Future Work

- Use information about user history
- Use A/B Testing to see if assumption about coupon influence is valid
- Implement in real-time

Thank you!
Questions?

Appendix

Final Model

- Algorithm: Logistic Regression
- Number of Features: 11
- Parameters:
  - Class Weights: 3:1
  - C: 10
  - Penalty: L1
  - Solver: liblinear

Distribution of Predicted Probabilities