CSCE 581 Project Report

November 30th, 2023

Tyler Beetle

Problem

According to a study by the American College Health Association, only 18.4% of college students reported eating 3 or more servings of fruit daily and 30.4% reported eating 3 or more servings of vegetables daily (Soldavini, 2021). This dietary pattern is often influenced by various constraints, including financial limitations, dietary restrictions, and a lack of cooking experience. Notably, 54.3% of students report that they only cook occasionally or they never cook (Soldavini, 2021). This is observed at the University of South Carolina in Columbia, SC, where a number of observed peers receive the majority of their diets from commercial restaurants on or off campus. Understanding the challenges faced by college students in maintaining a healthy diet, it becomes evident that providing effective cooking guidance can significantly contribute to improved nutrition. As a response to this identified problem, we propose the development of a recipe recommender chatbot tailored specifically for college students. Student populations often juggles academic commitments, part-time jobs, and social activities, leaving minimal time for meal planning and preparation. This recipe recommender chatbot aims to provide useful recommendations for college students.

The primary objective of this chatbot is to offer personalized recipe recommendations based on individual preferences, dietary restrictions, and lifestyle. By allowing users to input their preferred ingredients or other criteria, the chatbot aims to provide a curated list of recipes that not only meet their nutritional needs but also align with their cooking capabilities. This machine learning algorithm addresses the common barrier of limited cooking experience among college students. The recipe recommender chatbot will furnish users with comprehensive information for each suggested recipe, including a detailed description, a list of ingredients, and step-by-step cooking directions. This approach not only empowers students to make informed and healthier food choices but also encourages them to explore and enjoy the cooking process. Overall, the recipe recommender chatbot serves as a practical solution to solve the challenge of decreasing nutrition in the 18-25 year old demographic by promoting healthier eating habits among college students.

Related Work

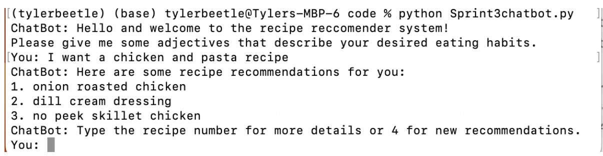
Recognition of food, ingredients and recipe recommendations are the context of numerous machine learning based algorithms. However, the majority of said algorithms are used for image recognition for recipe recommendation. The majority of these recipe recommendations utilize convolutional neural networks (CNN's) for object recognition and classification. Through research, several articles were identified that discuss recipe recommendations similar to the proposed recommendation chatbot. The first study identified was the development of the RecipeIS model which is based on the visual recognition of ingredients. RecipeIS leverages a specific type of CNN called a ResNet-50, which is identified as a frequent CNN type that is leveraged in image classification. The model operates in two parts. The first part aims at identifying the food ingredients through a specified dataset, while the second is giving recipe recommendations based on the ingredient detection. A second related work was identified from Masaryk University in the form of a content based recommender system. This iteration of a recipe recommender is text-based, but it is also strictly ingredient based. The recommender is the subject of a graduate level thesis project, so it is not a deployed commercial project. The project discusses multiple methods of developing a result, but they decide to use vectorization, specifically cosines similarity, as the primary method for developing a result. They drive the cosine matching strictly off of the ingredients and no other categorization. A third recommender system from the M.B. Vivek, N. Manju, and M.B. Vijay of JSS Academy of Higher Education & Research in India develops a text based recommendation system, but uses collaborative filtering for its model training. It primarily drives its training from the preferences of users and mapping users to other users. They leverage Euclidean Distance mapping, which is the comparison between two data points that have been graphed out. They map the users preferences to other similar user preferences in order to drive recommendation results. The students of JSS Academy do this by leveraging a neighborhood system of 200 neighborhoods where the inputted user is compared to a neighborhood of 200 neighbors.

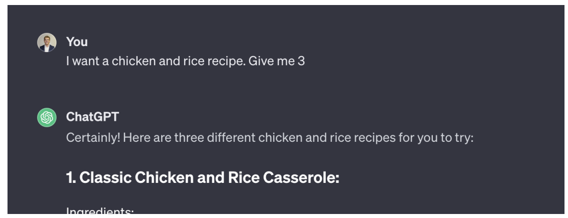
Approach

We took a two stepped approach to the text based recipe recommender system. The first step of the system is to take in the users input as to what their desired recipe recommendation is. The user's query is then pointed to the model which is trained on top of a dataset scrapped from Food.com. The dataset was obtained from Kaggle, where its original size was 230,000. The data was then cleaned to get rid of the unnecessary columns and to cut the size of the dataset down. The new dataset contains 30,000 entries of recipes, each containing individual ingredients, a description, and steps for the recipe. Each of these individual entries is vectorized using the sci-kit learn library, meaning that each individual entry is turned into a numeric vector. The entries are vectorized by taking each ingredient and the recipe steps and grading that using cosine similarity. These vectors are graphed on a two dimensional plane and then stored for later. The user's input is vectorized using scikit-learn as well, which takes in the entirety of the user's query. The user's query is then passed to comparison against the vectorization of the dataset. The query vector is compared against the vectors from the dataset and the top three indexed vectors from the dataset are returned to the user. The vectorization uses a grading scale that tokenizes that emulates cosine similarity in the form of 0 to 1. The user then gets to select which recipe match is their choice, and the ingredients, description, and steps of the recipe are returned to the user. The usage of cosine similarity enables the system to compare the user's query against all instances in the dataset without having to manage a dynamic user collaborative system. In essence, the system's design combines the power of machine learning with a user-centric interface, aiming to streamline and enhance the culinary discovery experience for users seeking personalized and delicious recipes.

Evaluation

Prior to conducting any analysis of the recipe recommender, we initially inferred that ChatGPT would be stronger than the developed recipe recommender. This is due to the sheer manpower and computing capabilities that Open AI has. we imputed 25 of the same prompts into both my ChatGPT and my recipe recommender, 3 of which we used in the working demo video. We used a binary grading system of 0 to 1 for each individual prompt for both ChatGPT and my recipe recommender. An example prompt I used was "I want a chicken and pasta recipe." Here are how the examples look in both formats:





I then took the results of the 3 results and determined their association with the prompted query. From the three responses, I either gave them a point if the three recommendations had 2 or 3 valid recommendations and 0 points if there were 0 or 1 valid recommendations. After doing this 25 times, I determined that ChatGPT was accurate nearly 100% of the time while my model was only accurate 76% of time based on my preset criteria.

Overall, it is evident that corporate Large Language Models (LLM) are significantly more accurate than my recipe recommendation model. This is due to multiple factors that put the larger corporations at an advantage against my development of models. One primary reason is due to the level of experience and education they have opposed to solely an undergraduate education. Another reason is due to their large financial backing and capabilities. To improve the recipe model in specific, there could be an implementation with specific measurements for all ingredients. It also probably could be improved with a different type of grading system that is more advanced than cosine similarity. The chatbot component could be improved with the implementation of Rasa to handle the basic interactions between user and bot. Overall, the recipe recommender system is a good building block but there are multiple areas for improvement.

Working Demo

Attached is a link to a working demo of the recipe recommender with chit chat and 3 individual example prompts. link.

References

Rodrigues, Miguel Simões, et al. "RecipeIS—Recipe Recommendation System Based on Recognition of Food Ingredients." *Applied Sciences*, no. 13, MDPI AG, July 2023, p. 7880. *Crossref*, doi:10.3390/app13137880.

Soldavini, Jessica, and Maureen Berner. "Characteristics Associated with Cooking Frequency among College Students." *International Journal of Gastronomy and Food Science*, Elsevier BV, Apr. 2021, p. 100303. *Crossref*, doi:10.1016/j.ijgfs.2021.100303.

Vivek, M. B., et al. "Machine Learning Based Food Recipe Recommendation System." *Proceedings of International Conference on Cognition and Recognition*, Springer Singapore, 2017, pp. 11–19, http://dx.doi.org/10.1007/978-981-10-5146-3_2.