

Predicting Customer Purchase Likelihood Using Machine Learning

-Anujkumar Yadav MIP-ML-18

Objective and Problem Statement

- **Objective:** To build a predictive model that estimates whether a customer will make a purchase based on demographic and behavioral data.
- **Problem Statement:** Using historical customer data, we want to predict purchase likelihood to support business decision-making and improve customer targeting.
- **Why This Matters:** Helps focus marketing efforts on high-potential customers, leading to increased conversions and revenue.

Dataset overview

Column Name	Description
Age	Age of customer
Gender	Male(0) or Female(1)
Annual Income	customers yearly income
product category	type of product
time spent on website	total min spend
loyalty program	membership status
Discount availed	discount used
purchase status	(1) yes, (0) no

Model Selection and Comparison

Evaluated three different models: Logistic Regression, Random Forest, and Gradient Boosting.

Model	Cross-Validation Accuracy
Logistic Regression	0.8167
Random Forest	0.9133
Gradient boosting	0.9167

Conclusion: Gradient Boosting performed the best, with the highest accuracy across validation folds.

Hyperparameter Tuning and Best Model

- Hyperparameter Tuning:
 1. Objective: To fine-tune the best model (Gradient Boosting) for optimal performance.
 2. Best Parameters Found:
 - a. n_estimators: 100 (example value)
 - b. learning_rate: 0.1 (example value)
 - c. max_depth: 5 (example value)
- Benefit of Tuning:

Optimized Gradient Boosting for better accuracy and reliability on unseen data.

Model Evaluation and Example Prediction

- Classification Report on Test Set:
 1. metrics: Precision, Recall, F1-Score, Accuracy.

Example Prediction:

- "Given a sample customer profile (e.g., Age 25, Female, Income \$50,000), the model predicted: Yes, likely to purchase."
- Key Takeaway: "The model effectively distinguishes high-potential buyers, supporting targeted marketing efforts."

Classification Report on Test Set:				
	precision	recall	f1-score	support
0	0.94	0.98	0.96	172
1	0.97	0.92	0.94	128
accuracy			0.95	300
macro avg	0.96	0.95	0.95	300
weighted avg	0.95	0.95	0.95	300

Conclusion and Future Work

- Conclusion: The Gradient Boosting model accurately predicts customer purchase likelihood, allowing for improved business strategies and resource allocation.
- Business Impact: By focusing on likely buyers, the model can help increase conversion rates and customer engagement.
- Future Improvements: Consider adding more behavioral features or testing additional algorithms to refine accuracy further.

Thank You