



UNIVERSITY OF COLOMBO, SRI LANKA

UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING

IS4103/SCS4204/CS4104 – Data Analytics

Academic Year 2024 - Semester 1

Assignment 1

Assignment Tasks

1. Download the “recipes.json” dataset from UGVLE and preprocessing the data.
2. Create a word cloud by using TF-IDF values of each ingredient in the recipes in the given dataset.
3. Show the number of recipes that require a “saucepan” to prepare that food.
4. Plot a bar graph to present the number of recipes written by each author.
5. List down the names of foods that are recommended by the recipes’ authors to make with kids (involve kids to make them). Use the GPT-3 model (OpenAI API) to automatically identify such foods.
6. Foods with the following ingredients are the most delicious foods as said by “customer A”. Recommend the best 12 foods for “customer A” in descending order with a suitable graph to show the reasons behind your recommendations.
 - sweet potato
 - chocolate
 - dried cranberries
 - raspberries
 - icing sugar
 - ginger
 - cherries
 - brandy
7. Download “img1.jpg” and “img2.jpg” images from UGVLE.
8. Draw the best 25 matches between img1.jpg and img2.jpg by using SIFT related methods in openCV and save the result as “sift-result.jpg”
9. Draw the best 25 matches between img1.jpg and img2.jpg by using ORB related methods in openCV and save the result as “orb-result.jpg”
10. Prepare a pdf by comparing the two methods: SIFT and ORB by using the results in sift-result.jpg and orb-result.jpg

Assignment Submission

Submit a .zip file that contains the following:

1. “recipes.json” dataset file
2. “img1.jpg”, “img2.jpg”, “sift-result.jpg”, “orb-result.jpg” image files
3. Seven Python files (or seven Jupyter Notebook files) for the assignment tasks 2, 3, 4, 5, 6, 8, 9.
 - You can name these files as “task2.py”, “task3.py”, ...
4. One PDF file for the assignment task 10

Tips to achieve good marks:

- Follow Python coding best practices
- Document your approaches and any assumptions you made as comments in the code.
- Avoid hard coding.