-CAPSTONE 2 PROJECT PROPOSAL-

Next-Order Predictions with Model Stacking and RNN-LSTM: An Advanced Approach to Instacart User Behavior

-Ryan Mennemeier

Introduction:
Understanding and anticipating customer purchasing behavior is critical for improving user experience and optimizing inventory management in e-commerce platforms. This project uses the Instacart dataset to predict future purchases, enabling personalized recommendations and better operational decisions. By leveraging model stacking, RNN-LSTM networks, and Bayesian Optimization, we seek to enhance prediction accuracy and robustness.

Problem Statement:
The objective is to build a predictive model capable of forecasting the next 'n' purchases for users based on their historical order data. We will implement and compare various machine learning models, including logistic regression, SVD, XGBoost, and RNN-LSTM networks, and enhance model performance through model stacking and Bayesian optimization.

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 predicting next-order purchases, 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 model stacking and Bayesian optimization.
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. **Predictive Modeling:** Focus on developing models that accurately predict next-order purchases.
2. **Model Comparison:** Compare multiple machine learning approaches to identify the best-performing models.
3. **Optimization Techniques:** Use advanced optimization methods to enhance model performance.
4. **Scalability:** Ensure that the solution can handle large-scale data efficiently.

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