# Customer Promotion Response Classification Report

## 1. Introduction

The objective of this project is to predict whether a customer will respond to a promotional offer by leveraging historical transaction records and promotion metadata. This constitutes a binary classification problem, where accurate prediction of promotional response enables marketing teams to optimize campaign targeting, reduce promotional expenditure, and maximize overall return on investment.

The dataset comprises multiple components. The training dataset (train_history) contains customer and promotion identifiers, promotion date, store and region information, and the target label active, indicating whether the customer responded to the promotion. The test dataset (test_history) follows the same structure but excludes the response label. A detailed transactions dataset (transactions) provides each customer's historical purchase records, including product category, brand, quantity, expenditure amount, and transaction date. Additionally, a promotion metadata file (promos) includes attributes of each promotional offer, such as product category, manufacturer, brand, promotion quantity, and promotion value.

## 2. Methodological Approach

The primary objective is to predict customer response to retail promotions by leveraging a rich set of leakage-free features that capture historical transactional behavior, brand and category loyalty, store-level engagement, pricing sensitivity, and short- and long-term competitive dynamics. The overall approach consists of three steps: feature engineering, model selection, and feature selection.

Feature engineering is performed under strict leakage-free principles, ensuring that only customer transactions preceding each promotion are used to generate predictive variables. The pipeline captures a broad spectrum of behavioral and contextual dimensions, including temporal patterns, long-term purchase behavior, brand and category affinity, store-level engagement, recency, frequency, monetary activity, and short-term competitive pressures. High-cardinality identifiers such as customer ID, store, product category, brand, manufacturer, and promotion ID are encoded using smoothed target encoding, preserving predictive signal while mitigating overfitting. Complementary features—including ratios, interactions, log-transformed variables, and diversity metrics—contextualize individual behavior relative to store and category baselines, resulting in a rich, temporally-valid feature set suitable for promotional response prediction.

Model selection is conducted using LightGBM with hyper-parameter tuning, where a grid search explores combinations of tree structure parameters, regularization coefficients, and sampling ratios. Models are evaluated based on AUC on a validation set separated via stratified sampling to maintain the response distribution. Early stopping is applied to ensure efficient and stable convergence while preventing overfitting during tuning. Neural network is not considered here because of the simple structure of the problem, and also the already over-fitting pattern of the results from LightGBM.

Following model selection, an attempt was made to perform feature selection based on importance scores derived from the tuned LightGBM model. Variables with consistently low importance were considered for removal to reduce dimensionality, minimize noise, and improve generalization. However, removing these low-importance features did not improve validation AUC even after fine-tuning of the threshold. As a result, all features were retained in the final model, despite slightly higher risk of overfitting, with validation AUC treated as the primary performance metric.

## 3. Implementation Detail of Feature Engineering

The feature engineering pipeline produces a rich set of temporally-valid, behaviorally-informed variables designed to capture multiple dimensions of customer activity and promotional context. Temporal descriptors derived from the promotion date, including day-of-week, month, and weekend indicators, encode seasonality and weekly behavioral patterns, providing insight into how timing influences promotional response. High-cardinality identifiers such as customer ID, store, brand, category, manufacturer, and promotion ID are target-encoded using a Bayesian-style smoothing approach, preserving predictive signal while mitigating instability in low-frequency categories.

Central to the pipeline is the merging of past-only transactions, which pairs each promotion with the customer's historical purchase record strictly prior to the promotion date. This enables extraction of comprehensive RFM-style metrics: recency features quantify the days elapsed since the most recent transaction at the overall, brand, category, and manufacturer levels; frequency features capture total transaction counts, unique categories and brands purchased, and recent activity within a 30-day window; monetary features summarize cumulative and average quantities and amounts spent, as well as spending variability. Short-term recency windows identify recently active customers, while diversity metrics measure the breadth of shopping behavior across categories and brands. Store-level aggregates contextualize customer activity relative to overall store performance, including average transaction quantity, amount, and total transaction volume. Brand- and category-level aggregates further summarize historical affinity to the promoted brand or category, including transaction counts, total spend, recency, loyalty scores, and the share of brand transactions within a category.

Price sensitivity features complement these behavioral metrics by quantifying the customer's responsiveness to promotional value and quantity relative to their historical spending patterns, as well as the consistency of their transaction amounts through the coefficient of variation. To account for scaling differences and synergistic effects, ratio and interaction features relate individual behavior to store-level baselines or combine complementary signals, such as recency weighted by brand loyalty or the product of category and brand response rates. Log transformations of skewed count and monetary variables reduce variance and improve separability, enhancing model robustness. Finally, competitive context is incorporated by counting promotions in the same store and category over recent windows, capturing short-term promotional congestion that may diminish the effectiveness of the focal offer. This comprehensive, leakage-free approach ensures that all features reflect realistic, past-only customer behavior while integrating temporal, behavioral, price-sensitivity, and contextual dimensions critical for predicting promotional response.

## 4. Results and Interpretation

The predictive model was developed using a LightGBM classifier with an 80/20 train–validation split, achieving strong generalization and stable convergence. Training AUC reached 0.8751 and validation AUC 0.7134. The training–validation gap is high, suggesting overfitting from including all features compared with a reduced feature set. However, the validation AUC was treated as the primary and sole performance metric, so the overfitting gap is trade-off.

Feature importance analysis indicates that customer behavioral history is the primary driver of promotional response. Aggregate measures of past engagement, including total and average spending, cumulative quantities, and transaction counts, are most predictive, reflecting overall purchasing intensity and monetary activity. Temporal recency metrics and recent-purchase counts further highlight that recently active customers are more likely to respond. Normalized behavioral ratios relative to store-level aggregates adjust for differences in store traffic and scale, distinguishing genuinely high-intensity customers from those shopping at busier locations, while store-level aggregates such as average spend and historical response rates provide additional contextual information.

Brand-level features, including recent brand purchase recency, cumulative brand spending, and brand loyalty scores, are also highly influential, indicating that responsiveness is shaped by both general engagement and product-specific preferences. Manufacturer-level variables and minor calendar indicators, including promotion day, month, and short-window competition, contributed minimally and showed limited predictive value.

In summary, promotion responsiveness is driven primarily by historical engagement, recent activity, and brand affinity, contextualized by store performance and long-term competitive pressures. The final model integrates these behavioral, brand, store, and competition signals to achieve robust predictive performance while preserving interpretability.

## 5. Challenges and Future Directions

Large-scale transactional retail data presents several inherent challenges. High sparsity and irregular purchase histories can produce unstable RFM metrics, which were partially mitigated through smoothing and aggregation. High-cardinality categorical variables, including customer IDs, brands, and promotions, required smoothed target encoding to preserve predictive signal while controlling for overfitting. Preventing temporal leakage necessitated careful pre-promotion filtering, balanced against runtime and memory constraints. The extensive feature set also raised interpretability concerns, and a trade-off was made between maximizing validation AUC and managing overfitting risk. Collectively, these considerations highlight the necessary balance between model sophistication, computational efficiency, and practical memory usage.

Future work could explore richer temporal dynamics, such as rolling trends, seasonality adjustments, and event-driven signals. Customer segmentation or cohort-based modeling may enable more personalized promotional strategies. Ensemble approaches and advanced model architectures could further improve predictive accuracy, while automated feature selection and unsupervised learning techniques might uncover latent customer or product structures to support enhanced supervised prediction and scalable deployment.

## 6. AI Usage Acknowledgement

ChatGPT and Claude were used as a supplementary tool for code drafting, debugging suggestions, and clarification of documentation during model development. All generated code and explanations were independently reviewed, validated, and adapted before inclusion in the project.

Example prompts used during development included:

• "Help refactor this LightGBM training loop to include early stopping."

• "Propose memory-efficient feature engineering for RFM-like customer features on transaction history."