**Feature Engineering**

This report presents a feature engineering approach for predicting customer loyalty in e-commerce by identifying potential loyal customers and focusing promotional efforts on this segment. By extracting various features from a dataset containing promotional shopping events, the proposed approach aims to enhance the performance of machine learning models for predicting customer loyalty. The results of this study can be employed by e-commerce platforms and merchants to optimize their promotional strategies and maximize profits.

In the competitive e-commerce market, merchants often struggle to retain customers due to the prevalence of one-time buyers attracted by promotions. To maximize return on investment (ROI) and minimize promotion costs, it is essential for merchants to differentiate between one-time buyers and potential loyal customers. In this report, we propose a feature engineering approach to extract meaningful features from a dataset containing promotional shopping events. These features can be utilized to train machine learning models for predicting customer loyalty.

The dataset used for this study comprises user interactions, user information, and merchant information from an e-commerce platform. It includes data on promotional shopping events, capturing various aspects of user behavior and merchant performance.

Feature engineering is a critical step in the development of machine learning models, as it allows for the creation of new features that provide insights into customer behavior and preferences. In this report, we focus on the extraction of features relevant to customer loyalty. Features we came up with include:

* **Click-to-purchase ratio**: This feature represents the likelihood of a user making a purchase after clicking on a product. It can help identify users who show a higher inclination towards making purchases, which may indicate potential loyalty.
* **Unique users**: The number of unique users that have interacted with a merchant provides a measure of the merchant's popularity. A larger number of unique users might suggest that the merchant is successful in attracting customers and has the potential to convert them into loyal customers.
* **Unique categories**: Users who have interacted with a wide range of categories could be more likely to make repeat purchases as they may have diverse interests. Tracking the number of unique categories a user has interacted with can provide insights into their preferences and help in tailoring promotional offers to target their interests effectively.
* **Unique merchants**: Users who have interacted with multiple merchants may be more open to trying new products and exploring various options on the platform. By analyzing the number of unique merchants a user has interacted with, we can identify users who may be more likely to make repeat purchases from different merchants, thus increasing their potential for loyalty.
* **Unique brands**: Users who engage with a variety of brands might have a higher propensity to make repeat purchases, as they may be interested in exploring different products and options. Understanding the number of unique brands a user has interacted with can help in designing promotional strategies that cater to their preferences, thereby increasing the likelihood of customer loyalty.
* **Merchant's overall click-to-purchase ratio**: This feature reflects the overall performance of a merchant in converting clicks into purchases. A higher ratio might indicate that the merchant is effective at closing sales, which could lead to increased customer loyalty.
* **Merchant's overall add-to-cart-to-purchase ratio**: Similar to the click-to-purchase ratio, this feature measures a merchant's ability to convert users who add items to their cart into actual buyers. A higher ratio suggests that the merchant effectively encourages users to complete their purchases.
* **Merchant's overall add-to-favorite-to-purchase ratio**: This ratio measures the merchant's ability to convert users who add items to their favorites into actual buyers. It can help identify merchants who are successful at turning user interest into sales, potentially leading to increased customer loyalty.
* **Interaction count and interaction summary**: These features capture the number of interactions between each user and merchant. A higher interaction count might indicate a stronger relationship between the user and the merchant, increasing the likelihood of customer loyalty.
* **Day of the week and month interactions**: Understanding user behavior patterns during different days of the week and months can help identify potential opportunities for targeted promotions. By extracting the day of the week and month from the time\_stamp, we can analyze the number of interactions (clicks, add-to-cart, purchases, add-to-favorites) for each user per day of the week and month. This information can help in recognizing specific days or months when users are more likely to engage with the platform and make purchases, which can be used to optimize promotional strategies to increase customer loyalty.
* **Demographic group**: The demographic group feature combines age range and gender information to create a categorical feature that can be used to identify differences in purchasing behavior across various demographic groups. This information can help tailor promotional offers to specific segments, increasing the likelihood of repeat purchases.
* **User's lifetime value (LTV)**: LTV represents the total revenue generated by a user since their first purchase. Identifying high-LTV customers who are more likely to make repeat purchases can help merchants focus their promotional efforts on these valuable customers.
* **Frequency of previous purchases**: The number of times a user has made a purchase before the current promotional event. Customers with a higher frequency of past purchases are more likely to be loyal customers.
* **Time since last purchase**: The time elapsed since a user's most recent purchase. Users who have made a purchase more recently might be more likely to make another purchase.
* **Average days between purchases**: This feature helps identify users with a consistent purchasing pattern, who are more likely to make another purchase within the next 6 months.
* **User's average order value**: The average value of a user's past orders. Users who tend to spend more per order might be more valuable to target.
* **User's promotional engagement**: The number of promotional events a user has participated in or the number of promotions redeemed by the user. Users who engage more with promotions might be more likely to make repeat purchases.
* **Category or product affinity**: The user's preference for certain product categories or specific items. This information can help tailor promotional offers to individual user preferences, increasing the likelihood of repeat purchases.
* **User's browsing behavior**: The number of pages viewed, time spent on the platform, or the bounce rate of a user. Users with more active browsing behavior might be more engaged and more likely to make repeat purchases.

These features were chosen based on their relevance to customer loyalty and their ability to provide insights into user behavior and preferences. They capture various aspects of user interaction, merchant performance, and user demographic information.

Feature engineering plays a crucial role in improving the performance of machine learning models for predicting customer loyalty. By extracting meaningful features from the dataset, we gained insights into user behavior and preferences, enabling merchants to focus their promotional efforts on potential loyal customers.