Coupon and Offer Prediction

Machine Learning for hungry customers

Artur Magalhães R. dos Santos

August 13, 2025

Data Science Technical Case

Data Analysis

Available Data

- **17,000** unique customers
- **10** unique offers
- **306K+** transaction events

Offer Types

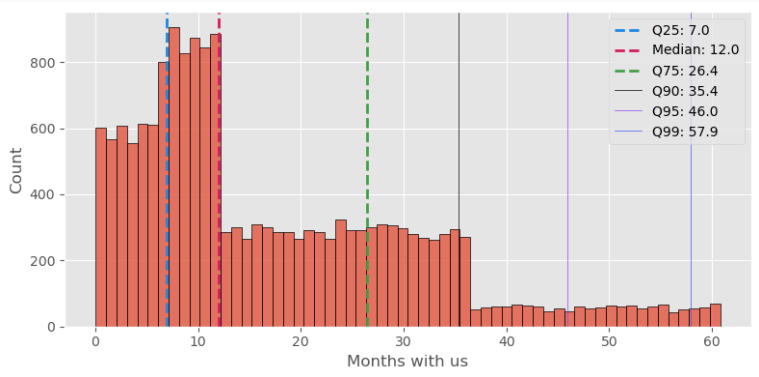
- BOGO (Buy One Get One)
- Discount
- Informational

Distribution Channels

- Email, Mobile, Social, Web

Key Insights

- Ages concentrated between 40-80 years
- ~35% male gender, ~35% female, ~30% not informed
- Diversified tenure: new to extremely experienced



Processing

Feature Engineering

- Target: Offer completion within deadline
- Historical: Past spending & transactions
- Rolling: 30-day metrics
- Segmentation: 5 tenure groups
- Channels: Distribution dummies

Dataset

- **86,432** observations
- **29** features
- Split at 80% (day 21)

Key Features

- `hist_spent`: Historical spending
- `rolling_spent_30d`: 30-day spending
- `hist_completion_rate`: Historical rate
- `tenure_segments`: Customer segments
- `offer_type`: Offer type
- `discount_value`: Discount value

Pipeline

- Categorical $\rightarrow$ 'missing'
- OneHot encoding
- No scaling needed

Modeling

Algorithm

- Random Forest
- Temporal split (training day 21)
- Complete sklearn pipeline

Performance Metrics

Metric	Class 0	Class 1
Precision	0.85	0.76
Recall	0.75	0.85
F1-Score	0.80	0.81

ROC-AUC: 89%

Accuracy: 80%

Most Important Features

1. Historical spending (`hist_spent`)
2. 30-day rolling spending
3. Discount value
4. Tenure (months with us)
5. Historical conversion rate
6. Offer type
7. Distribution channels

Model Insights

- Historical behavior is a strong predictor
- Discount offers $>$ conversion
- Experienced customers more likely

Next Steps

Model Improvements

- **Algorithms**: LightGBM, Neural Networks
- **Embeddings**: Use textual embeddings from offer description
- **Ensemble**: Model combination
- **Feature Selection**: Extend feature set
- **Hyperparameter Tuning**: Random or Bayesian Search

Feature Engineering

- Feature interactions
- Age-based features
- Temporal seasonality
- Behavior clustering

Production

- **Logging**: Add logging and experiment tracking
- **MLOps**: Automated pipeline
- **Monitoring**: Data drift detection
- **A/B Testing**: Production validation

Additional Analyses

- Class balance (training vs test) and improve splitting
- Outlier analysis (age = 118)
- Customer segmentation
- ROI by offer type

Next model target: 92%+ ROC-AUC