

Recipe Swiper

Al Romero and Mary Deignan
December 8, 2024

1 Project Description, Background, and Motivation

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

Our final product will be a user-friendly interface designed to bring more joy to meal planning. Users will have the option to input their dietary preferences, and then be able to swipe through a list of recipes within those preferences. As the user swipes, our product will update a classification system which attempts to predict the next recipes the user will like best.

When a user “likes” a recipe, it will be saved to another page within the application, so users can return to recipes they’ve liked for future reference.

2 Data Description

The data obtained from this project was scraped from Allrecipes.com. The scraping code was adapted from Dmitry Zub’s GitHub¹.

Data was obtained using our custom python script, `scraping.py`. This file collects recipe links from Allrecipes and saves them to a txt file in the `src/scraping/links` folder. It then extracts the recipe data, such as ingredients and directions, and saves that information to the `scraped_recipes.json` file in the `src/scraping/recipes` folder. The scraper also collects an image for each recipe, when available. These images are saved as .png files in the `src/scraping/images` folder. The images can be matched to the corresponding recipe with the `image_filename` column in the JSON file. The JSON file can then be uploaded to a pandas DataFrame and modified. The scraped data for each recipe has 7 different categories: `basic_info`, `prep_data`, `ingredients`, `nutritions`, `directions`, `image_filename`, and `category_url`. The data without any modifications can be seen in Figure 1.

¹Dmitry Zub, *allrecipes-us-recipes-by-state-analysis*, GitHub, <https://github.com/dimitryzub/allrecipes-us-recipes-by-state-analysis/tree/main>, accessed November 1, 2024.

	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...	5401053-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 1/2 cups gr...	{'calories': '41', 'fat': '10g', '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 x 7 columns

Figure 1: Raw Data

We separated these groups so each feature has its own column (`prep_time` and `cook_time` have their own columns). 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.

2.1 Feature Modification

Our project goals require some feature modification in order to provide useful information about each recipe. We've conducted three major feature modifications so far: **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 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.

Some of our other variables may require feature modification in the future. For example, the nutrition variables are currently stored as strings (for example, ‘3g’ of protein). These adjustments will be made if they’re deemed necessary in the future. For now, they are just options. We’re considering automating this part of our workflow, but the information is quite specific to the data we have scraped, so some may not be valuable for future datasets.

	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-e9807995b74fe58bd2fc56b541010.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...	1059053-70a5c73ef725427e9770a803c423cd7.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 ounces) can sweetened condensed milk (bu...	[Preheat the oven to 350 degrees F (175 degree...	9281485-authentic-mexican-com-bread-Use-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...	5414541-1e498b617c84c479998b9903009a851.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 1/2 cups rolled oats, 1 1/2 cups sifted pastry ...	[Preheat the oven to 350 degrees F (175 degree...	7368576-869e126b14a489bcbcbdf6b3c827f6.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 p...	[Preheat the oven to 350 degrees F (175 degree...	688015-3c19325f6a64093b08928c40e47b2e.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

We then saved the data to our data folder as `recipe_data.csv`. The featured dataset can be seen in Figure 2.

2.2 Interface Design

We decided to build an application mimicking the user-friendly, intuitive swiping system found in dating apps like Tinder. This would be useful, since many people already know how to interact with such apps. The interface currently works by swiping randomly through the entire dataset of recipes. This could be improved with pages for storing user preferences and liked recipes, alongside a small tutorial on the main page showing new users how to utilize the tool.

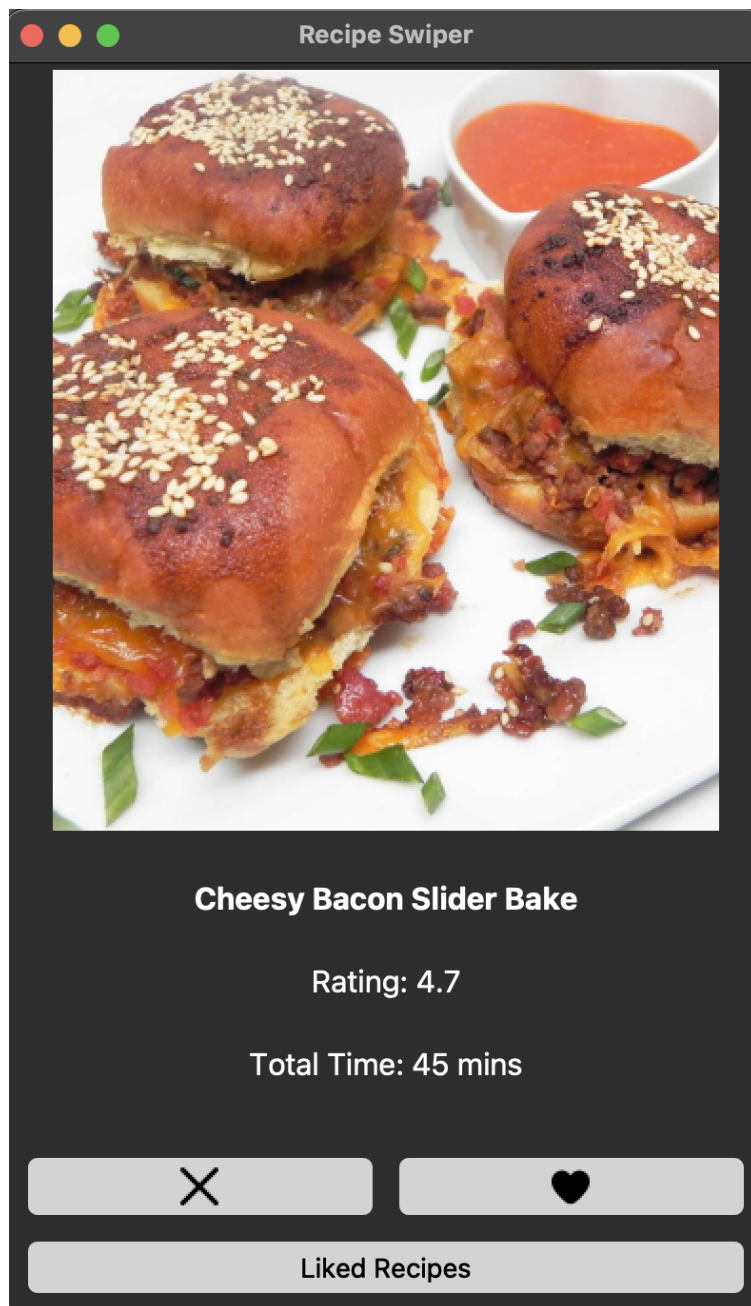


Figure 3: Interface Example

3 Progress and Next Steps

So far, we have scraped the data from Allrecipes, documented our scraping code, and worked on adding relevant features to our data. We have the basic “swiping” implementation completed, where each swipe stores the selected recipe as “liked” or “disliked”. We have also created a separate page where users can see all the recipes they have liked. Users can now

select a recipe from this list and view the title, rating, ingredients, and directions. We are in the process of properly formatting this information as well as deciding what additional information to display.

Our main priority is currently stress testing our interface to ensure the main functionality is working properly. We also would like to connect our interface to a recommender system that predicts the items the user is likely to “like” next. We are currently thinking of utilizing cosine similarity to compare the different recipes, but there are some challenges to implementation (the user is providing a list of preferences, which may not all represent the same type of recipe, and needs to connect to any specified preferences).

4 Group Work

Al has worked on the feature engineering as well as ideas for the recommender system and helping with the interface. Mary has worked on scraping and documenting the scraping code as well as building the interface.