Snackr: A swipe-based recipe recommendation app

Al Romero and Mary Deignan
Automations and Workflows
December 17, 2024

1 Introduction

Home cooks around the world are constantly faced with the question of what to cook. This decision quickly becomes overwhelming as cooks factor in ingredients, cooking time, nutrition goals, and personal preference. This project intends to make finding new recipes fun, not exhausting.

Our final product is a user-friendly interface designed to bring more joy to meal planning. Rather than searching online endlessly, users can simply swipe through a list of recipes. As users interact by swiping, our application's recommendation engine learns from their preferences to suggest recipes they're most likely to enjoy. The user can take a closer look at each recipe with an information page before making their decisions. When a user "likes" a recipe, it is saved to another page within the application, so users can return to recipes they have liked for future reference. This page contains all of the necessary information about the recipe including directions, ingredients, and nutrition information.

2 Data Description

The data obtained from this project was scraped from Allrecipes.com. The scraping code was adapted from Dmitry Zub's GitHub [3]. Additional information on how we scraped and prepared the data can be found in the Methods section. Our raw data can be seen in Figure 1.

3 Methods

3.1 scraper.py and Feature Modification

Data was obtained using our custom python script, `src/scraping/scraping.py`. The script operates in three main parts.

The link collection stage begins by accessing the Allrecipes A-Z categories page. Using the `get_category_links()` function, it extracts URLs for all recipe categories. For each category, `get_recipe_links_from_category()` navigates to collect all recipe URLs on the category page. These URLs are stored in `src/scraping/links/links.txt` as tab-separated values, pairing each recipe URL with its category URL.

In the data extraction stage, the `extract_recipe_data()` function processes each recipe URL to collect structured data in seven categories: `basic_info`, `prep_data`, `ingredients`, `nutritions`, `directions`, `image_filename`, and `category_url`. The `basic_info` column

contains recipe details including title, category, rating, and rating count. The `prep_data` column includes preparation-related information such as cooking time and servings.

In the storage and organization stage, recipe data is structured and saved to a JSON file. This file is stored in `scraping/recipes/scraped_recipes.json`. Recipe images are downloaded as PNG files to `scraping/images/`. Image filenames in the JSON file correspond directly to the saved image files, allowing easy matching between recipes and their images. The data without any modifications can be seen in Figure 1.

	basic_info	prep_data	ingredients	nutritions	directions	image_filename	category_url
0	{'title': 'Grilled Cheese Sloppy Joes', 'categ...	{'prep_time': '10 mins', 'cook_time': '1 hr', ...	[1 pound lean ground beef, 2 tablespoons dried...	{'calories': '585', 'fat': '39g', 'carbs': '32...	[Heat a large skillet over medium-high heat. C...	5410153-e9807f995fb74fe58bd2fcd56b541010.jpg	https://www.allrecipes.com/recipes/14930/main-...
1	{'title': 'Our 14 Best Empanada Recipes', 'cat...	{}	[]	{}	[]	5589810-3f53ac235a5346838b2ec079e34283e1.jpg	https://www.allrecipes.com/recipes/23131/appet...
2	{'title': 'Creamy Butternut Squash Pasta', 'ca...	{'prep_time': '15 mins', 'cook_time': '20 mins...	[1 (16 ounce) package ziti or other pasta, 1 t...	{'calories': '316', 'fat': '21g', 'carbs': '26...	[Bring a large pot of lightly salted water to ...	9408195-e2b738fc048f47bcb47c99f58890ba0b.jpg	https://www.allrecipes.com/recipes/550/pasta-a...
3	{'title': 'Golden Yam Brownies', 'category': '...	{'prep_time': '30 mins', 'cook_time': '30 mins...	[1 cup butter, 1 cup packed brown sugar, 1 cup...	{'calories': '221', 'fat': '10g', 'carbs': '32...	[Preheat the oven to 350 degrees F (175 degree...	1059053-70a5c73ef725427e9770ad083c423e87.jpg	https://www.allrecipes.com/recipes/2452/fruits...
4	{'title': 'Authentic Mexican Corn Bread', 'cat...	{'prep_time': '15 mins', 'cook_time': '45 mins...	[1 (14 ounce) can sweetened condensed milk (su...	{'calories': '448', 'fat': '14g', 'carbs': '74...	[Preheat the oven to 350 degrees F (175 degree...	9281485-authentic-mexican-corn-bread-Lisa-Lode...	https://www.allrecipes.com/recipes/342/bread/q...
...	...	...	...	...	...	...	...
19073	{'title': 'Hash Brown and Bacon Omelet Cups', ...	{'prep_time': '10 mins', 'cook_time': '35 mins...	[cooking spray, 3 cups frozen shredded hash br...	{'calories': '311', 'fat': '22g', 'carbs': '17...	[Preheat the oven to 425 degrees F (220 degree...	7567917-hash-brown-omelet-cups-KH-AR15501-4x3-...	https://www.allrecipes.com/recipes/205/meat-an...
19074	{'title': 'Strawberry Freezer Jam', 'category'...	{'prep_time': '10 mins', 'cook_time': '10 mins...	[4 cups halved fresh strawberries, 1 ½ cups gr...	{'calories': '41', 'carbs': '10g'}	[Gather all ingredients., Dotdash Meredith Foo...	N/A	https://www.allrecipes.com/recipes/1961/side-d...
19075	{'title': 'Chewy Chocolate Chip Oatmeal Cookie...	{'prep_time': '20 mins', 'cook_time': '12 mins...	[1 cup butter, softened, 1 cup packed light br...	{'calories': '145', 'fat': '8g', 'carbs': '17g...	[Gather the ingredients. Preheat the oven to 3...	N/A	https://www.allrecipes.com/recipes/851/dessert...
19076	{'title': 'Sweet Restaurant Slaw', 'category':...	{'prep_time': '15 mins', 'additional_time': '2...	[1 (16 ounce) bag coleslaw mix, 2 tablespoons ...	{'calories': '200', 'fat': '12g', 'carbs': '23...	[Gather the ingredients., Combine coleslaw mix...	N/A	https://www.allrecipes.com/recipes/96/salad/
19077	{'title': 'Easy Apple Cobbler', 'category': 'R...	{'prep_time': '10 mins', 'cook_time': '30 mins...	[3 tablespoons white sugar, 1 teaspoon ground ...	{'calories': '168', 'fat': '1g', 'carbs': '41g...	[Preheat the oven to 350 degrees F (175 degree...	9402881-827c42f75932475fb553990c838ddf83.jpg	https://www.allrecipes.com/recipes/361/dessert...

19078 rows × 7 columns

Figure 1: Raw Data

We separated these category groups so each feature has its own column (`prep_time` and `cook_time` have their own columns rather than being together in `prep_data`). We then removed any rows where there was no matching image so all recipes that appear have a corresponding image. We also removed all rows where any blank or null values appeared as well as their corresponding image. This was done to reduce the size of the data so this project can be stored on GitHub. After removing these rows, we reduced our dataset from 19,078 recipes to 6,360.

Our project goals required some feature modification in order to provide additional useful information about each recipe. We conducted three major feature modifications: **verb_count, ingredient_count, and yield_servings_merge**.

1. `verb_count` indicates the number of verbs in the `directions` column of a recipe. Since recipes sometimes hide a large number of steps in a single "step" designation, we used this to provide a more accurate picture of the number of steps necessary for a single recipe. This was determined using `nltk` part-of-speech tagging. This estimation is imperfect (due partially to the difference in sentence structure between recipes and normal sentences), but is close enough to be valuable.

2. `ingredient_count` indicates the number of items in the `ingredients` column of a recipe. This entry is formatted as a list of strings, so it simply counts the number of

strings in the list. This attribute helps to quantify the ingredients list, so users can tell from a glance if the recipe requires a large amount of ingredients without having to parse the list itself.

3. `yield_servings_merge` indicates the yield of the recipe from the column `yield`, if it exists. If the yield is not available, it indicates the servings for the recipe from the column `servings`. We thought that the `yield` column contained more objective data, without making assumptions about how much a serving is. However, this data was sometimes missing, so we decided to supplement it with data from the `servings` column, which is still useful, but more subjective. We then saved the data as `data/recipe_data.csv`. The featured dataset can be seen in Figure 2.

	ingredients	directions	image_filename	category_url	title	category	rating	rating_count	cook_time	total_time	yield	servings	calories	fat	carbs	protein	verb_count	ingredient_count	yield_servings_merge
0	[1 pound lean ground beef, 2 tablespoons dried...	[Heat a large skillet over medium-high heat. C...	5410153-e9807f995fb74fe58bd2fcd56b541010.jpg	https://www.allrecipes.com/recipes/14930/main-...	Grilled Cheese Sloppy Joes	Sloppy Joes	4.8	4	1 hr	1 hr 10 mins	8 sandwiches	8	585	39g	32g	27g	13	8	8
3	[1 cup butter, 1 cup packed brown sugar, 1 cup...	[Preheat the oven to 350 degrees F (175 degree...	1059053-70a5c73ef725427e9770ad083c423e87.jpg	https://www.allrecipes.com/recipes/2452/fruits...	Golden Yam Brownies	Yams	4.5	497	30 mins	1 hr	24 servings	24	221	10g	32g	2g	11	12	24
4	[1 (14 ounce) can sweetened condensed milk (su...	[Preheat the oven to 350 degrees F (175 degree...	9281485-authentic-mexican-corn-bread-Lisa-Lode...	https://www.allrecipes.com/recipes/342/bread/q...	Authentic Mexican Corn Bread	Cornbread	4.3	50	45 mins	1 hr	1 pan cornbread	12	448	14g	74g	11g	9	11	12
6	[6 (5-inch) corn tortillas, 3 cups chopped coo...	[Place a paper towel on a microwave-safe plate...	5414541-1e498b617cf84c479998b9903009a851.jpg	https://www.allrecipes.com/recipes/17874/main-...	Traditional Mexican Street Tacos	Tacos	4.4	9	0 mins	10 mins	6 tacos	2	697	30g	44g	64g	1	6	2
7	[1 ½ cups rolled oats, 1 ½ cups sifted pastry ...	[Preheat the oven to 350 degrees F (175 degree...	7368576-869e126fb14a489bbcb6df6fb3c827f6.jpg	https://www.allrecipes.com/recipes/841/holiday...	Gramma's Date Squares	Christmas Cookies	4.6	476	25 mins	50 mins	12 squares	12	363	13g	64g	4g	13	9	12
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
19061	[3 large eggs, 2 cups white sugar, 1 cup veget...	[Preheat the oven to 350 degrees F (175 degree...	459186-chocolate-zucchini-muffins-Pam-Ziegler-...	https://www.allrecipes.com/recipes/348/bread/q...	Chocolate Zucchini Muffins	Zucchini Bread	4.7	245	20 mins	35 mins	2 dozen muffins	24	219	10g	30g	3g	11	14	24
19066	[3 cups all-purpose flour, 1 teaspoon baking s...	[Preheat the oven to 350 degrees F (175 degree...	688015-3c19325f6a6d4f93b0f8928cf0e47b2e.jpg	https://www.allrecipes.com/recipes/348/bread/q...	Chocolate Chip Orange Zucchini Bread	Zucchini Bread	4.7	476	50 mins	1 hr 10 mins	2 9x5-inch loaves	20	339	18g	42g	4g	10	14	20
19071	[2 pounds ground chuck, 2 tablespoons	[Place the ground chuck in a large skillet,	2463212-tacos-de-matamoros-Allrecipes-Magazine...	https://www.allrecipes.com/recipes/17874/main-...	Tacos De Matamoros	Tacos	4.3	111	1 hr 30 mins	2 hrs	6 servings	6	629	39g	33g	38g	14	14	6

Figure 2: Modified Data

3.2 Recommendation system: `classifier.py`

The recommendation algorithm is based on a graph convolutional network, where the embeddings and the similarity between embeddings make up the graph. The goal of the recommendation algorithm is to make predictions while disrupting the user flow as little as possible. To do this, there are intervals in place to give the model time to train and make predictions.

The model trains on the first four selections by the user, then it begins predicting once the user has swiped ten times. Once predictions are made, a counter is reset in the interface, so that the model trains after the next four swipes and predicts after ten swipes. For efficiency, the model runs on separate QThreads from the interface for both training and prediction. QThreads allow the model to run while other actions within the interface occur. The model itself is composed of two classes: `RecipeDataClassification` and `TwoLayerGCN`.

`RecipeDataClassification` is the main class, which trains and makes predictions based on input data.

Training begins by creating BERT embeddings of each recipe's directions text, which captures both ingredients and cooking methods. For each recipe, these embeddings are averaged into a single vector for efficient comparison. A graph network is then constructed where nodes are recipes and edges represent recipe similarity, calculated using Euclidean distance between their embedding vectors. The TwoLayerGCN classifier, based on Kipf et al.'s 2017 work [1], processes this graph structure over 100 training epochs to learn patterns in user preferences. After training, the model is saved for later predictions.

For prediction, the system selects 20 random recipes and processes them through the same embedding pipeline. It converts directions to BERT embeddings and constructs a graph based on recipe similarities. The saved TwoLayerGCN then evaluates these recipes, outputting values between 0 and 1 representing the predicted likelihood of user preference for each recipe.

3.3 interface.py

The interface is the main route through which users can interact with the project. It was written using PyQt5, a package for designing apps with Python. The main goal we had when creating this interface was to build an easy-to-use and simple interface that users enjoyed swiping through. We modeled the interface design after popular swipe-based applications like Tinder, and created three major 'pages': the swiping interface, the 'liked recipes' page, and the 'further information' page with information about individual recipes. Users can swipe on recipes, view details about recipes they've 'liked' in the past, and view more information about the current recipe before they decide whether to swipe left or right. The swiping page has click-and-drag swiping as well as buttons for selecting user preferences, along with a button that provides more information about each recipe. The information button additionally provides many of the statistics calculated during feature modification. The Liked Recipes page provides more information about each recipe, including directions.

We wanted to implement a recommendation system within the interface that would adapt to the user's preferences without interfering with the user's experience. To do this, much of the recommendation system is run using QThreads, a PyQt5 class that encourages efficient resource management. This meant that we could run computationally expensive model training and prediction code without interfering with the user's experience. Essentially, the recommendation system is called based on the number of times the user has swiped. The model is first trained once there have been four swipes, then it makes its first predictions at ten swipes. Predictions are made on a random set of twenty recipes from the original dataframe `data/recipe_data.csv`. Once predictions have been successfully made, they are presented to the user in the order of most likely to like to least likely, and then the swipe counter is reset, so the process can continue.

4 Results

Our web scraping implementation successfully gathered a comprehensive dataset of 6,360 recipes from Allrecipes.com, which was used to create an interactive recipe swiping applica-

tion. The interface design focuses on presenting recipe information in a clear and user-friendly format.

Figure 3 shows the main interface page where users can view individual recipes. Each recipe is displayed with its image, title, rating, and total preparation time. The interface includes interaction buttons for accepting or rejecting recipes that can be used rather than the swiping feature. This page also shows an info button next to the recipe title. When this button is selected, a recipe details preview page comes up, as seen in Figure 4. This main page (Figure 3) also shows the "Liked Recipes" button which takes users to the liked recipes page, as shown in Figure 5. The liked recipes page allows users to easily view all recipes they have liked since they began using the application. Finally, users can select a given recipe title from the liked recipes list to view all relevant information needed to prepare the recipe, as seen in Figure 6. This recipe details page shows users all the information from the details preview page as well as the directions. Figure 7 shows a diagram laying out how the pages are connected.

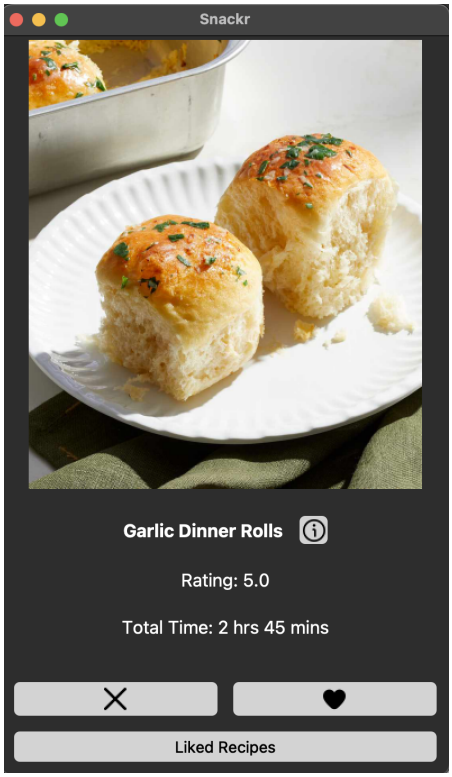

Figure 3: The main page of the interface. Users can swipe to like or dislike a recipe, click to display more information about the recipe, or open a list of their liked recipes.

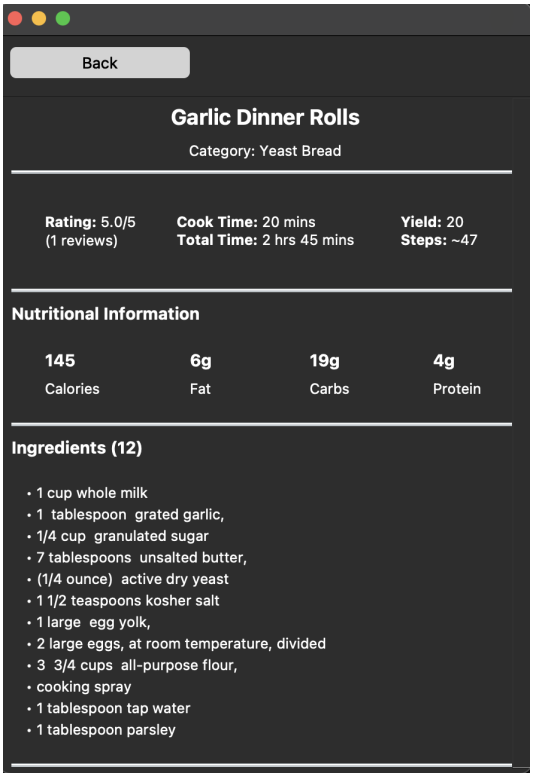

Figure 4: The recipe details preview page. Users can view ingredients, preparation time, and other key recipe information before deciding to like or dislike.

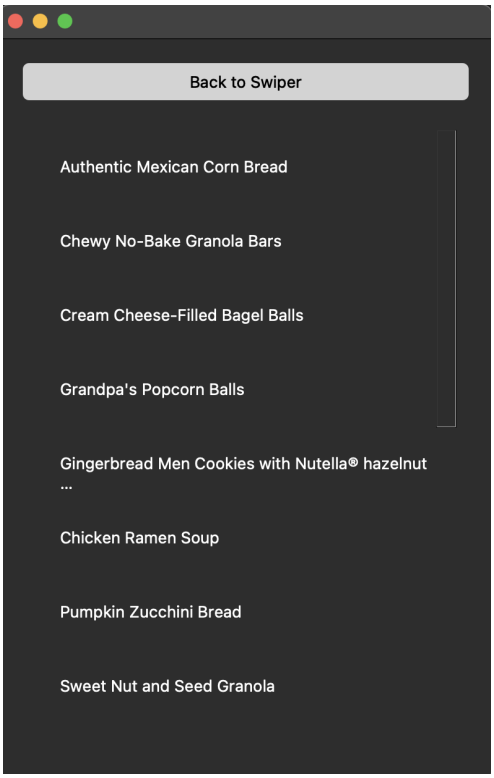

Figure 5: The liked recipes page showing a scrollable list of all recipes the user has liked. Users can click any recipe title to view its full details.

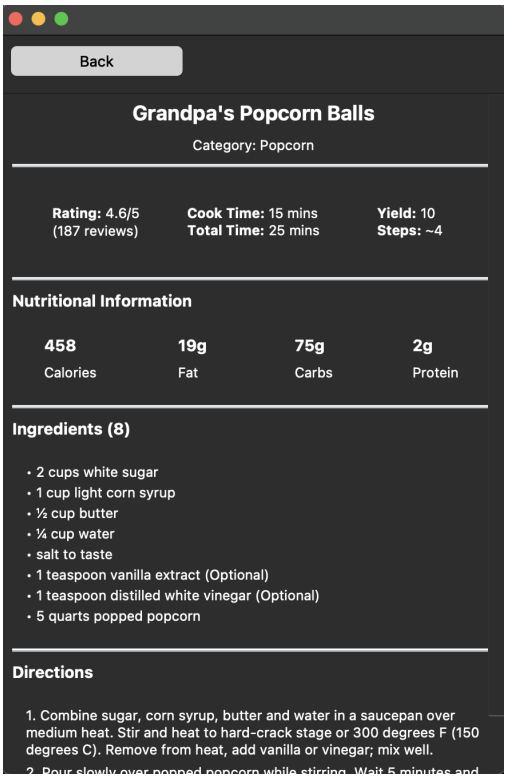

Figure 6: The complete recipe details page for liked recipes. Users can view all recipe information including full preparation directions and ingredient measurements.

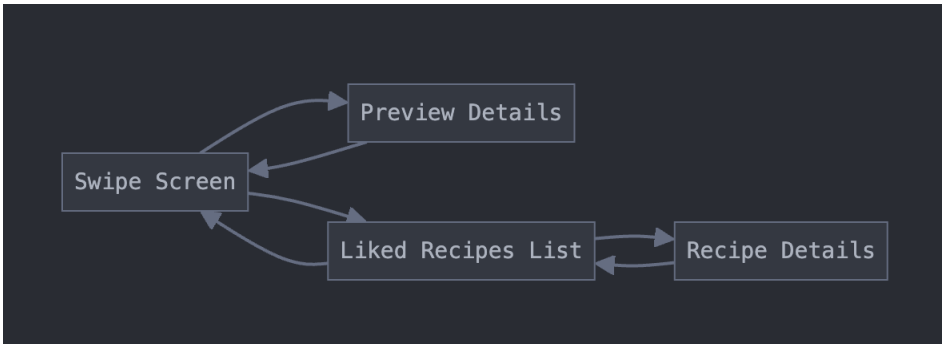

Figure 7: Diagram showing transitions between pages in the interface.

5 Future Work

Future work will focus on improving the recommendation system with more sophisticated algorithms, conducting more thorough stress testing of the interface, particularly around the

implementation of the recommender, and refining both the interface design and the code for improved performance. The recommender could be improved by using batch embedding to speed up training and prediction times, and a more sophisticated method of comparison between individual embeddings.

6 Contributions

Al has worked on feature engineering, built the recommender algorithm, implemented the recommender within the interface, and helped with the interface design and building. Mary has worked on scraping and documenting the scraping code, building the interface, and helping with the recommender algorithm.

References

[1] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

[2] Shikhar Vashishth. Gnns for nlp. `https://github.com/svjan5/GNNs-for-NLP`, 2019.

[3] Dimitry Zub. Allrecipes us recipes by state analysis. `https://github.com/dimitryzub/allrecipes-us-recipes-by-state-analysis`, 2022.