



# Information Retrieval and Text Mining Project

Ingredient Based Recipe Finder

Team 4 - Namitha Bhaskar, Dhairya Dutt, Vaishakh Dineshan



# Overview

**Text Based Information Retrieval system.**

**What is a Text Based Information Retrieval system?**

- A text-based information retrieval system is a software application designed to search and retrieve specific information from a collection of textual documents.
- These systems utilize techniques such as indexing, querying, and ranking to efficiently locate and present relevant documents based on user input.

In short - Retrieval is based on the text i.e ingredients given by the user.

# Data

- Data for this project has been taken from recipeNLG, a novel dataset of cooking recipes.
- The dataset has columns like recipe name, ingredients used, directions to
- make the recipe and recipe url which takes us to the webpage
- The dataset has a total of about 2 million records.

```
print(recipe_df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2231142 entries, 0 to 2231141
Data columns (total 4 columns):
#   Column      Dtype
---  -
0    title      object
1  ingredients  object
2  directions  object
3    link      object
dtypes: object(4)
```

```
print(recipe_df.shape)
```

```
(2231142, 4)
```

Vocabulary Size: 3785

```
print(recipe_df.head(5))
```

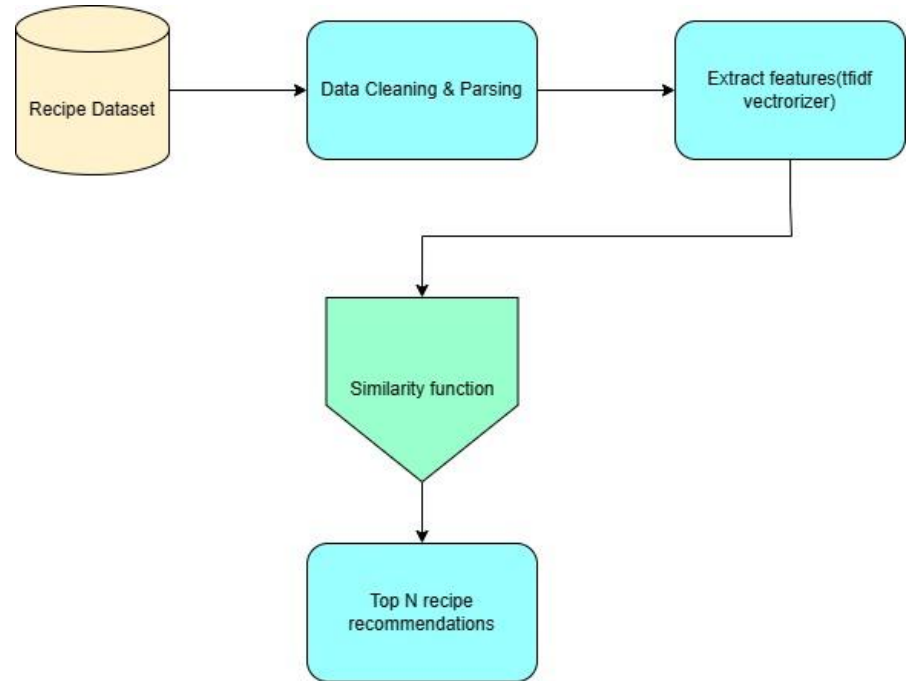
	title	ingredients \
0	No-Bake Nut Cookies	["1 c. firmly packed brown sugar", "1/2 c. eva...
1	Jewell Ball'S Chicken	["1 small jar chipped beef, cut up", "4 boned ...
2	Creamy Corn	["2 (16 oz.) pkg. frozen corn", "1 (8 oz.) pkg...
3	Chicken Funny	["1 large whole chicken", "2 (10 1/2 oz.) cans...
4	Reeses Cups(Candy)	["1 c. peanut butter", "3/4 c. graham cracker ...

	directions \
0	["In a heavy 2-quart saucepan, mix brown sugar...
1	["Place chipped beef on bottom of baking dish....
2	["In a slow cooker, combine all ingredients. C...
3	["Boil and debone chicken.", "Put bite size pi...
4	["Combine first four ingredients and press in ...

	link
0	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=44874">www.cookbooks.com/Recipe-Details.aspx?id=44874</a>
1	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=699419">www.cookbooks.com/Recipe-Details.aspx?id=699419</a>
2	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=10570">www.cookbooks.com/Recipe-Details.aspx?id=10570</a>
3	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=897570">www.cookbooks.com/Recipe-Details.aspx?id=897570</a>
4	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=659239">www.cookbooks.com/Recipe-Details.aspx?id=659239</a>

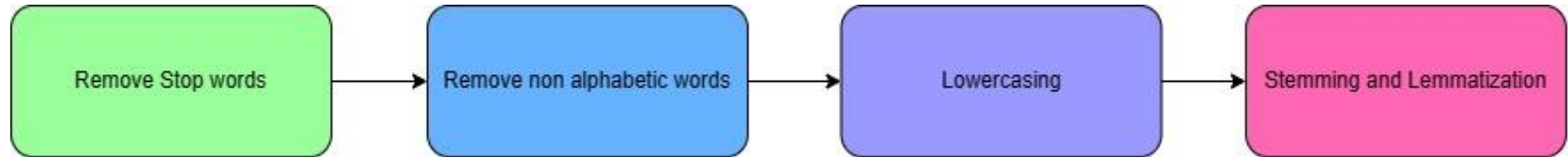
# Methodology

- The Top N recipes are Top 5 recipes for the ingredients given as input from the user.



# Preprocessing

- Initial step is analysing the ingredients and formulate a list of possible words and ingredients that would be treated as stop words.
- Followed by preprocessing the ingredients column present in the dataset.
- Preprocessing includes normalisation, tokenization of words, removing punctuations and numbers followed by stop words removal.



# Feature Extraction : TF-IDF Vectorizer



```
print(tfidf_recipe)
```

```
(0, 1390)    0.15808714433794496
(0, 1157)    0.35713630777565214
(0, 880)     0.2149288405263349
(0, 834)     0.27129741090029275
(0, 530)     0.4819292100827813
(0, 498)     0.43739889696886863
(0, 187)     0.21432712481669183
(0, 115)     0.5080902929396166
(1, 1524)    0.313123731888858
(1, 1313)    0.2227126407409637
(1, 914)     0.2506159505420913
(1, 359)     0.332864214664976
(1, 262)     0.49060934416238133
(1, 253)     0.22192345975252448
(1, 225)     0.34752752324143626
(1, 159)     0.31413116263985646
(1, 137)     0.41151470300800524
(2, 359)     1.0
```

- Converting raw data into meaningful features for analysis or modeling
- Transforming and selecting relevant information from the data
- Reducing dimensionality of the data
- Enhancing the performance of algorithms, especially in classification and retrieval tasks.

Feature extraction is done for each ingredient in each recipe.

# Cosine Similarity

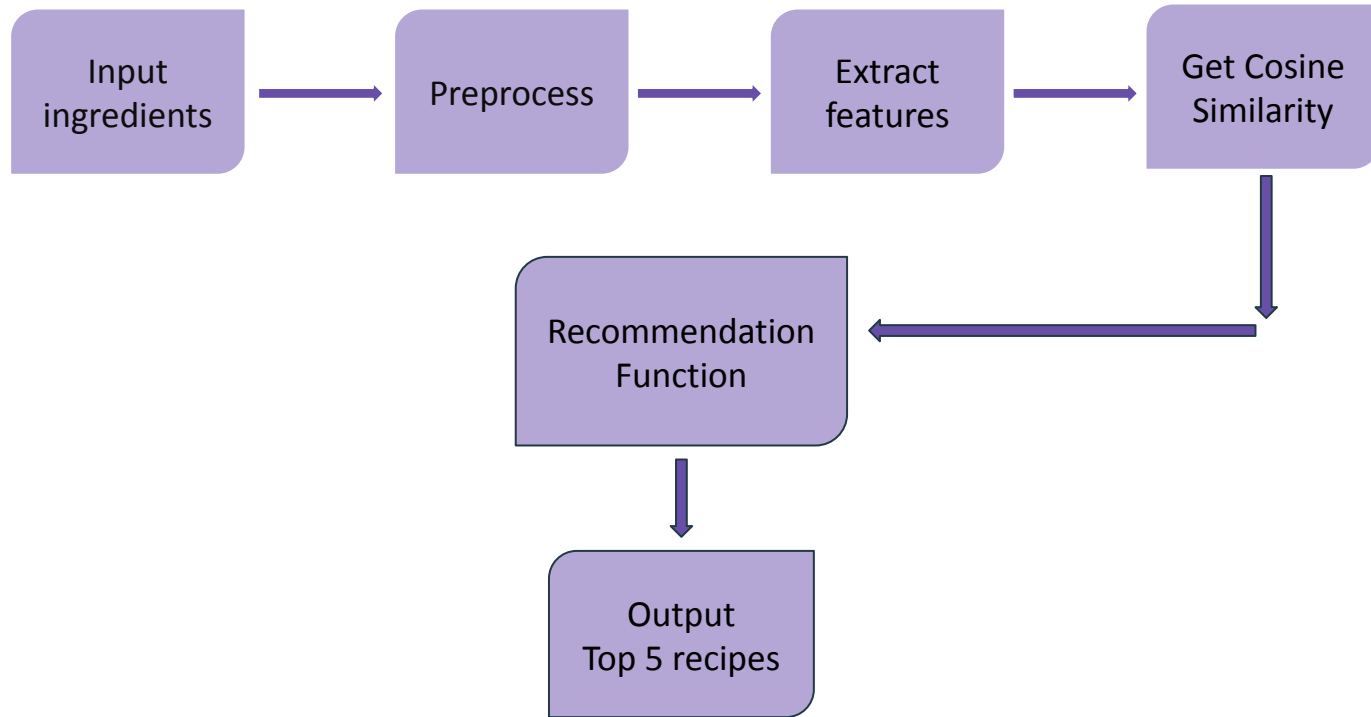
Used to get the similarity for vectors obtained from TF-IDF Vectorizer.

Cosine similarity provides a geometric measure of similarity that is robust to the magnitude of vectors.

## Why Cosine Similarity?

- Invariant to vector magnitude
- Efficient for sparse data
- Emphasis on content similarity
- Geometric interpretability
- Widely used in text mining
- Robustness to outliers and noise

# Top N Recommendations





# Results

## Top 5 recipes

```
# test ingredients
test_ingredients1 = "ground beef, pasta, spaghetti, tomato pasta sauce, bacon, onion, zucchini, and, cheese"
```

## Output

	recipe \	score
0	Spaghetti Pasta Sauce Recipe - Allrecipes.com	0 0.885
1	Zucchini and Shells Recipe - Allrecipes.com	1 0.871
2	Ricotta Stuffed Squash Recipe - Allrecipes.com	2 0.737
3	Meatball Casserole Recipe - Allrecipes.com	3 0.654
4	Stringy Macaroni Recipe - Allrecipes.com	4 0.646

# Results

## Text to Image

Cleaned recipe names to get image

Spaghetti Pasta Sauce  
Zucchini and Shells  
Ricotta Stuffed Squash  
Meatball Casserole  
Stringy Macaroni

```
from diffusers import StableDiffusionPipeline
import torch

model_id = "dreamlike-art/dreamlike-photoreal-2.0"
pipe = StableDiffusionPipeline.from_pretrained(model_id, torch_dtype=torch.float16)

import torch
from diffusers import StableDiffusionPipeline

# Load the pre-trained model
model_id = "runwayml/stable-diffusion-v1-5"
pipe = StableDiffusionPipeline.from_pretrained(model_id)
```

For Zucchini and Shells



For Meatball Casserole



For Spaghetti Pasta Sauce



For Ricotta Stuffed Squash



For stringy Pasta



```
# test ingredients
test_ingredients = "cherry , ale , jell o cherry flavor gelatin , almond , miniature marshmallow"
```

	recipe	ingredients \
0	Double Cherry Delight	cherry,ale,jell o cherry flavor gelatin,almond...
1	Cherry Jello Salad	box cherry jell o,cherry pie filling,cream che...
2	Cherry Salad	cherry,box cherry jell o,sugar,pineapple,pecan
3	Cranberry Punch	cranberry,ale,cherry jell o,lemonade
4	Penny Punch	cherry flavored gelatin,lemon flavored gelatin...

	score	url
0	1.000	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=703381">www.cookbooks.com/Recipe-Details.aspx?id=703381</a>
1	0.517	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=662334">www.cookbooks.com/Recipe-Details.aspx?id=662334</a>
2	0.508	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=677459">www.cookbooks.com/Recipe-Details.aspx?id=677459</a>
3	0.495	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=321508">www.cookbooks.com/Recipe-Details.aspx?id=321508</a>
4	0.454	<a href="http://www.cookbooks.com/Recipe-Details.aspx?id=138148">www.cookbooks.com/Recipe-Details.aspx?id=138148</a>

```
# test ingredients
test_ingredients1 = "ground beef, pasta, spaghetti, tomato pasta sauce, bacon, onion, zucchini, and, cheese"
test_ingredients2 = "shallot, chicken thigh, sesame, soy, rice noodle, tofu, onion, spinach, seaweed nori sheet, lime"
recs = RecSys(test_ingredients1)

pd.set_option('display.max_colwidth', None)
print(recs)
```

```

                                recipe \
0  Spaghetti Pasta Sauce Recipe - Allrecipes.com
1  Zucchini and Shells Recipe - Allrecipes.com
2  Ricotta Stuffed Squash Recipe - Allrecipes.com
3  Meatball Casserole Recipe - Allrecipes.com
4  Stringy Macaroni Recipe - Allrecipes.com
```

```
ingredients \
```

```

0
1 (28 ounce) jar prepared pasta sauce,1 onion, chopped,1 teaspoon olive oil,1/2 cup prepared mincemeat, crumbled, or to taste,1/2 cup tomato puree, or to taste (optional),5 leaves fresh basil, chopped
1
1 (15 ounce) can tomato sauce,1 (8 ounce) package uncooked pasta shells,1 large zucchini, peeled and cubed,1 teaspoon dried basil,1 teaspoon dried oregano,1/2 cup white sugar,1/2 medium onion, finely chopped,1/4 cup grated Romano cheese,1/4 cup olive oil,1/8 teaspoon crushed red pepper flakes,2 cups water,3 cloves garlic, minced
2
                                1 (10 ounce) package frozen chopped spinach, thawed and drained,1 (24 ounce) jar tomato and basil pasta sauce,1 clove garlic, minced,1 small onion, finely chopped,1 tablespoon chopped parsley,1/2 teaspoon dried basil,1/2 teaspoon dried oregano,1/2 teaspoon salt,1/4 cup grated Parmesan cheese,2 cups ricotta cheese,2 eggs, beaten,2 tablespoons butter, melted,8 zucchini, halved lengthwise,ground black pepper to taste
3 1 (1/2 ounce) refrigerated Italian-style cooked turkey meatballs, halved,1 (15 ounce) container ricotta cheese,1 (16 ounce) package penne pasta,1 cup chopped roasted red peppers,1 cup grated Parmesan cheese,1 teaspoon kosher salt,1/2 teaspoon fennel seeds,1/4 cup tomato paste,2 (24 ounce) jars tomato-basil pasta sauce,2 cups shredded mozzarella cheese,2 large eggs, lightly beaten,2 tablespoons olive oil,3 cloves garlic, minced,3/4 cup chopped fresh basil, divided,cooking spray
4
1 (16 ounce) can canned tomatoes,1 (16 ounce) package cottage cheese,1 (8 ounce) can tomato sauce,1 (8 ounce) package medium seashell pasta,1 1/4 teaspoons dried oregano,1 small onion, minced,1/2 cup grated Parmesan cheese,1/2 pound mozzarella cheese, shredded,1/2 teaspoon salt,1/4 teaspoon ground black pepper
```

# ChatGPT Verification

```
# test ingredients
test_ingredients1 = "ground beef, pasta, spaghetti, tomato pasta sauce, bacon, onion, zucchini, and, cheese"
```

Spaghetti Pasta Sauce  
Zucchini and Shells  
Ricotta Stuffed Squash  
Meatball Casserole  
Stringy Macaroni



ChatGPT

Here's a basic recipe for spaghetti pasta sauce:

Ingredients:

- 2 tablespoons olive oil
- 1 onion, finely chopped
- 2 cloves garlic, minced
- 1 can (28 ounces) crushed tomatoes
- 1 can (6 ounces) tomato paste
- 1 teaspoon dried oregano
- 1 teaspoon dried basil
- 1/2 teaspoon salt, or to taste
- 1/4 teaspoon black pepper, or to taste
- 1/4 teaspoon crushed red pepper flakes (optional, for heat)
- 1 tablespoon sugar (optional, to balance acidity)
- Fresh basil leaves, chopped (for garnish, optional)



ChatGPT

Here's a recipe for Zucchini and Shells pasta:

Ingredients:

- 8 ounces medium pasta shells
- 2 medium zucchinis, diced
- 2 tablespoons olive oil
- 3 cloves garlic, minced
- 1/4 teaspoon red pepper flakes (optional, for heat)
- Salt, to taste
- Black pepper, to taste
- 1/4 cup grated Parmesan cheese, plus more for serving
- 2 tablespoons chopped fresh parsley, for garnish (optional)

```
# test ingredients
test_ingredients1 = "ground beef, pasta, spaghetti, tomato pasta sauce, bacon, onion, zucchini, and, cheese"
```

Here's a recipe for Meatball Casserole:

Ingredients:

- 1 pound ground beef
- 1/2 cup breadcrumbs
- 1/4 cup grated Parmesan cheese
- 1/4 cup chopped fresh parsley
- 1 egg
- 2 cloves garlic, minced
- 1/2 teaspoon salt
- 1/4 teaspoon black pepper
- 1 tablespoon olive oil
- 1 onion, chopped
- 2 cloves garlic, minced
- 1 can (28 ounces) crushed tomatoes
- 1 teaspoon dried oregano
- 1 teaspoon dried basil
- Salt, to taste
- Black pepper, to taste
- 8 ounces cooked pasta (such as penne or spaghetti)



ChatGPT

Here's a recipe for Ricotta Stuffed Squash:

Ingredients:

- 4 small to medium-sized squash (such as zucchini or yellow squash)
- 1 cup ricotta cheese
- 1/2 cup grated Parmesan cheese
- 1 egg, lightly beaten
- 2 cloves garlic, minced
- 1/4 cup chopped fresh basil
- Salt, to taste
- Black pepper, to taste
- 1 cup marinara sauce
- 1/2 cup shredded mozzarella cheese
- Fresh basil leaves, for garnish (optional)



# Bells and Whistles

- Implementation of techniques learned in class
- Compilation of recipe-specific stop words
- Integration of Text-to-Image synthesis functionality using Stable Diffusion Pipeline(Python package)
- Incorporation of ChatGPT for verification purposes

# Future Scope

- Improve the images that are obtained as part of the text to image synthesis.
- We can expand this project to have a user experience by creating a proper interface.
- We can expand this project to take in audio prompts and give out recipe.





THANK YOU !

