CookSmart

Final Project Presentation

Course : DATA 515

Date : 3/16/2021

Team : Marc Mascarenhas
Sandeep Tiwari
Sandhya Tharanian

Problem Statement

- What recipe can I make with my ingredients?
- Can we reverse the traditional cooking approach by searching for recipes based on list of ingredients instead of selecting the recipe and then creating a list of ingredients to buy them at the grocery?

Proposed Solution

Our project aims to address this idea via a recipe recommendation system.

1. Given any recipe dataset, build a model to find most relevant recipes that match the user’s ingredients list.
2. Perform unsupervised Topic Modeling on the recipes to group recipes into topics.
3. Rank the ingredients based on its frequency i.e. each successive ingredient in the list is weighted incrementally less.
4. Create a search algorithm that utilizes similarity scoring to rank recipes according to the greatest similarity to the user-input list of ingredients and returns recipe recommendations based on the scores.

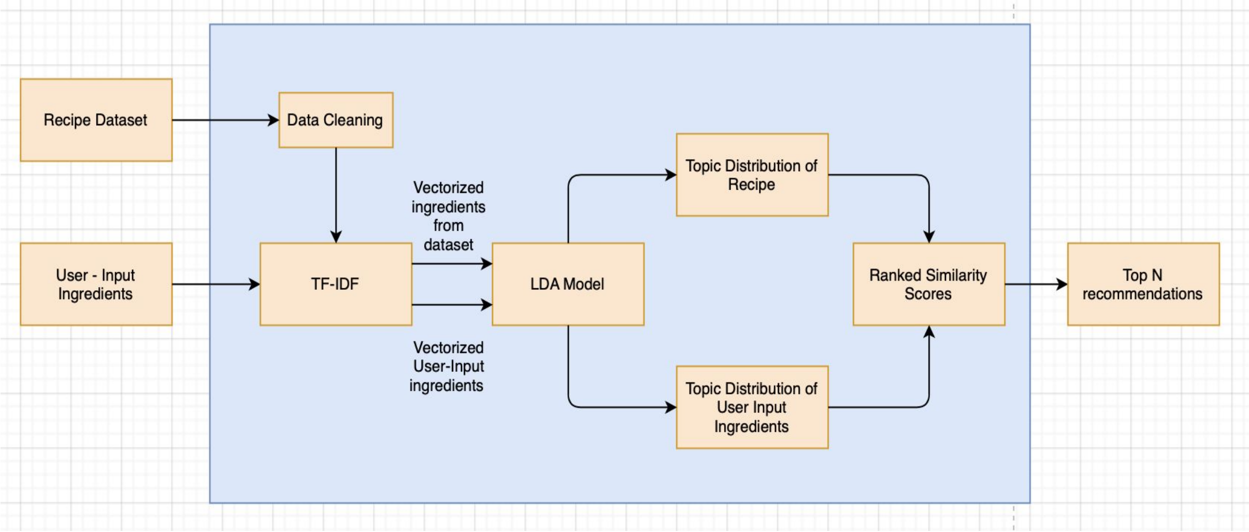
Target Users

- Anyone with an appetite to cook
    - Save time searching for recipes based on a list of ingredients
    - Find relevant and interesting recipes
- UX/UI Developer
    - Proto-type user friendly UI
    - Save time by using existing libraries to build out a web interface or an app
- Data analyst transitioned to a Data Scientist
    - To understand how the existing model works.
    - Learn by incorporating additional features to the existing libraries

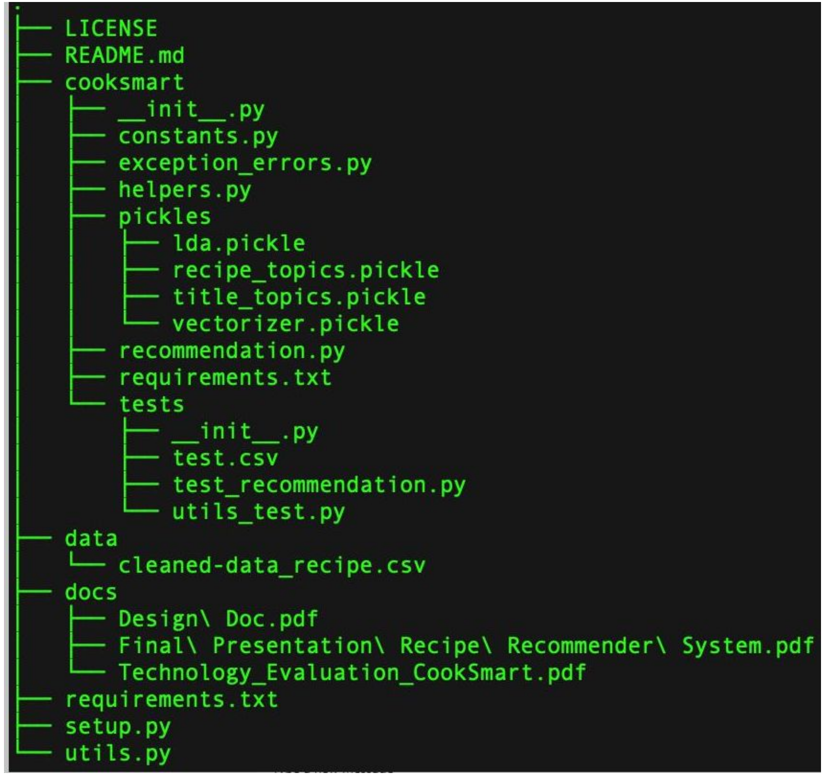
Recipe Dataset

1. The kaggle dataset[1] used for our recommendation system was created by scraping AllRecipes.com, a popular social network recipe site.
2. For each recipe, the dataset contains a corresponding row with recipe id, name, average ratings of reviewers, image url, review nums, ingredients, cooking directions, nutritions, and reviews.
3. Dataset includes 49,698 recipes with 38,131 ingredients
4. Data Features:
   - Recipe Name
   - Ingredients
   - Cooking Directions

Design


Recipe Dataset
Data Cleaning
User - Input Ingredients
TF-IDF
Vectorized ingredients from dataset
Vectorized User-Input ingredients
LDA Model
Topic Distribution of Recipe
Topic Distribution of User Input Ingredients
Ranked Similarity Scores
Top N recommendations

Directory Structure



```
.
├── LICENSE
├── README.md
├── cooksmart
│   ├── __init__.py
│   ├── constants.py
│   ├── exception_errors.py
│   ├── helpers.py
│   ├── pickles
│   │   ├── lda.pickle
│   │   ├── recipe_topics.pickle
│   │   ├── title_topics.pickle
│   │   └── vectorizer.pickle
│   ├── recommendation.py
│   ├── requirements.txt
│   └── tests
│       ├── __init__.py
│       ├── test.csv
│       ├── test_recommendation.py
│       └── utils_test.py
├── data
│   └── cleaned-data_recipe.csv
├── docs
│   ├── Design\ Doc.pdf
│   ├── Final\ Presentation\ Recipe\ Recommender\ System.pdf
│   └── Technology_Evaluation_CookSmart.pdf
├── requirements.txt
├── setup.py
└── utils.py
```

Demo

Learnings

1. Learned about the LDA algorithm for topic modeling (both scikit and gensim implementations)
2. Advantages of decoupling the fit and recommendation when constructing a model - avoiding having to fit the model on the given dataset every single time.
3. How to construct a Python package
4. Best software practices, i.e. collaborating with Git, documentation, component diagrams
5. Continuous integration using Travis

Future Enhancements

1. In its current form, the recommender system does not handle recommendations based on priority of the input ingredients which might produce undesired results which can be worked upon to make it better.
2. Expand the model to include additional features such as desserts, vegetarian, non-vegetarian, nutrition based.
3. Incorporate support for other topic model algorithms such as PCA or MALLET.
4. Include an UI to extend support for non-technical users.

References

[1] Dataset - https://www.kaggle.com/elisaxxygao/foodrecsysv1

[2] Scikit Learn LDA - https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html

THANK YOU!!