## **High-Level Project Document: Predicting Customer Satisfaction for E-commerce Platforms**

**Project Overview**

This project aims to develop a model to predict customer satisfaction with a product purchased on an e-commerce platform. By understanding customer sentiment, the platform can personalize recommendations, improve product selection, and ultimately enhance customer experience.

**Data Acquisition**

* **Orders**
* **Products**
* **Sellers**
* **Product Categories**
* **Customer Locations**
* **Order Items**
* **Order Payments**
* **Customer Reviews**

**Data Preprocessing and Feature Engineering:**

1. **Merging Datasets: The datasets are merged to create a comprehensive view of customer purchases and reviews.**
2. **Missing Values: Missing values are removed to ensure data quality.**
3. **Binary Classification: Review scores are converted into a binary format (satisfied: 1, dissatisfied: 0) for the prediction model.**
4. **Feature Selection: Relevant features such as product ID, seller ID, price, freight value, and product category are chosen to represent the purchase experience.**
5. **Encoding Categorical Variables: Categorical features like product category are converted into numerical values for the model.**

**Model Development and Evaluation**

1. **Data Split: The data is divided into training and testing sets. The training set is used to train the model, and the testing set evaluates the model's performance.**
2. **Normalization: Feature values are normalized to ensure all features are on a similar scale.**
3. **Model Training: A Random Forest Classifier model is trained on the training data. This model learns the relationship between purchase characteristics and customer satisfaction.**
4. **Prediction and Evaluation: The model is used to predict customer satisfaction for the testing data set. The accuracy and classification report are used to assess the model's performance.**