-CAPSTONE 2 PROJECT PROPOSAL-

Customer Behavior Analysis for Instacart

-Ryan Mennemeier

Introduction:
Understanding and anticipating customer purchasing behavior is critical for improving user experience and optimizing inventory management in e-commerce platforms. By leveraging customer segmentation, customer churn prediction, and Customer Lifetime Value (CLV) prediction, we will provide actionable insights to enhance marketing strategies, optimize inventory management, and optimize revenue generation. The project will employ various data science methodologies, including clustering and classification models.

Problem Statement:
The objective is to build predictive models capable of identifying distinct customer segments, forecasting customer churn, and predicting Customer Lifetime Value based on historical order data. These models will enable Instacart to tailor their marketing strategies, implement targeted retention efforts, and optimize revenue generation.

Context:
This project is situated within the broader scope of e-commerce and data science applications. Predictive modeling of customer behavior can significantly impact business decisions, marketing strategies, and inventory management, leading to enhanced customer satisfaction and operational efficiency. The Instacart dataset provides a rich source of information for developing these predictive models.

Criteria for Success:

1. **Model Accuracy:** Achieve high accuracy and reliability in segmenting customers, predicting customer churn, and estimating CLV, measured by metrics such as F1 score, precision, recall, and ROC AUC.
2. **Performance Improvement:** Demonstrate a significant improvement over baseline models through advanced techniques like feature engineering and hyperparameter tuning.
3. **Scalability:** Ensure the solution is scalable and can handle large datasets efficiently.
4. **Stakeholder Satisfaction:** Deliver insights and results that meet the needs and expectations of key stakeholders.

Scope of Solution Space:

1. **Customer Segmentation:** Focus on developing models that accurately segment customers based on purchasing behavior.
2. **Churn Prediction:** Build models to predict customer churn, enabling proacting retention strategies.
3. **CLV Prediction:** Develop models to estimate the lifetime value of customers to inform marketing and sales strategies.

Constraints within the Solution Space:

1. **Data Limitations:** The quality and completeness of the Instacart dataset may impose constraints on the model's performance.
2. **Computational Resources:** The availability of computational resources may limit the complexity of models and the extent of hyperparameter tuning.
3. **Time Constraints:** Project timelines may affect the depth of analysis and the extent of model optimization.
4. **Implementation Constraints:** Practical considerations for deploying the models in real-world scenarios may impose additional constraints.

Key (Potential) Stakeholders:

1. **E-commerce Platforms:** Companies like Instacart that rely on accurate demand forecasting and personalized recommendations.
2. **Data Scientists and Analysts:** Professionals who will use the insights and methodologies developed in this project.
3. **Business Decision Makers:** Executives and managers who will leverage the predictive models for strategic decisions.
4. **Customers:** End-users who will benefit from improved recommendations and shopping experiences.

Data Sources:
Kaggle(Instacart Market Basket Analysis)
https://www.kaggle.com/c/instacart-market-basket-analysis/data