

VegVista

~ a happy vegan ~

By Anupa Jayakody

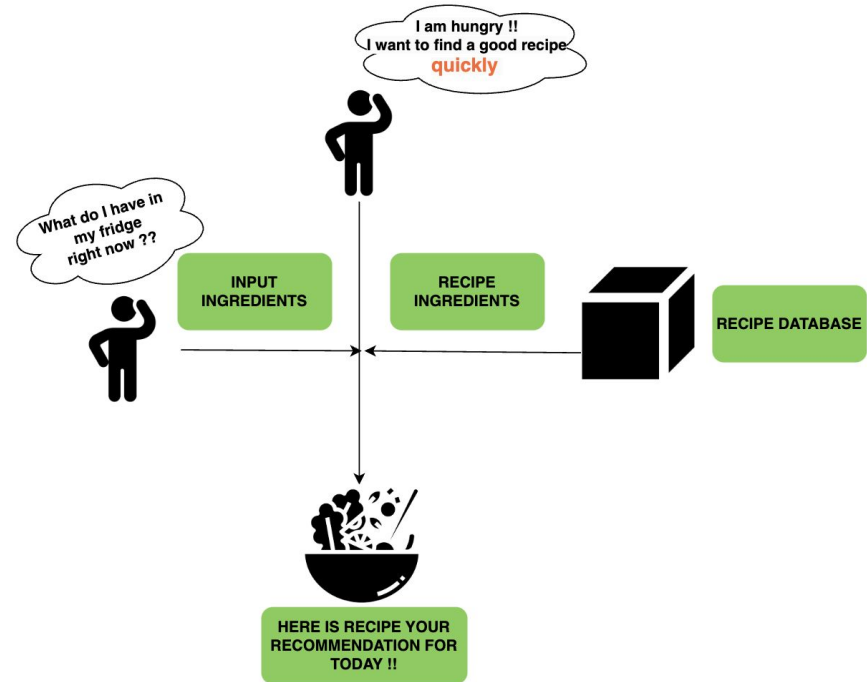
PROBLEM AREA & PROJECT OVERVIEW

With majority of the recipe sources being catered to non-vegans, scope of the project is to make you 'A Happy Vegan' by developing a solution to recommend vegan recipes and make them available at your fingertips. This can be used by any non-vegan, who wishes to follow a vegan recipe as well.

Project scope is to develop a 'Recipe Recommender' for any user to easily find a recipe based on the ingredients of their liking.

Project will target to deliver 2 results;

1. Recommend a vegan recipe
2. Recommend a vegan recipe based on the ingredients preferred



PROJECT VISION WITH DATA SCIENCE

- The solution will be developed with the use of Machine Learning languages; NLP and Recommender systems.
- This will be done by mapping the input ingredients to the recipe ingredients and recommending the recipes with the highest similarity.

Github link:

<https://github.com/anupa-jayakody/Capstone>

POTENTIAL IMPACT

A vegan recipe in 2 mins

- A web/mobile application can be developed to search recipes that will reduce the time taken by each user to find a recipe on the internet.
 - This will provide you a recipe or recipes within 1-2 mins where as generic searching will require you to read through the full recipe to gauge whether you can prepare them with the ingredients you have or whether you will like it).
-

DATASET

Source :

Kaggle (www.vegunary.com)

No.of records:

~1300 recipes

Data :

1. Recipe title
2. Recipe ingredients
3. Preparation method
4. URL

Pre processing & EDA :

1. There were 9 duplicate records, no null values in the data columns and data was cleaned using NLTK library; tokenization, lemmatization, removing stop words, lower casing, removing additional characters, digits and common words.
2. Based on the titles; it can be seen that database consist of all types of recipes, main meals, breakfast, snacks, desserts etc.
3. Most common ingredients used in the recipes are ; garlic, onion, olive, coconut, lemon, tomato, almonds.
4. Longest recipe is one with ~50 ingredients and the shortest is ~4.
5. Most common recipe types are; salads, soups, burgers, tofu, quinoa, chickpea, potato, lasagna, chocolate based recipes.
6. Least common ingredients are; cornflakes, parsley, hash browns.
7. Most key ingredients are usually used with another word that is associated with an ingredient. E.g olive oil

NEXT STEPS

1. Since both ingredients and titles need to be analysed further by considering word combinations /sequence of occurrence, advanced NLP techniques such as n-grams to be used. e.g: the most common ingredient 'red' doesn't imply an ingredient alone.
2. Clustering analysis to be done to cluster similar types of recipes to categorise them and give an option to the user to select the type of recipe.
3. Similarity between input ingredients and database ingredients to be mapped using cosine similarity or a similar method.