FINAL PROJECT PART-1 (Proposal)

Submitted by RATAN TEJASWI VADAPALLI

Course: INTRODUCTION TO NLP FOR DATA SCIENCE
Course Code: DSCI-D 590
Professor: Olga Scrivner

1. PROJECT TITLE

The title of the project is:
FlavorGuru: Personalized Recipe Recommendations based on Flavor Profile.

2. TEAM

The only member of the team is **Ratan Tejaswi Vadapalli**. The name of the team is **LoneWolf.**

3. PROJECT DESCRIPTION

Objectives:
FlavorGuru aims to provide personalized recipe recommendations based on a user's taste preferences, dietary restrictions, past likes/dislikes. The system will analyze flavor combinations, ingredients used, and user's feedback to suggest recipes that match the user's palate. This can be done by first getting to know what the user actually prefers to have and find those ingredients and then see if the ingredients of any other recipe have a good match and this might be a good match for the user's liking. With the power of NLP, FlavorGuru will be able to understand what the user prefers, with the help of feedback. It also predicts potential recipe hits for the users, making their food experience better. This mainly helps users save time when they try to find something to cook without actually knowing whether they might like it.

Usefulness:

FlavorGuru is useful because it helps the users identify their flavor profile and save time exploring new recipes according to their taste. There are various recipe recommender systems out there in the internet, such as Yummly or Epicurious, but most of them base their recommendations on broad categories or keyword searches. The uniqueness of FlavorGuru is that it takes an NLP-driven approach into finding the user's flavor profiles by understanding patterns in ingredient combinations and analyzing the ingredients in the recipes users like and gives the users suggestions which are tailored according to their palate. Some of the target users of this application will be home cooks, culinary students who are trying to explore new recipes, amateur cooks, people with dietary restrictions trying to find new recipes. The distinctive features in this application include understanding a flavor from the combination of ingredients.

Data

- Origin: The origin of this data is the Epicurious – Recipes with Rating and Nutrition dataset from Kaggle which was compiled by Hugo Darwood from the Epicurious website. This dataset was created to explore different factors affecting people's enjoyment of food and cooking.
- Format: The data is in csv format.
- Initial Description: Contains recipe details like ingredients, calories, macros and has user ratings.
- Labeled Data: Yes, the dataset has labeled data in the form of user ratings.
- The number of records in the dataset is 20052, which is the number of recipes. The number of columns in the data are 680, which include title, ratings, macros, calories, and individual ingredients. Upon examining the dataset using pandas, it is found out that there are 4188 records with some missing information or null values in some columns.
- Cleaning Requirements: It requires text-preprocessing and cleaning noisy data. The missing nutritional values in some recipes needs to be addressed, ingredients must be standardized and check if there are duplicate ingredients with different names, check for duplicates in the recipes, and extraction of meaningful data from different ingredients.

Functionalities

- **NLP Functions:**
  i. **Flavor Profile Extraction:** This includes analyzing the ingredients to determine the flavor profile of a recipe and later match it with the flavor profile of a user according to their palate. An extension to this could be using unsupervised learning to cluster recipes into different flavor categories.
  ii. **Recipe Similarity:** Finding similarity among different recipes according to the ingredients, ratings and macros.
  iii. **Feedback Loop:** Use the user's feedback and past choices to enhance recommendations.

- **User Interaction:**
  i. **Dietary Preferences:** Users can set their dietary preferences like vegetarian, low calorie, vegan, gluten-free, etc. so that they only get filtered recipes.
  ii. **Ingredient Recipe Search:** Users can input a list of ingredients and search for recipes which can be cooked using those ingredients, or most of those ingredients.
  iii. **User Feedback:** Users can give feedback on recommendations or even recipes in general to refine future recommendations.

Communication and Sharing

This project is an individual project, so there is no need for a communication channel. However, it requires a disciplined project versioning and tracking. For this reason, a GitHub repository has been created to maintain the code and other artifacts. The GitHub repository currently contains the readme file with the project description and the csv file of the data in a zip format, which has been compressed due to large size and size restrictions on GitHub.

GitHub Link: https://github.iu.edu/ravadap/d590nlp_flavorguru
Kaggle Link: https://www.kaggle.com/datasets/hugodarwood/epirecipes/

4. PERSONAL CONTRIBUTION STATEMENT

I plan to spend a minimum of 6 hours per week on this project and adjust my schedules accordingly. The next plan is to breakdown the project into different tasks, set deadlines for each task and create a JIRA board and all the tasks in the board. I will work to finish those goals according to the deadlines set, so that the project will be on track to completion. Proper planning and discipline are necessary to achieve a task, so, I will try to be as disciplined as possible to work on the project as planned.

5. REFERENCES

Kaggle: https://www.kaggle.com/datasets/hugodarwood/epirecipes/
APA Guidelines: https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/general_format.html#:~:text=General%20APA%20Guidelines,title%20and%20the%20page%20number.