Low-Level Design

Python Web Scraping

Written By	Bhavya Shah
Document Version	0.1
Last Revised Date	16 – Feb -2023

FOOD RECOMMENDATION LLD i ii LOW LEVEL DESIGN (LLD)

Document Control

Change Record:

Version	Date	Author	Comments
0.1	16 – Feb - 2023	Bhavya Shah	Introduction & Architecture defined

Contents

1. Introduction

1.1. What is Low-Level design document?

The goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for Food Recommendation System. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

1.2. Scope

Low-level design (LLD) is a component-level design process that follows a step-by step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design wor.

3. Architecture Description

3.1. Data Description

Recipe 1M+ dataset is the biggest publicly available recipe dataset. The information each recipe contains is separated in two JavaScript Object Notation (JSON) files. This dataset contains 1029715 recipes including 1480 different ingredients.

3.2. Web Scrapping

In order to create a more complete recipe collection we will need some more datasets which will contain Nutritional value of recipes along with Ratings and total Calories.

3.3. Data Transformation

In the Transformation Process, we will convert our original dataset which is in JSON format to CSV Format. And will merge it with the Scrapped dataset.

3.4. Data Insertion into Database

a. Database Creation and connection - Create a database with name passed. If the database is already created, open the connection to the database.
b. Table creation in the database.
c. Insertion of files in the table

3.5. Export Data from Database

Data Export from Database - The data in a stored database is exported as a CSV file to be used for Data Pre-processing and Model Training.

3.6. Data Pre-processing

Data Pre-processing steps we could use are Null value handling, stop words removal, punctuation removal, Tokenization, Lemmatization, TFIDF, Imbalanced data set handling, Handling columns with standard deviation zero or below a threshold, etc.

3.7. Data Clustering

K-Means algorithm will be used to create clusters in the pre-processed data. The optimum number of clusters is selected by plotting the elbow plot. The idea behind clustering is to implement different algorithms to train data in different clusters. The K-means model is trained over preprocessed data and the model is saved for further use in prediction

3.10. Model Building

After clusters are created, we will find the best model for each cluster. For each cluster, algorithms will be passed with the best parameters derived from Grid-Search. We will calculate the AUC scores for models and select the model with the best score. Similarly, the models will be selected for each cluster. All the models for every cluster will be saved for use in Recommendation.

3.11. Data from User

Here we will collect physiological data from user such as user height and weight, heart rate, burned calories, daily physical activity level; as well as information directly provided by the user such as daily food intake

3.12. Data Validation

Here Data Validation will be done, given by the user

3.13. User Data Inserting into Database

Collecting the data from the user and storing it into the database. The database can be either MySQL or Mongo DB.

3.14. Data Clustering

The model created during training will be loaded, and clusters for the user data will be predicted.

3.15. Model Call for Specific Cluster

Based on the cluster number, the respective model will be loaded and will be used to predict/Recommend the data for that cluster.

3.16. Recipe Recommendation & Saving Output in Database

After calling model Recipe/Output will be recommended, this output will be saved in Database and it will beused to show the same Output if other users provide the same data.

3.17. Deployment

We will be deploying the model to AWS.
This is a workflow diagram for the Recipe Recommendation..

4. Unit Test Cases

Test Case Description	Pre-Requisite	Expected Result
Scraped data from the ODI website	1. Web URL should be defined	Web URL should be accessible
Verify whether data is proper or not.	1. URL is necessary	Data should be scraped properly and should be stored in the dataframe and save in csv format.