Data Science Mini Project

Cuisine & Recipe Recommendations

Group Members

Pratik Patil
MIS: 112003104
Batch: T3
Division: 2

Prerna Tulsiani
MIS: 112003114
Batch: T3
Division: 2

Problem Statement

Classifying the cuisine from a given list of ingredients and recommending recipes which can be made from the available ingredients and nutrition requirements of the user.

Project Objectives

1. Classifying cuisine from the given list of ingredients.
2. Recommending Recipes from the given list of ingredients.
3. Searching Recipes and get more information about the recipe.
4. Auto-completion when user is typing and particular recipe.
5. Search similar recipes according to nutrition values.

Project Requirements

User Frontend	Made using Flask
Cuisine Classification	Logistic Regression and Voting Classifier

Recipe Recommendation	TF-IDF Vectorizer
Similar Recipes according to nutrition	Average of cosine, euclidean and hamming distance
Auto-completion feature	Done using our Dataset Recipes and Javascript
Integrating ML model with Frontend	Training the model and then dump it into a pickle file and loading the file whenever required in the frontend

Process Flow

Sr. No.	Name	Techniques used
1	*Data Scraping*	*Selenium and Beautiful Soup*
2	*Data Pre-processing*	• *Removing unnecessary characters* • *Lemmatization* • *Removing Stop Words* • *NLTK POS Tagging* • *Normalizing all nutrient values*
3	*Exploratory Data Analysis*	*Box plots, Histograms, SNS Count Plot, Bar Plot*
4	*Data Modeling*	*Logistic Regression, Voting Classifier, Bi-gram Model, TF-IDF Vectorizer*
5	*Performance Evaluation*	*Accuracy and Precision*

GitHub Link

pt3002/Cuisine-and-Recipe-Recommendation (github.com)

Data Scraping

Website 1 - Allrecipes | Recipes, How-Tos, Videos and More

- Static Website where next page navigations and clicking of buttons was not required
- Hence BeatifulSoup used for scraping
- Code - Git Repo Link - Cuisine-and-Recipe-Recommendation/BeautifulSoup.py at master · pt3002/Cuisine-and-Recipe-Recommendation · GitHub

Website 2 - RecipeDB (iiitd.edu.in)

- Very Dynamic website which needed user inputs for giving results and many new page navigations were required.
- Used Selenium for Scraping
- Code - Git Repo Link - Cuisine-and-Recipe-Recommendation/Selenium.py at master · pt3002/Cuisine-and-Recipe-Recommendation · GitHub

Features in Scraped Data -

- Cuisine
- Dish Name
- Ingredients
- Calories
- Carbs
- Fat
- Protein
- Recipe URL
- View Data - Cuisine-and-Recipe-Recommendation/cuisine_final.csv at master · pt3002/Cuisine-and-Recipe-Recommendation (github.com)

Data Pre-processing

- **Removing unnecessary characters from ingredient string**

```
def removing_special_characters(x):

    # remove anything inside paranthesis
    x = re.sub(r"\([^\)]+\)", '', x)

    # remove anything containing a digit
    x = re.sub(r"\S*\d\S*", '', x)

    # make everything lowercase
    x = x.lower()
```

```
    # lemmatize all ingredients in the string
    x = lemmatize(x)

    # remove non-word characters except for , and -
    x = ' '.join(re.findall(r"[-,''\w]+", x))

    # clean excess whitespace
    x = re.sub(r"\s+", ' ', x).strip()

    return x
```

- **Performing Lemmatization on Ingredients**

```
def lemmatize(x):
    for word in re.findall(r"[a-z]+", x):
        x = x.replace(word, wnl.lemmatize(word, 'n') if 's' in word[-3:] else word)
    return x
```

- **Removing Stop words from ingredients and doing NLTK POS Tagging**

```
# set of stop words to filter out
stop_words = set(stopwords.words('english'))

# not-needed terms
not_needed = ['chopped', 'all-purpose', 'divided', 'or', 'to', 'taste', 'sliced', 'a', 'needed', 'finely', 'chopped', 'mi

# define a set of POS tags to keep
keep_tags = set(['NN', 'NNS', 'NNP', 'NNPS'])

cleaned_ingredients = []

for iterator in range(len(cuisine_df.index)):
    if(iterator in cuisine_df.index):
        ingredients = ingredient_list[iterator]
        cleaned_tokens = []
        for ing in ingredients.split(", "):
            individual_cleaned_token = []
            for i in ing.split():
                tokens = word_tokenize(i.lower())
                tagged_tokens = nltk.pos_tag(tokens)
                for token, tag in tagged_tokens:
                    if tag in keep_tags and token not in stop_words and token not in not_needed:
                        individual_cleaned_token.append(token)
            if(individual_cleaned_token):
                cleaned_tokens.append(" ".join(individual_cleaned_token))
        new_ing_list = ", ".join(cleaned_tokens)
        cuisine_df.loc[iterator, 'Ingredients'] = new_ing_list

cuisine_df.to_csv('test_data.csv', index=False)
```

- **Normalizing Nutrients of all recipes individually according to the serving they have per gram.**

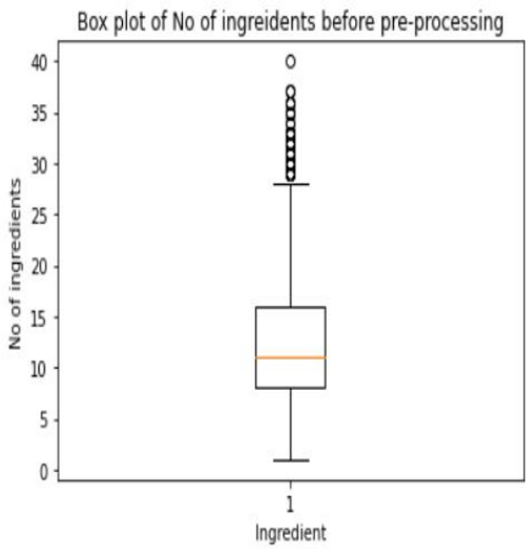
Normalized Nutrients Data Code - [Cuisine-and-Recipe-Recommendation/Nutrition_Normalized.csv at master · pt3002/Cuisine-and-Recipe-Recommendation (github.com)]

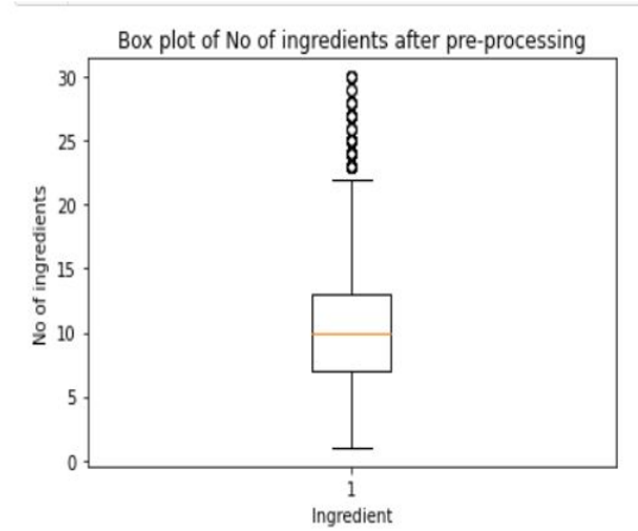
After performing all pre-processing steps:

Final Data: [Cuisine-and-Recipe-Recommendation/cuisine_preprocess.csv at master · pt3002/Cuisine-and-Recipe-Recommendation (github.com)]

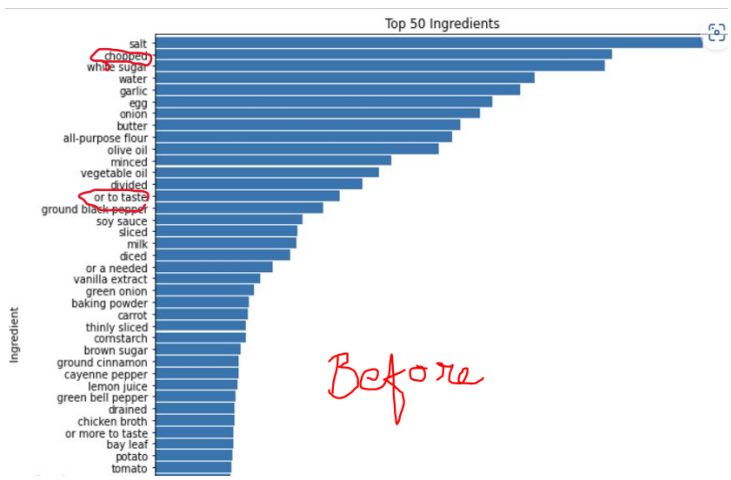
Exploratory Data Analysis

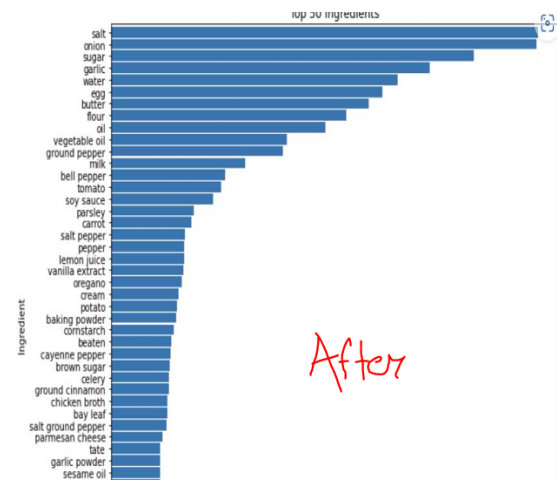
- Box plot of Number of Ingredients before and after pre-processing

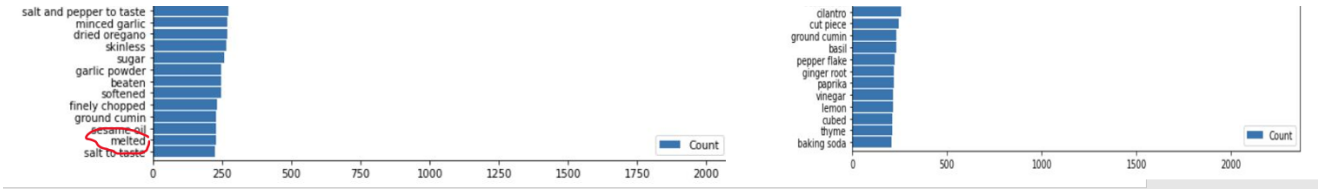




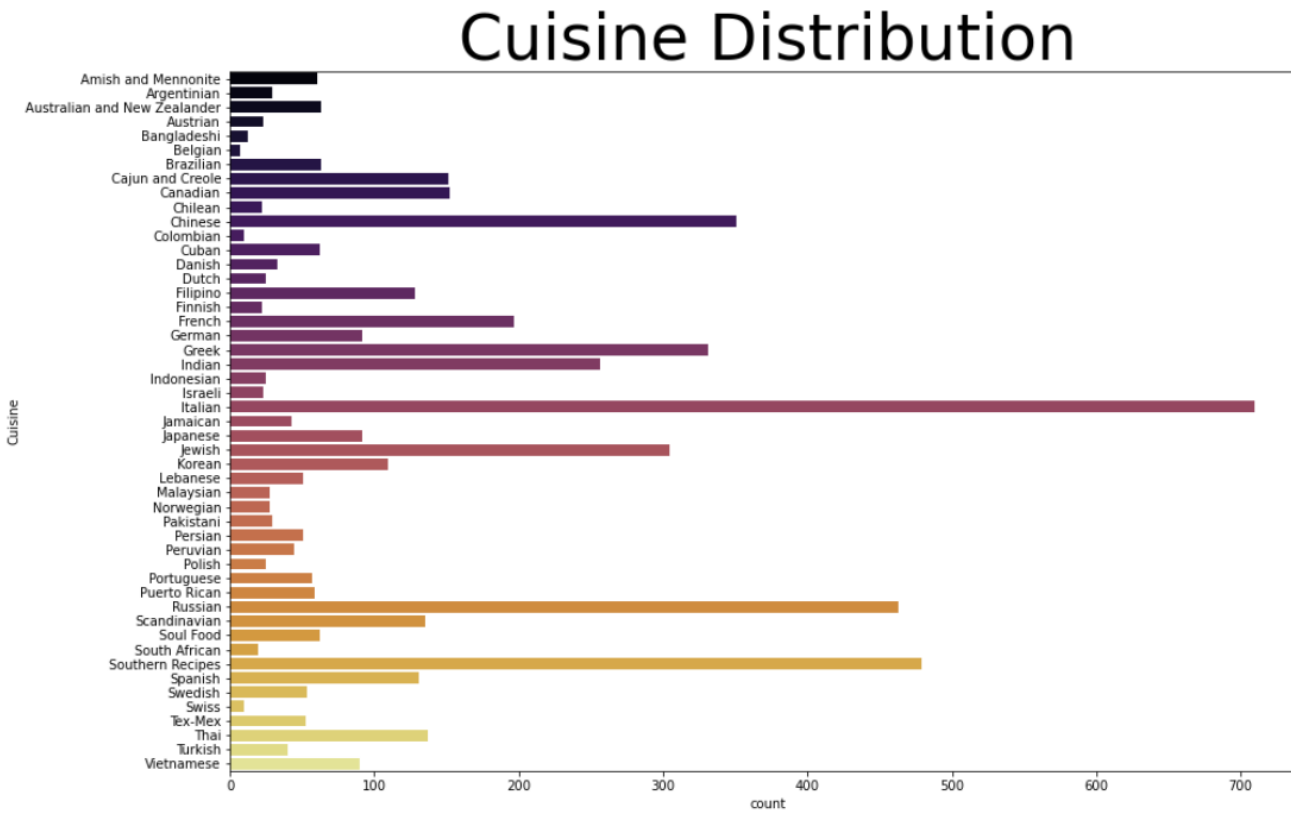
- Find out unnecessary ingredient values that are even present in the data by plotting ingredients with maximum occurrence
  - For e.g. : Chopped, according to taste, diced etc, all these words are unnecessary and need to be removed manually



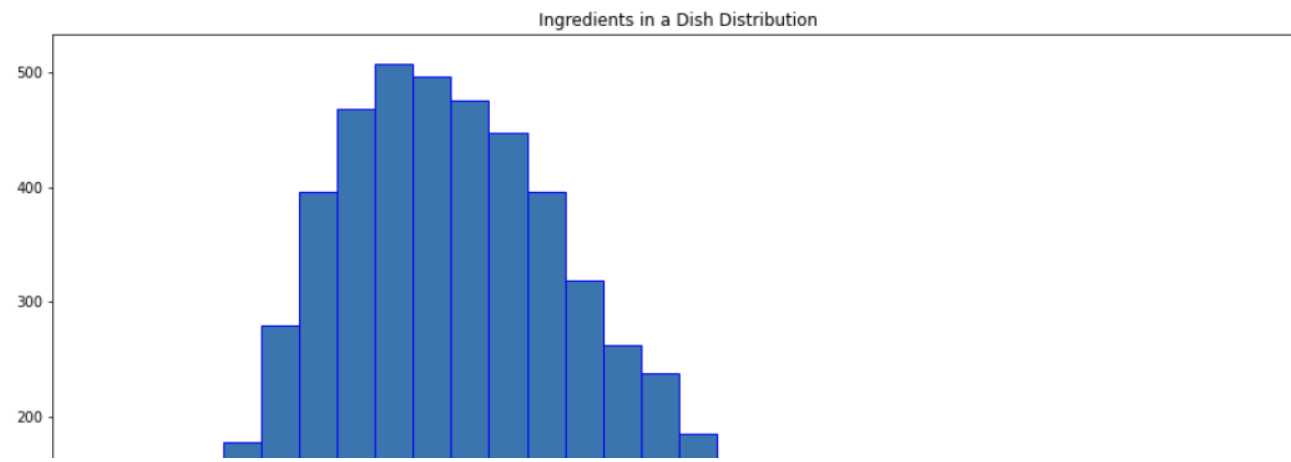


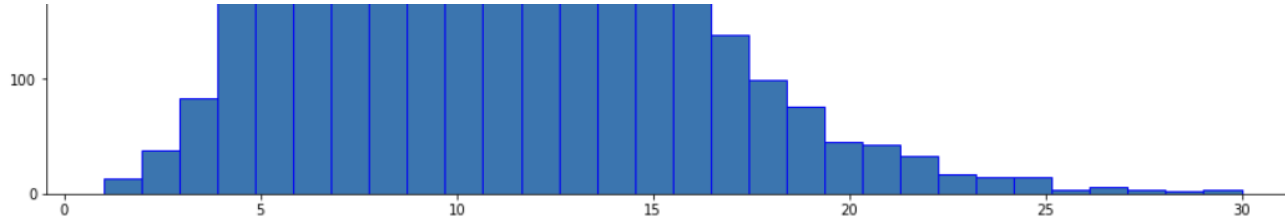


- Finding out Cuisines and Number of Recipes each cuisine has by SNS Count Plot

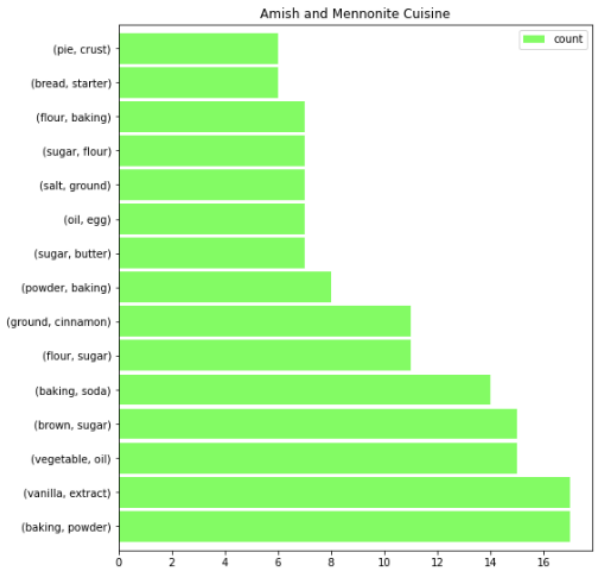


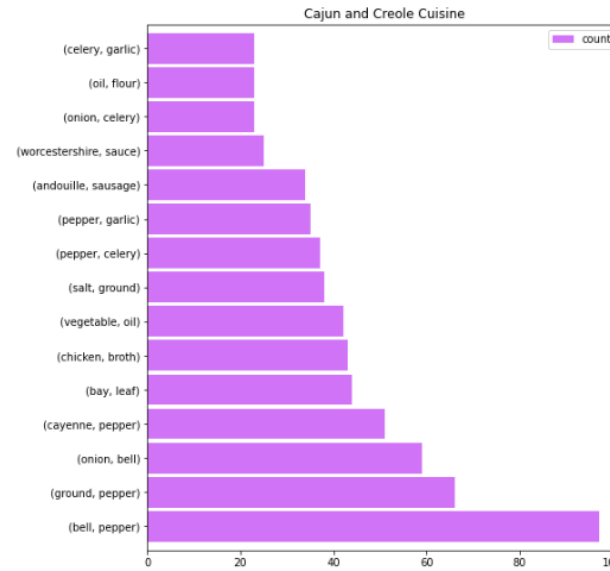
- Plotting Ingredients in a Dish Distribution

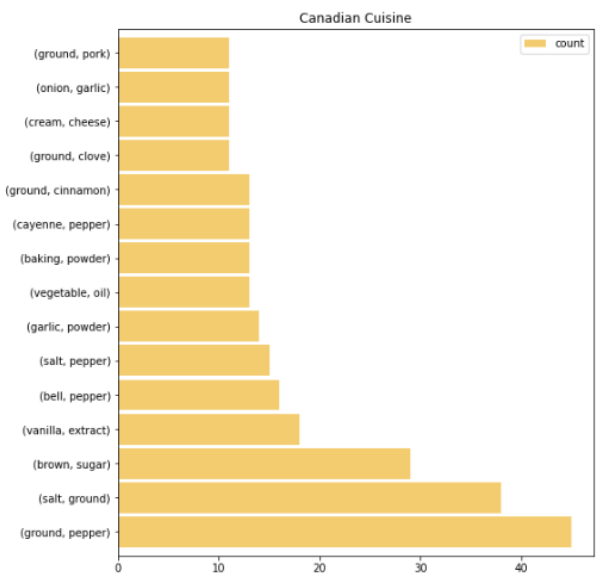


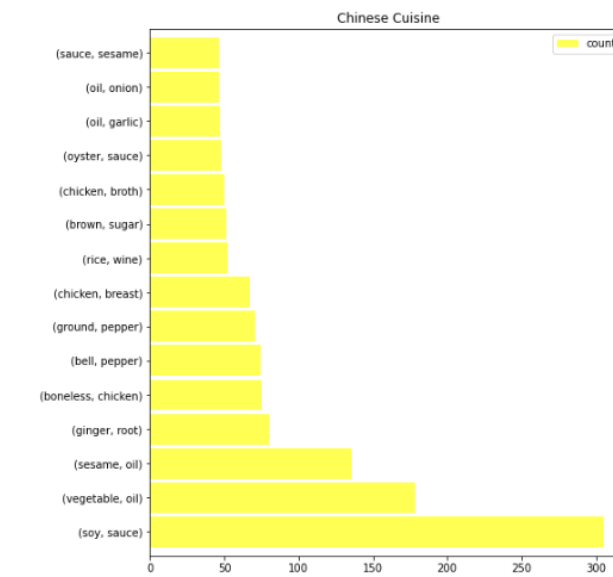


- Plotting bar plots of top 10 bi-grams of 4 example cuisines









Code of all these plots is -

[Cuisine-and-Recipe-Recommendation/Cuisine Prediction.ipynb at master · pt3002/Cuisine-and-Recipe-Recommendation · GitHub]

Data Modeling

For Cuisine Classification -

Logistic Regression and Voting Classifier

In [91]:
```
from sklearn.preprocessing import LabelEncoder

encoder = LabelEncoder()
y_transformed = encoder.fit_transform(cuisine_df.Cuisine)
```

In [92]:
```
# Logistic Regression

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_train_vectorized, y_transformed , random_state = 0)

from sklearn.linear_model import LogisticRegression

clf1 = LogisticRegression(C=10,dual=False)
clf1.fit(X_train , y_train)
clf1.score(X_test, y_test)
```

Out[92]: 0.5467255334805003

In [93]:
```
from sklearn.svm import SVC
from sklearn.ensemble import VotingClassifier
vclf=VotingClassifier(estimators=[('clf1',LogisticRegression(C=10,dual=False)),('clf2',SVC(C=100,gamma=1,kernel='rbf',probab
vclf.fit(X_train , y_train)
vclf.score(X_test, y_test)
```

Out[93]: 0.5695364238410596

Code - [Cuisine-and-Recipe-Recommendation/cuisinetfidf.py at master · pt3002/Cuisine-and-Recipe-Recommendation (github.com)]

Link of .pkl file - [Cuisine-and-Recipe-Recommendation/cuisine.pkl at master · pt3002/Cuisine-and-Recipe-Recommendation · GitHub]

For Recipe Recommendation -

TF - IDF Vectorizer

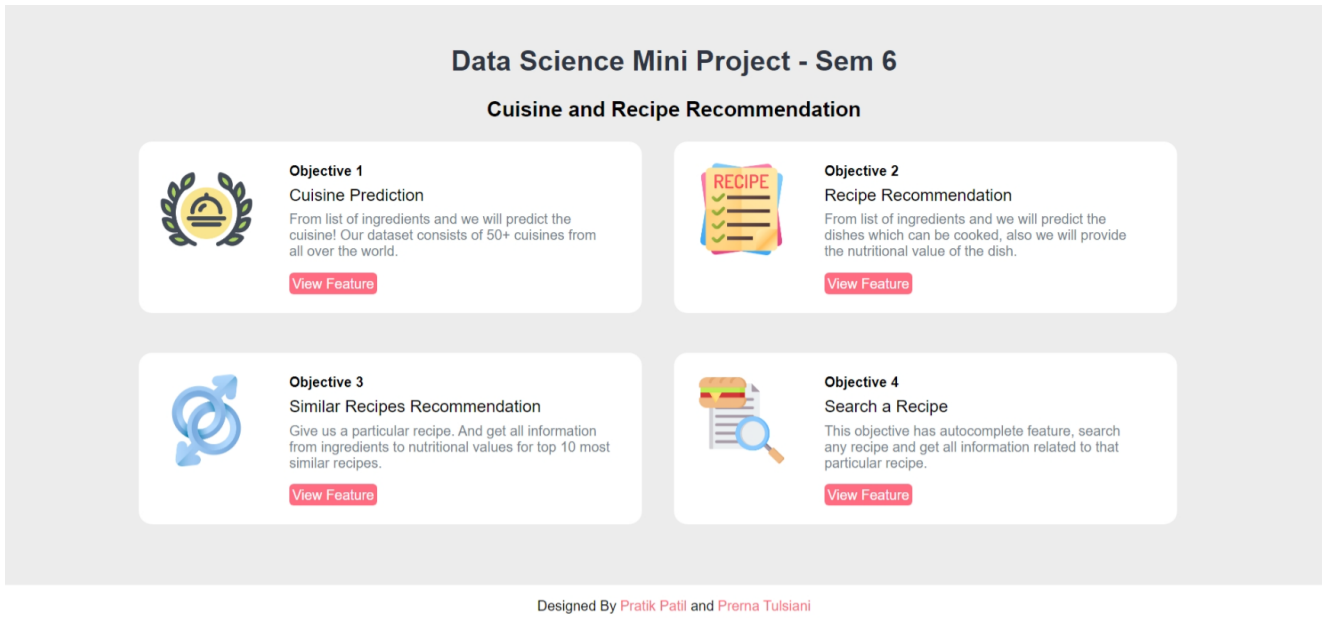
```
# TF-IDF feature extractor
tfidf = TfidfVectorizer()
tfidf.fit(df_recipes['Ingredients'])
tfidf_recipe = tfidf.transform(df_recipes['Ingredients'])

# save the tfidf model and encodings
with open(config.TFIDF_MODEL_PATH, "wb") as f:
    pickle.dump(tfidf, f)
```

```
with open(config.TFIDF_ENCODING_PATH, "wb") as f:
    pickle.dump(tfidf_recipe, f)
```

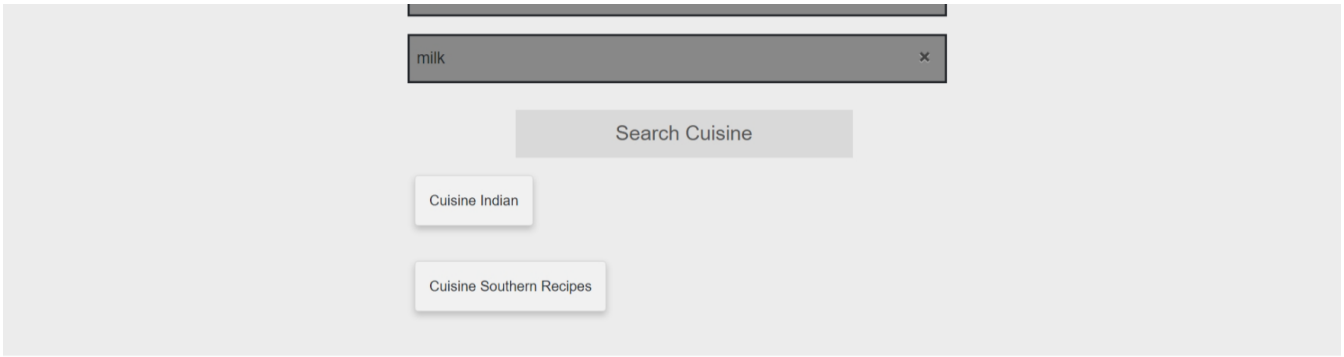
Code - [Cuisine-and-Recipe-Recommendation/tfidfencoder.py at master · pt3002/Cuisine-and-Recipe-Recommendation (github.com)]

Link of .pkl file - [Cuisine-and-Recipe-Recommendation/tfidf.pkl at master · pt3002/Cuisine-and-Recipe-Recommendation · GitHub]

Front End Screenshots

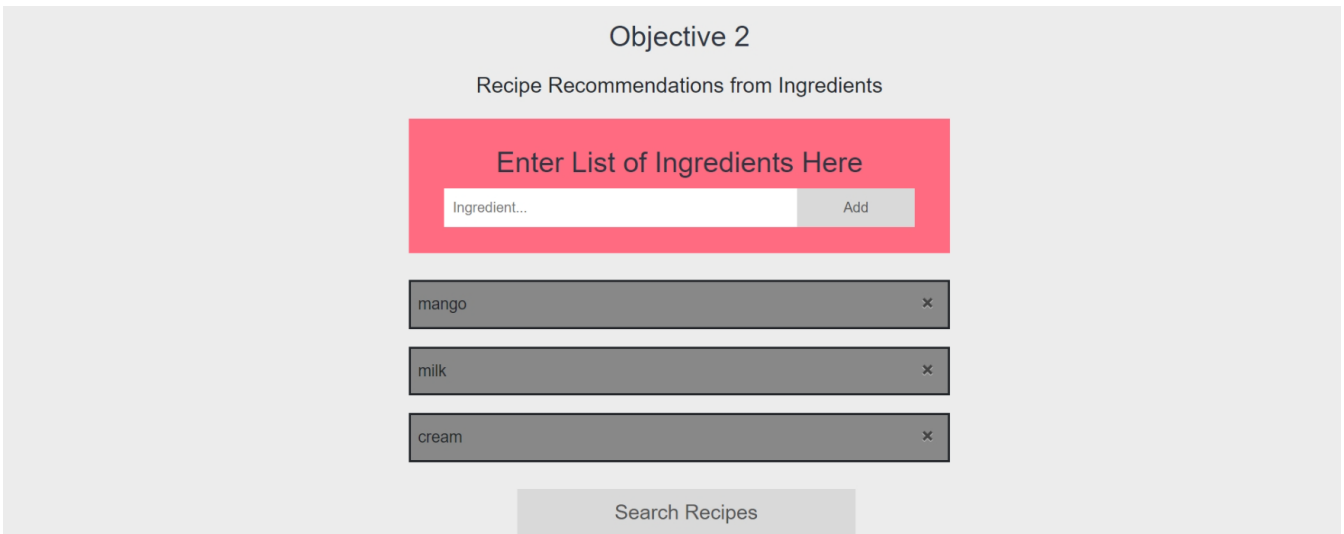


Objective 1 - Cuisine Classification



Designed By Pratik Patil and Prerna Tulsiani

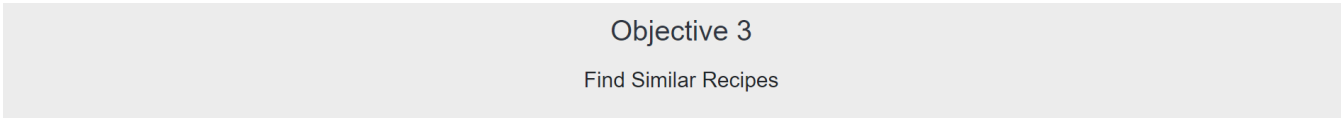
Objective 2 - Recipe Recommendation

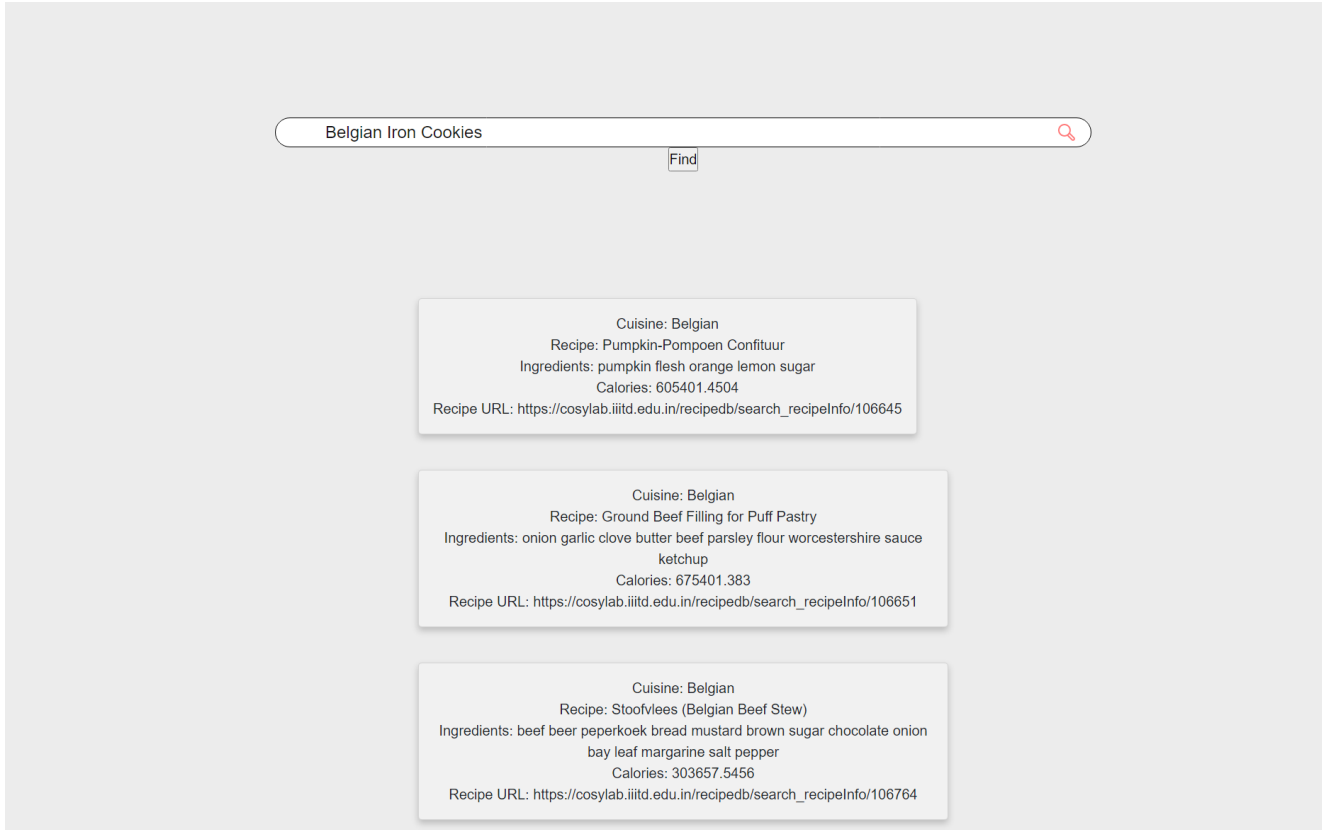


Cuisine	Recipe	Ingredients	Calories	URL
Indian	Restaurant-Style Mango Lassi	mango pulp mango slice juice plain yogurt milk ice cube	121	https://www.allrecipes.com/recipe/54319/restaurant-style-mango-lassi/
Indian	My Mango Lassi	mango plain yogurt cold milk cream confectioner sugar ground cardamom	228	https://www.allrecipes.com/recipe/264349/my-mango-lassi/
Indian	Mango Lassi-Come-Home	yogurt vanilla yogurt milk pureed mango honey mango nectar ground cardamom	207	https://www.allrecipes.com/recipe/206172/mango-lassi-come-home/
Indian	Easy Mango Lassi	plain milk yogurt milk mango sugar ground cardamom	195	https://www.allrecipes.com/recipe/202402/easy-mango-lassi/
Mexican	Mango Salsa Steak Fajitas	taco vegetable oil steak mango salsa mango juice lime jalapeno chile onion cilantro leaf	693.3845	https://cosylab.iiitd.edu.in/recipedb/search_recipeInfo/5979

Designed By Pratik Patil and Prerna Tulsiani

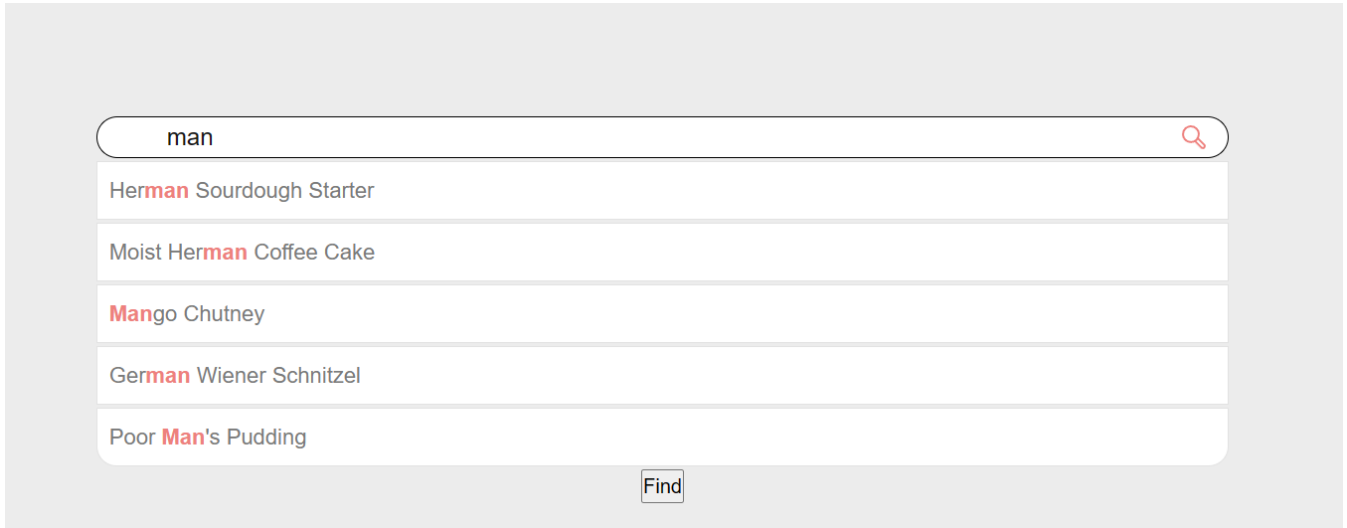
Objective 3 - Similar Recipes according to Nutrient Values





Objective 4 - Search a Recipe

Objective 5 - AutoComplete Feature



ALL CSV FILES -
https://drive.google.com/drive/folders/1Iu6_C0X1fq29Dh9e0SwWluv7vBbGiUaY?usp=sharing