

Keto/Vegan Diet classifier

Argmax, a consulting firm specializing in search and recommendation solutions with offices in New York and Israel, is hiring entry-level Data Scientists and Machine Learning Engineers.

At Argmax, we prioritize strong coding skills and a proactive, “get-things-done” attitude over a perfect resume. As part of our selection process, candidates are required to complete a coding task demonstrating their practical abilities.

In this task, you’ll work with a large recipe dataset sourced from Allrecipes.com. Your challenge will be to classify recipes based on their ingredients, accurately identifying keto (low-carb) and vegan (no animal products) dishes.

Successfully completing this assignment is a crucial step toward joining Argmax’s talented team.

```
In [1]: from opensearchpy import OpenSearch
        from decouple import config
        import pandas as pd

        client = OpenSearch(
            hosts=[config('OPENSEARCH_URL', 'http://localhost:9200')],
            http_auth=None,
            use_ssl=False,
            verify_certs=False,
            ssl_show_warn=False,
        )
```

Recipes Index

Our data is stored in OpenSearch, and you can query it using either Elasticsearch syntax or SQL.

Elasticsearch Syntax

```
In [2]: query = {
        "query": {
            "match": {
                "description": { "query": "egg" }
            }
        }
    }

    res = client.search(
        index="recipes",
        body=query,
        size=2
    )
```

```
hits = res['hits']['hits']
hits
```

```
Out[2]: [{ '_index': 'recipes',
  '_id': 'KBehQJcBmKMcD7RGUg3P',
  '_score': 3.9817066,
  '_source': {'title': 'Genuine Egg Noodles',
    'description': 'These egg noodles are the original egg noodles. ',
    'instructions': ['Combine flour, salt and baking powder. Mix in eggs and enough wa
ter to make the dough workable. Knead dough until stiff. Roll into ball and cut into
quarters. Using 1/4 of the dough at a time, roll flat to about 1/8 inch use flour as
needed, top and bottom, to prevent sticking. Peel up and roll from one end to the oth
er. Cut roll into 3/8 inch strips. Noodles should be about 4 to 5 inches long dependi
ng on how thin it was originally flattened. Let dry for 1 to 3 hours.',
    'Cook like any pasta or, instead of drying first cook it fresh but make sure wate
r is boiling and do not allow to stick. It takes practice to do this right.'],
    'ingredients': ['2 cups Durum wheat flour',
      '1/2 teaspoon salt',
      '1/4 teaspoon baking powder',
      '3 eggs',
      'water as needed'],
    'photo_url': 'http://images.media-allrecipes.com/userphotos/560x315/966604.jpg'}},
  {'_index': 'recipes',
    '_id': 'Ph2hQJcBmKMcD7RGpSI4',
    '_score': 3.6793466,
    '_source': {'title': 'Eggs Creole Over Toast',
      'description': 'Tasty egg scramble.',
      'instructions': ['Melt butter in a skillet over medium-high heat. Cook and stir mu
shrooms, onion, and pepper in melted butter until soft, about 5 minutes; add tomatoes
and capers and bring to a simmer.',
      'Pour beaten eggs over vegetable mixture; cook and stir until eggs are scrambled
and no longer runny, about 5 minutes. Spoon egg mixture over toast.'],
      'ingredients': ['1 tablespoon butter',
        '1/4 cup chopped mushrooms',
        '1 tablespoon chopped onion',
        '1 tablespoon chopped green bell pepper',
        '1 cup crushed tomatoes',
        '1 tablespoon capers',
        '6 eggs, beaten',
        '4 slices bread, toasted'],
      'photo_url': 'http://images.media-allrecipes.com/userphotos/560x315/2579003.jp
g'}}}]
```

SQL syntax

```
In [3]: query = """
SELECT *
FROM recipes
WHERE description like '%egg%'
LIMIT 10
"""

res = client.sql.query(body={'query': query})
df = pd.DataFrame(res["datarows"], columns=[c["name"] for c in res["schema"]])
df
```

Out[3]:

	description	ingredients	instructions	photo_url	title
0	A delicious stuffed chicken recipe that uses a...	[1 tablespoon vegetable oil, 1/2 onion, finely...	[Lightly oil grill and preheat to medium high...	http://images.media-allrecipes.com/userphotos/...	Stuffed Chicken
1	A delicious and pretty veggie dish to serve at...	[3 white onions, sliced to 1/4 inch thickness,...	[Lightly oil grill and preheat to high., Place...	http://images.media-allrecipes.com/userphotos/...	Yum Yum Veggie Foils
2	While you can get some smoked flavor by adding...	[1 eggplant, sliced into 1/2 inch rounds, 2 re...	[Brush vegetables with oil to coat., Prepare s...	http://images.media-allrecipes.com/userphotos/...	Smoky Grilled Vegetables
3	A healthy way to grill veggies! Makes a great ...	[1/2 cup thickly sliced zucchini, 1/2 cup slic...	[Place the zucchini, red bell pepper, yellow b...	http://images.media-allrecipes.com/userphotos/...	Marinated Veggies
4	These little parcels are made with seasoned gr...	[3 tablespoons olive oil, 1 pound ground beef,...	[In a large, heavy saute pan, heat olive oil o...	http://images.media-allrecipes.com/userphotos/...	Puerto Rican Meat Patties
5	Stroganoff is a wonderful dish for guests; it ...	[1/4 cup butter, 1 1/2 pounds sirloin tip, cut...	[Melt butter in a large skillet over medium he...	http://images.media-allrecipes.com/userphotos/...	Beef Stroganoff I
6	This is a recipe my Aunt always makes. It has...	[1/4 cup shortening, 2 pounds lean beef chuck,...	[Melt shortening in large skillet over medium ...	http://images.media-allrecipes.com/userphotos/...	Beef Paprika
7	Black beans and rice are the logical accompani...	[2 teaspoons cumin seeds, 1/2 teaspoon whole b...	[Heat a small, heavy skillet over medium heat....	http://images.media-allrecipes.com/userphotos/...	Cuban Pork Roast I
8	These are fabulous as a side dish with stir-fr...	[1 pound ground pork, 1 teaspoon ground ginger...	[Season pork with ginger and garlic powder and...	http://images.media-allrecipes.com/userphotos/...	Best Egg Rolls
9	Pork cubes braised in a tangy honey, ginger an...	[2 pounds boneless pork loin, cubed, 2 tablesp...	[In a large skillet heat oil and brown pork cu...	http://images.media-allrecipes.com/userphotos/...	Honey Pork Oriental

Task Instructions

Your goal is to implement two classifiers:

1. Vegan Meal Classifier
2. Keto Meal Classifier

Unlike typical supervised machine learning tasks, the labels are not provided in the dataset. Instead, you will rely on clear and verifiable definitions to classify each meal based on its

ingredients.

Definitions:

1. **Vegan Meal:** Contains no animal products whatsoever (no eggs, milk, meat, etc.).
2. **Keto Meal:** Contains no ingredients with more than 10g of carbohydrates per 100g serving.
For example, eggs are keto-friendly, while apples are not.

Note that some meals may meet both vegan and keto criteria (e.g., meals containing avocados), though most meals typically fall into neither category.

Example heuristic:

```
In [4]: def is_ingredient_vegan(ing):  
        for animal_product in "egg meat milk butter veel lamb beef chicken sausage".split():  
            if animal_product in ing:  
                return False  
        return True  
  
def is_vegan_example(ingredients):  
    return all(map(is_ingredient_vegan, ingredients))  
  
df["vegan"] = df["ingredients"].apply(is_vegan_example)  
df
```

Out[4]:

	description	ingredients	instructions	photo_url	title	vegan
0	A delicious stuffed chicken recipe that uses a...	[1 tablespoon vegetable oil, 1/2 onion, finely...	[Lightly oil grill and preheat to medium high....	http://images.media-allrecipes.com/userphotos/...	Stuffed Chicken	False
1	A delicious and pretty veggie dish to serve at...	[3 white onions, sliced to 1/4 inch thickness,...	[Lightly oil grill and preheat to high., Place...	http://images.media-allrecipes.com/userphotos/...	Yum Yum Veggie Foils	False
2	While you can get some smoked flavor by adding...	[1 eggplant, sliced into 1/2 inch rounds, 2 re...	[Brush vegetables with oil to coat., Prepare s...	http://images.media-allrecipes.com/userphotos/...	Smoky Grilled Vegetables	False
3	A healthy way to grill veggies! Makes a great ...	[1/2 cup thickly sliced zucchini, 1/2 cup slic...	[Place the zucchini, red bell pepper, yellow b...	http://images.media-allrecipes.com/userphotos/...	Marinated Veggies	True
4	These little parcels are made with seasoned gr...	[3 tablespoons olive oil, 1 pound ground beef,...	[In a large, heavy saute pan, heat olive oil o...	http://images.media-allrecipes.com/userphotos/...	Puerto Rican Meat Patties	False
5	Stroganoff is a wonderful dish for guests; it ...	[1/4 cup butter, 1 1/2 pounds sirloin tip, cut...	[Melt butter in a large skillet over medium he...	http://images.media-allrecipes.com/userphotos/...	Beef Stroganoff I	False
6	This is a recipe my Aunt always makes. It has...	[1/4 cup shortening, 2 pounds lean beef chuck,...	[Melt shortening in large skillet over medium ...	http://images.media-allrecipes.com/userphotos/...	Beef Paprika	False
7	Black beans and rice are the logical accompani...	[2 teaspoons cumin seeds, 1/2 teaspoon whole b...	[Heat a small, heavy skillet over medium heat....	http://images.media-allrecipes.com/userphotos/...	Cuban Pork Roast I	True
8	These are fabulous as a side dish with stir-fr...	[1 pound ground pork, 1 teaspoon ground ginger...	[Season pork with ginger and garlic powder and...	http://images.media-allrecipes.com/userphotos/...	Best Egg Rolls	False
9	Pork cubes braised in a tangy honey, ginger an...	[2 pounds boneless pork loin, cubed, 2 tablesp...	[In a large skillet heat oil and brown pork cu...	http://images.media-allrecipes.com/userphotos/...	Honey Pork Oriental	True

Limitations of the Simplistic Heuristic

The heuristic described above is straightforward but can lead to numerous false positives and negatives due to its reliance on keyword matching. Common examples of incorrect classifications include:

- "Peanut butter" being misclassified as non-vegan, as "butter" is incorrectly assumed to imply dairy.
- "eggless" recipes being misclassified as non-vegan, due to the substring "egg."
- Animal-derived ingredients such as "pork" and "bacon" being incorrectly identified as vegan, as they may not be explicitly listed in the keyword set.

Submission

1. Implement Diet Classifiers

Complete the two classifier functions in the `diet_classifiers.py` file within this repository. Ensure your implementation correctly identifies "keto" and "vegan" meals. After implementing these functions, verify that the Flask server displays the appropriate badges ("keto" and "vegan") next to the corresponding recipes.

Note

This repo contains two `diet_classifiers.py` files:

1. One in this folder (`nb/src/diet_classifiers.py`)
2. One in the Flask web app folder (`web/src/diet_classifiers.py`)

You can develop your solution here in the notebook environment, but to apply your solution to the Flask app you will need to copy your implementation into the `diet_classifiers.py` file in the Flask folder!!!

```
In [5]: def is_ingredient_keto(ingredient):  
        # TODO: complete  
        return False  
  
        def is_ingredient_vegan(ingredient):  
            # TODO: complete  
            return False
```

For your convenience, you can sanity check your solution on a subset of labeled recipes by running `diet_classifiers.py`

```
In [ ]: ! python diet_classifiers.py --ground_truth /usr/src/data/ground_truth_sample.csv
```

2. Repository Setup

Create a **private** GitHub repository for your solution, and invite the GitHub user `argmax2025` as a collaborator. **Do not** share your implementation using a **forked** repository.

3. Application Form

Once you've completed the implementation and shared your private GitHub repository with argmax2025, please fill out the appropriate application form:

1. [US Application Form](#)
2. [IL Application Form](#)

Your application will not be considered complete until this form is submitted.

Evaluation process

Your submission will be assessed based on the following criteria:

1. **Readability & Logic** – Clearly explain your approach, including your reasoning and any assumptions. If you relied on external resources (e.g., ingredient databases, nutrition datasets), be sure to cite them.
2. **Executability** – Your code should run as is when cloned from your GitHub repository. Ensure that all paths are relative, syntax is correct, and no manual setup is required.
3. **Accuracy** – Your classifiers will be evaluated against a holdout set of 20,000 recipes with verified labels. Performance will be compared to the ground truth. data.

Next steps

If your submission passes the initial review, you'll be invited to a 3-hour live coding interview, where you'll be asked to extend and adapt your solution in real time.

Please make sure you join from a quiet environment and have access to a Python-ready workstation capable of running your submitted project.

In []: