High Level Design Document

For

“Market Basket Project on E-Commerce”

Document Version Control

SL. No	Version No.	Date	Prepared By	Reviewed By	Comments
01	1.0	12th April, 2022	Irfan Razi		

Signature :

_____

Prepared by : Irfan Razi

Intern – Data Scientist

Signature : _____

Reviewed by :

Signature :

_____

Approved by : Sudhanshu Kumar

Founder iNeuron, Chief AI Engineer & CEO,

iNeuron.ai

Table of Contents

List of Figures

17th Floor tower A Brigade Signature Towers,
Sannatammanahalli, Bengaluru, 562129, INDIA

GENERAL

1.1 Introduction

Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary details to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail.
- Describe the user interface being implemented.
- Describe the hardware and software interfaces.
- Describe the performance requirements.
- Include design feature and the architecture of the project.
- List and describe the non-functional attribute like:
    i. Security
    ii. Reliability
    iii. Maintainability
    iv. Portability
    v. Reusability
    vi. Application Compatibility
    vii. Resource Utilization
    viii. Serviceability

1.2 Scope of this Document

The HLD document presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

1.3 Abstract

Most customers do not post a review rating or any comment after purchasing a product which is a challenge for any E-commerce platform to perform If a company predicts whether a customer liked/disliked a product so that they can recommend more similar and related products as well as they can decide whether or not a product should be sold at their end. This is crucial for E commerce-based company because they need to keep track of each product of each seller, so that none of products discourage their customers to come shop with them again. Moreover, if a specific product has very few ratings and that too negative, a company must not drop the product straight away, may be many customers who found the product to be useful haven't actually rated it. Some reasons could possibly be comparing your product review with those of your competitors beforehand, gaining lots of insight about the product and saving a lot of manual data pre-processing, maintain good customer relationship with company, lend gifts, offers and deals if the company feels the customer is going to break the relation.

Objective of this case study is centred around predicting customer satisfaction with a product which can be deduced after predicting the product rating a user would rate after he makes a purchase.

2 General Description

2.1 Product Perspective

The Market Basket Review Scoring solution system is a Data Science-based Machine Learning model which help the client to predict the Review Score from the customer and improve their services based on that.

2.2 Problem Statement

To create an AI solution for predicting the Review Score by the customer who has not reviewed the product.

2.3 Proposed Solution

The solution proposed here is a data science model based on machine learning can be implemented to perform above mention use cases. First, we will collect the data from the customers who has reviewed the products and train and validate the model from it. Then, we will get the review score for those customers who has not reviewed their purchase order.

2.4 Further Improvements

Market Based Review Scoring solution system can be added with more use cases like product recommendation system for a particular customer, gift vouchers for regular customer suggestion, location-based service improvement, predict the high-demand time so that shopkeepers can be ready with their products availability, improve the profitability by understanding the majority of the Payment mode, customer behaviour analysis, cross-marketing on online stores, deliver targeted marketing.

2.5 Data Requirements

Data requirement completely depend on our problem statement.

We need data of the customers who has given review score for their purchases. We will be needing these many attributes

1. order_item_id: This attribute is the quantity of the product purchased. It is int type.
2. product_weight_g: This attribute is the weight of the item in grams. It is int type.
3. payment_installments: This attribute is the number of installments in which customer paid for the product. This is int type.
4. order_delivered_customer_time_in_days: This attribute is created from other attirbutes which is the timestamp for purchase order placed and delivery date. This is int type.

5. product_volume: This attribute is created from other attributes which are length, width and depth of the product. This is int type.

6. customer_seller_distance: This attribute has been created from the latitudes and longitudes of the customer and seller. This is int type.

7. order_purchase_year: We have data from year 2016 to 2018. This is int type.

8. total_payment: This is combined attribute of price and freight value for the product. This is int type.

9. order_status: The order might be delivered, in-processing, cancelled, etc. This attribute has been encoded to int type.

10. payment_type: Customer may make the payment via credit card, debit card, voucher, etc. This attribute has been encoded to int type.

11. product_category_name: There are wide range of product categories. This attribute has been encoded to int type.

12. Timing: This attribute is to understand at what time customer has placed the order. It could be Morning, Evening, Afternoon and Midnight. This attribute has been encoded to int type.

13. Season: This attribute gives the idea during which season the purchase has been made. It could be summer, winter, etc. This attribute has been encoded to int type.

2.6 Tools Used

Python programming language and frameworks like NumPy, Pandas, Scikit-learn, Matplotlib, seaborn, Plotly, Flask, etc are used to build the whole model.

- Visual Studio Code is used as IDE.
- Plotly, Matplotlib and Seaborn is used for Visualization.
- Github is used as Version Control System.
- Python and Flask are used for backend development.
- Heroku is used for model deployment.
- Docker hub is used for Dockerization.
- Circleci is used for CI-CD pipeline.

2.7 Constraints

The Market Basket Review Scoring solution system should be very accurate. It should not mislead any report. It should be as automated as possible and users are not required to know its working.

2.8 Assumptions

The main objective of the project is to implement the use cases as previously mentioned for new dataset that comes through the client which has this solution install in their system to capture people's review score. This way they can plan and focus on target marketing.

3 Design Details

3.1 Process Flow

For Customer Review Score, we will use machine learning base model. Below is the process flow diagram.

Figure 1 gives the detail regarding proposed methodology.

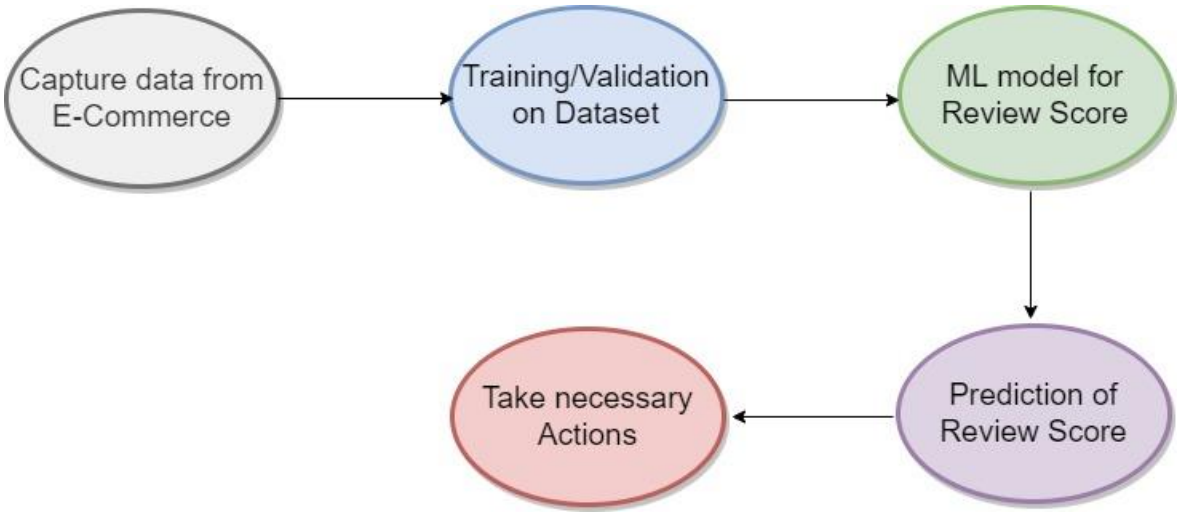

Figure 1: Proposed Methodology

3.1.1 Model Training and Evaluation

Figure 2 gives the detail regarding model training and evaluation process.

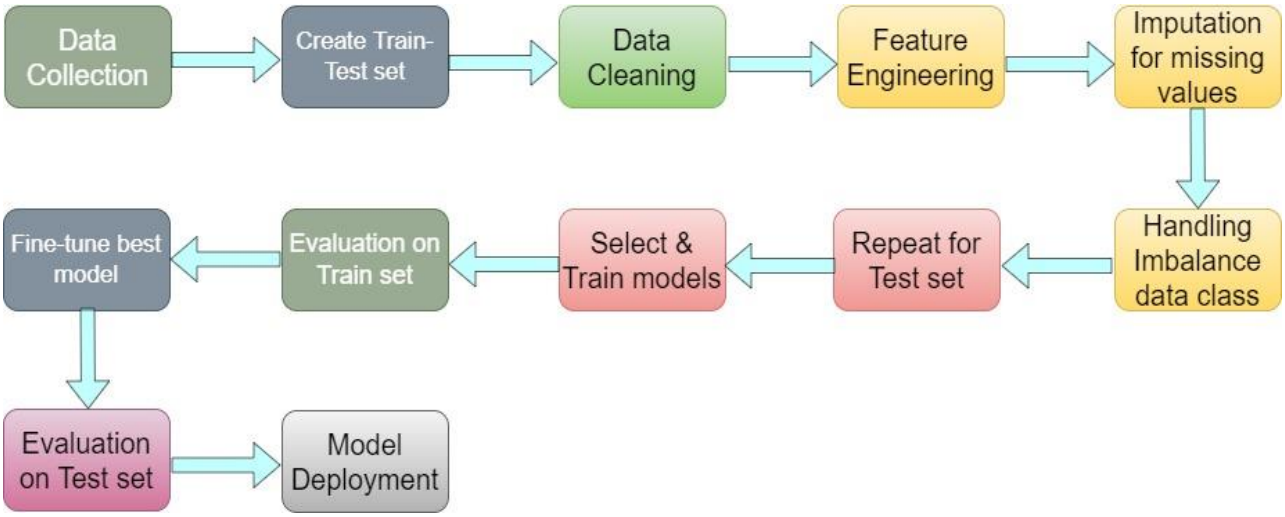

Figure 2: Model training and Evaluation process

3.1.2 Deployment Process

Figure 3 gives details regarding deployment process.

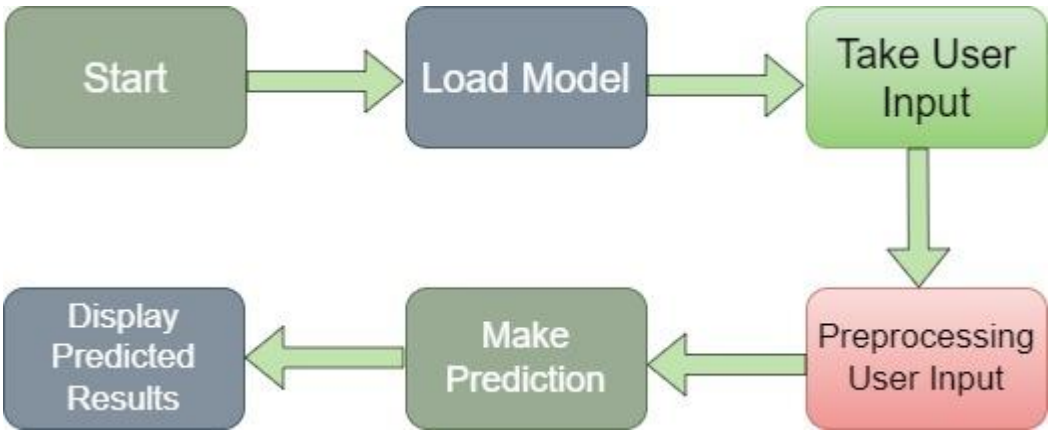

Figure 3: Deployment Process

3.2 Event Log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

1. The system identifies at what step logging is required.
2. The system should be able to log each and every system flow.
3. Developer can choose logging method. You can either choose database logging or file logging.
4. System should not hang even after using so many loggings. Logging is done just because we can easily debug issues. So, logging is mandatory to do.

3.3 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage. Errors are logged as well for faster debugging.

4 Performance

The machine learning based Review Score solution will be used for scoring the Review score by the customer who has not submitted any review so that necessary market plans can be proposed and executed. Also, model re-training is very crucial to improve performance.

4.1 Reusability

The code written and the components used should have the ability to be reused with no problems.

4.2 Application Compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

4.3 Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

4.4 Deployment

5 Conclusion

Market Basket Review Scoring solution will take E-Commerce data of those customers who have submitted their Review scores. And, then it will leverage its prediction to get the review score from the customers who has not submitted the review scores. With this, the stakeholders can get better insight about the customer sentiments and take necessary action to increase the profitability and better serviceability to the customers.

6 References

Brazilian E-Commerce Public Dataset by Olist:

url: https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce