**Experiment Use Case Document for DiscountMate Price Prediction and Product Recommendation Model**

**1. Introduction**

In the ever-competitive grocery industry, maintaining a customer-centric approach while optimizing profitability is crucial. We propose the development of two machine learning models: a **Price Prediction Model** and a **Product Recommendation Model**, designed to enhance the grocery shopping experience. The Price Prediction Model will provide dynamic pricing suggestions based on various factors, while the Product Recommendation Model will deliver personalized product recommendations to users. Both models aim to improve business outcomes by optimizing pricing strategies and enhancing customer retention through personalized shopping experiences.

**2. Dataset Overview**

The primary dataset leveraged for these models is the **Instacart Market Basket Analysis dataset**, sourced from Kaggle (Sparks, P., n.d.). This dataset provides valuable insights into grocery shopping behaviour, containing over **3 million orders** from **200,000+ Instacart users**, along with detailed information about the products, orders, and associated departments.

Key features of the dataset:

* Over **3 million** grocery orders from more than **200,000** Instacart users.
* Detailed product and order metadata.
* Temporal information, including time-of-day and order intervals, enabling robust time-based predictions and analyses.

**3. Experiment Objectives**

**Primary Objectives:**

1. **Price Prediction:**

* Predict optimal product prices based on historical pricing, category trends, and sales data.
* Provide actionable insights into pricing strategies for better competitiveness.

1. **Product Recommendation:**

* Build a system that recommends relevant products based on purchase history and trends.
* Increase user engagement by improving recommendation relevance and personalization.

**Secondary Objectives:**

* **Seasonal Trend Analysis:** Explore and analyze trends in product popularity and pricing fluctuations over different seasons.
* **User Retention:** Enhance the user experience by tailoring suggestions and pricing in real time, increasing engagement and retention.

**4. Methodology**

**4.1 Price Prediction Model (Price.ipynb)**

The goal of this model is to develop a machine learning pipeline to predict product prices based on several features, enabling dynamic pricing strategies. The model is trained on product category, historical sales, and pricing data.

**Data Preprocessing:**

* **Loading Data:** Import datasets containing product details and price history.
* **Handling Missing Data:** Impute missing values using techniques such as median imputation or removal of incomplete records.
* **Feature Engineering:**
  + Create features such as **Product Category** (e.g., Fruits, Vegetables), **Brand**, and **Product Attributes** (e.g., weight, organic status).
  + Utilize historical sales and pricing trends to generate relevant features.
* **Normalization:** Apply standardization (using StandardScaler or MinMaxScaler) to continuous numerical variables, ensuring that all features are on the same scale for models like Linear Regression.

**Exploratory Data Analysis (EDA):**

* **Visualizing Correlations:** Use correlation matrices to identify key relationships between features such as product category and price.
* **Outlier Detection:** Identify and handle outliers in price data through box plots and distribution analysis.

**Model Development:**

* **Linear Regression**: Interpretable model for price prediction based on the linear relationship between variables.
* **Random Forest Regressor**: Captures complex interactions between features and offers improved accuracy through ensemble learning.

**Model Evaluation:**

* **Mean Absolute Error (MAE)**: Measure the average error between predicted and actual prices.
* **Root Mean Squared Error (RMSE)**: More sensitive to larger errors, providing insights into model robustness.
* **R² Score**: Quantifies how well the model explains the variance in product prices.

**Price Prediction Output:**

* Once trained, the model provides price recommendations for new products, integrated into the grocery app for real-time pricing insights.

**4.2 Product Recommendation Model (product recom.ipynb)**

The Product Recommendation Model is designed to enhance the user experience by delivering personalized product suggestions based on past purchase behaviour.

**Data Preparation:**

* **Loading Data:** Transaction data, including user\_id, product\_id, and order\_id, is loaded for analysis.
* **User-Item Matrix:** A sparse binary matrix is created, with rows representing users and columns representing products, indicating purchase behavior.

**Collaborative Filtering:**

* **Item-Based Filtering:** Suggests products frequently purchased together by other users.
* **User-Based Filtering:** Identifies similar users based on past purchasing patterns and recommends items favoured by those users.

**Association Rule Learning:**

* **Market Basket Analysis:** Uses the Apriori algorithm to find product pairs or groups frequently purchased together. For instance, users who buy milk are often recommended bread.

**Evaluation Metrics:**

* **Precision**: Measures the proportion of relevant products among the top K recommendations.
* **Recall**: Assesses the system's ability to retrieve relevant products.
* **F1 Score**: Balances precision and recall to evaluate overall recommendation quality.

**Recommendation Output:**

* The model delivers real-time recommendations through the app based on a user's recent purchases, helping increase sales by promoting cross-selling opportunities.

**5. Experimental Setup**

**Data Splitting:**

* Both models are trained using an 80/20 train-test split, ensuring generalization to unseen data.
* **Temporal splitting** is used to account for time-sensitive purchasing trends.

**Model Tuning:**

* **Grid Search** and **Random Search** are used for hyperparameter tuning, optimizing model performance.
* Advanced techniques like **Bayesian Optimization** may also be used to explore the hyperparameter space efficiently.

**6. Ethical Considerations**

* **Data Privacy:** User purchase history and interactions are anonymized to ensure privacy and data security.
* **Fairness in Recommendations:** The model is designed to ensure that recommendations are not biased towards specific brands or product types, promoting a fair user experience.

**7. Integration with the DiscountMate App Platform**

The Price Prediction and Product Recommendation Models will be integrated via API endpoints, enabling seamless deployment into the existing grocery app infrastructure. A feedback loop will be incorporated, allowing the models to learn from real-time user interactions and continuously improve over time.

**8. Future Work and Scalability**

**Potential Enhancements:**

* **Reinforcement Learning** for adaptive pricing and recommendations.
* **Graph-Based Neural Networks** to improve user-product relationship modelling.
* **Context-Aware Recommendations** using data such as location, time of day, and external factors (e.g., weather).

**Scalability:**

* The system will be designed to scale to handle millions of users, with periodic model retraining to adapt to changes in user behaviour and market conditions.

**9. Expected Outcomes and Impact**

The models are expected to provide:

* Accurate price predictions that enhance profitability.
* Personalized recommendations that improve user engagement and retention.
* Insights into consumer behaviour that drive better business decisions.

By leveraging machine learning, the grocery app will be equipped with tools to offer dynamic pricing and personalized shopping experiences, positioning it as a market leader.

**References**

1. GitHub. (n.d.). melodygr/grocery\_recommendation. GitHub. Retrieved [31/07/2024], from <https://github.com/melodygr/grocery_recommendation>
2. Sparks, P. (n.d.). Instacart Market Basket Analysis. Kaggle. Retrieved [31/07/2024], from <https://www.kaggle.com/datasets/psparks/instacart-market-basket-analysis>