FoodFlix

Mentor: Hyesoo Youn

FoodFlix

A Food Recommender Engine

(Netflix for Food)

Team:

Kyle Bilton

Shudan Zhong

Javier Caravaca

Sudha Subramanian

Mentor: Hyesoo Youn

Introduction

- Objective
  - Build a food recommendation system from the ground up

- Why FoodFlix?
  - Make it easy for users to find healthy diet choices
  - Build weekly schedules or shopping lists

INPUT

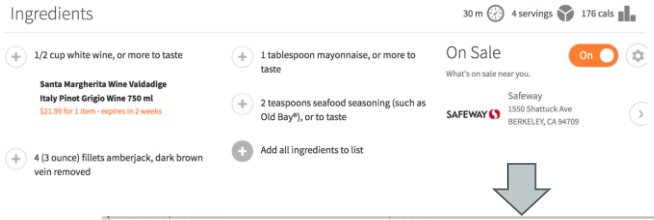
- User's height & weight
- Restrictions (eg. nut)
- User preferences (eg. pumpkin, milk)
- Feedback on recommended recipes (i.e., like or dislike)

OUTPUT

- Recommended Recipes

1. Obtaining data

- 18K recipes scraped from allrecipes.com using *Beautiful Soup*
  - Recipe with ingredients , Cooking time, Number of reviews, Rating, Nutrition information, Calories
- Data Cleaning
  - Removing Stop words, non-alphabetical character, recipes without ingredients, etc.

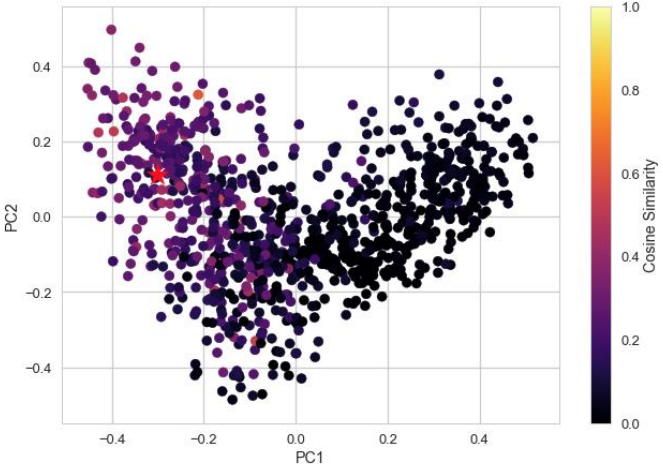


recipe_id	recipe_name	recipe_url	ingredients	cook_time	calorie_count	review_count	overall_rating
255000	baked-amberjack	https://www.allrecipes.com/recipe/255000/baked-amberjack/	['1/2 cup white wine, or more to taste', '4 (3 ounce) fillets amberjack, dark brown vein removed', '1 tablespoon mayonnaise, or more to taste', '2 teaspoons seafood seasoning (such as Old BayÂ®), or to taste']	30 m	176 cals	0 reviews	0

recipe_id	ingredients
237052	sweetened condensed milk heavy whipping cream ...
237363	rolled oats cashews shredded coconut dried mix...
231960	flour butter confectioners sugar walnuts confe...

2. Feature engineering

1. **TF-IDF**: term frequency inverse document frequency
2. **Word2vec**: average of ingredients' word vector
3. **Doc2vec:** train recipe vectors using their ingredients



2D Principal component weights for TF-IDF vectors

3. Model Development

1. **Find recipes with preferred ingredients using word search**
    a. Pumpkin → use recipes that contain pumpkin in ingredients
2. **Find recipes with calories within range set by BMI input**
    a. Height/Weight → BMI → Recommended calories range per meal → Filter recipes
3. **Remove recipes with restricted ingredients**
    a. Nut → Filter out recipes that contain nuts
4. **Find recipes similar to the 'liked' ones**, using cosine similarity of TF-IDF, or Word2Vec, or Doc2Vec
    a. Among the recipes that are filtered from (step 1, 2, 3), recommend more recipes that reflect user's personal preference shown by 'likes' and 'dislikes'

4. User interface development

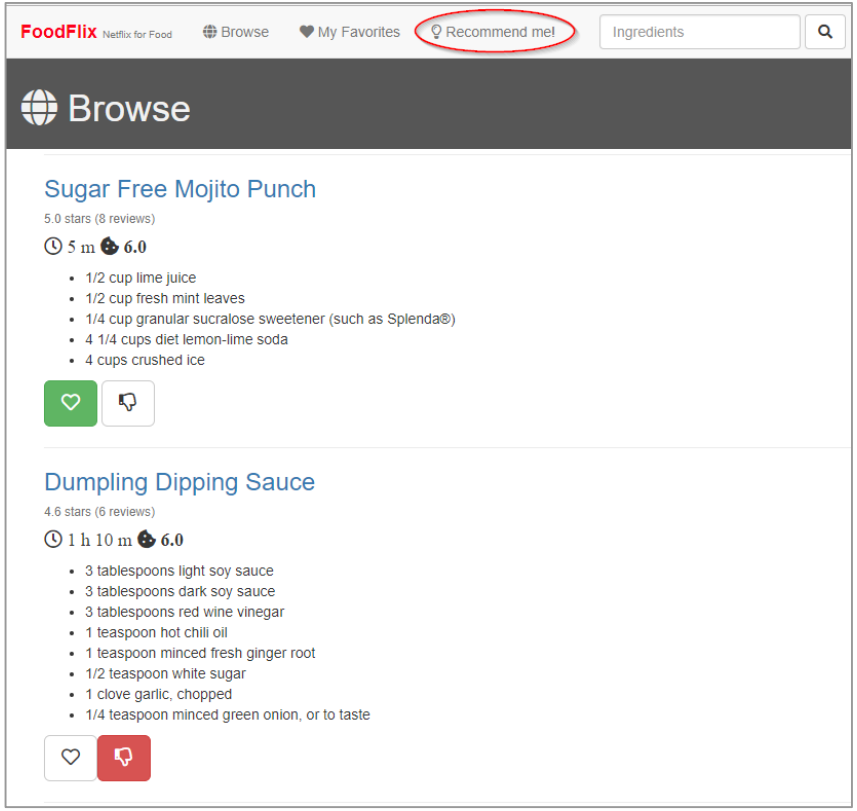
- Easy app/web development
- Individual sessions
- User Profile

- Compute BMI, Calories per day
- Model Training & Prediction

- Database to store recipe data
- Hold user account information

FoodFlix

https://foodflix-api-heroku.herokuapp.com/



- Login on website / validation
- Recipes ‘Liked’ or ‘Disliked’
- Invoke Recommender Engine
- Predict based on:
  - Likes / dislikes
  - Calories per day
- Provide top 3 recommendations

FoodFlix

LIVE DEMO

5. Conclusion

- Built a food recommendation system from data scraping to user-interface
- Evaluation results:
  - Blind test comparing FoodFlix results to randomly recommended recipes shows that it performs better than random guesses
  - Subjectively rate recommendations 0 - 5 based on how relevant they are

Mean recommendation rating (out of 5)

TF-IDF / Random	W2V / Random	D2V / Random
3.7 / 1.53	3.7 / 1.4	3.5 / 0.4

6. Potential improvements

- Perform more thorough and quantitative evaluation (A/B testing if possible)
- Ensemble of TF-IDF, W2V, and D2V may result in better recommendation
- Emphasis on healthy recommendations (eg. nutrition information)
- Classify recipes to use in filters and recommendations
- Recommend based on similar profiles (collaborative filtering)

Acknowledgements

- GDSO
- Hyesoo Youn
- Mentors

Any
Questions?

Graduate Data Science Organization

officers@gdso.berkeley.edu