

```
In [ ]: !pip install langdetect
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Requirement already satisfied: langdetect in /usr/local/lib/python3.8/dist-pa
ckages (1.0.9)
Requirement already satisfied: six in /usr/local/lib/python3.8/dist-packages
(from langdetect) (1.15.0)
```

```
In [ ]: import pandas as pd
import numpy as np
from langdetect import detect
```

This is the original dataset that we will use for this project

```
In [ ]: df = pd.read_csv("/content/drive/MyDrive/Food_ingredient_dataset/IndianFoodDat
asetCSV.csv")
```

```
In [ ]: df.head()
```

Out[]:

| | Srno | RecipeName | TranslatedRecipeName | Ingredients | TranslatedIngredients | PrepTimeInMins |
|---|------|---|---|---|---|----------------|
| 0 | 1 | Masala Karela Recipe | Masala Karela Recipe | 6 Karela (Bitter Gourd/ Pavakkai) - deseeded, S... | 6 Karela (Bitter Gourd/ Pavakkai) - deseeded, S... | 15 |
| 1 | 2 | टमाटर पुलियोगरे रेसिपी - Spicy Tomato Rice (Recipe) | Spicy Tomato Rice (Recipe) | 2-1/2 कप चावल - पका ले, 3 टमाटर, 3 छोटा चमच बी... | 2-1 / 2 cups rice - cooked, 3 tomatoes, 3 teas... | 5 |
| 2 | 3 | Ragi Semiya Upma Recipe - Ragi Millet Vermicel... | Ragi Semiya Upma Recipe - Ragi Millet Vermicel... | 1-1/2 cups Rice Vermicelli Noodles (Thin), 1 On... | 1-1/2 cups Rice Vermicelli Noodles (Thin), 1 On... | 20 |
| 3 | 4 | Gongura Chicken Curry Recipe - Andhra Style Go... | Gongura Chicken Curry Recipe - Andhra Style Go... | 500 grams Chicken, 2 Onion - chopped, 1 Tomato -... | 500 grams Chicken, 2 Onion - chopped, 1 Tomato -... | 15 |
| 4 | 5 | आंध्रा स्टाइल आलम पचड़ी रेसिपी - Adrak Chutney ... | Andhra Style Alam Pachadi Recipe - Adrak Chutn... | 1 बड़ा चमच चना दाल, 1 बड़ा चमच सफ़ेद उरद दाल, 2... | 1 tablespoon chana dal, 1 tablespoon white ura... | 10 |

Below, we will remove the rows with nan in them

```
In [ ]: df=df.dropna() #removing NA's  
df.shape
```

```
Out[ ]: (6865, 15)
```

```
In [ ]: df = df.reset_index()  
df = df.drop(['index'], axis=1)
```

```
In [ ]: def detect_lang(text): #detecting hindi language  
        return detect(text)
```

Now, we will also remove those rows where the TranslatedIngredients column has hindi letters

```
In [ ]: df_filtered = pd.DataFrame(columns = df.columns)  
for i in range(len(df)):  
    lang = detect(df['TranslatedIngredients'][i])  
    if lang=='en':  
        df_filtered = df_filtered.append(df.iloc[i], ignore_index = True)
```

```
In [ ]: len(df_filtered)
```

```
Out[ ]: 6276
```

```
In [ ]: columns = df_filtered.columns  
print(columns)
```

```
Index(['Srno', 'RecipeName', 'TranslatedRecipeName', 'Ingredients',  
      'TranslatedIngredients', 'PrepTimeInMins', 'CookTimeInMins',  
      'TotalTimeInMins', 'Servings', 'Cuisine', 'Course', 'Diet',  
      'Instructions', 'TranslatedInstructions', 'URL'],  
      dtype='object')
```

```
In [ ]: !pip install pying
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting pying
  Downloading PyIng-0.1.15.tar.gz (74.4 MB)
    |████████████████████████████████████████| 74.4 MB 1.2 MB/s
Collecting tflite-runtime>=2.5
  Downloading tflite_runtime-2.10.0-cp38-cp38-manylinux2014_x86_64.whl (2.5 M
B)
    |████████████████████████████████████████| 2.5 MB 46.2 MB/s
Requirement already satisfied: numpy in /usr/local/lib/python3.8/dist-package
s (from pying) (1.21.6)
Building wheels for collected packages: pying
  Building wheel for pying (setup.py) ... done
  Created wheel for pying: filename=PyIng-0.1.15-py3-none-any.whl size=744193
27 sha256=6a86ef300e827b7d0af599ee50bdf05a80fd8e17dddaa4b9b8731f6f4b783a4c
  Stored in directory: /root/.cache/pip/wheels/00/a3/19/5c50922ead9784d290c95
8608c4542e99fd988408d94a79f80
Successfully built pying
Installing collected packages: tflite-runtime, pying
Successfully installed pying-0.1.15 tflite-runtime-2.10.0
```

This is the library that we will use to parse ingredients

```
In [ ]: from PyIng import parse_ingredients
```

```
In [ ]: for i in range(len(df_filtered)):
    temp_ingreds = df_filtered['TranslatedIngredients'][i].split(',')
    parsed_ingredients = parse_ingredients(temp_ingreds)
    l = []
    if type(parsed_ingredients) == dict:
        l.append(parsed_ingredients['name'])
    else:
        for j in range(len(parsed_ingredients)):
            l.append(parsed_ingredients[j]['name'])
    df_filtered['TranslatedIngredients'][i] = l
```

Below, we will show how does parsed ingredient looks like

```
In [ ]: print(df_filtered['Ingredients'][5000])
        print(df_filtered['TranslatedIngredients'][5000])
```

```
1 Cauliflower (gobi) - cut into florets,1 teaspoon Fennel seeds (Saunf),1 Gin
ger,2 Tomatoes - finely chopped,2 Bay leaves (tej patta),1/2 cup Green Chutne
y (Coriander & Mint),1/2 cup Curd (Dahi / Yogurt),1/2 teaspoon Turmeric powde
r (Haldi),1 teaspoon Coriander Powder (Dhania),Salt,Red Chilli powder,Sunflow
er Oil - for cooking
['cauliflower', 'fennel seeds', 'ginger', 'tomatoes', 'bay leaves', 'green ch
utney coriander mint', 'curd dahi yogurt', 'turmeric powder', 'coriander powd
er', 'salt', 'red chilli powder', 'sunflower oil']
```

```
In [ ]: df_filtered.head()
```

```
Out[ ]:
```

| | Srno | RecipeName | TranslatedRecipeName | Ingredients | TranslatedIngredients | PrepTimeInMins |
|---|------|---|---|---|---|----------------|
| 0 | 1 | Masala Karela Recipe | Masala Karela Recipe | 6 Karela (Bitter Gourd/ Pavakkai) - deseeded,S... | [gourd, salt, onion, flour, turmeric powder, r... | 15 |
| 1 | 2 | टमाटर पुलियोगरे रेसिपी - Spicy Tomato Rice (Recipe) | Spicy Tomato Rice (Recipe) | 2-1/2 कप चावल - पका ले,3 टमाटर,3 छोटा चमच बी... | [rice cooked, tomatoes, belle, salt, chickpea ... | 5 |
| 2 | 3 | Ragi Semiya Upma Recipe - Ragi Millet Vermicel... | Ragi Semiya Upma Recipe - Ragi Millet Vermicel... | 1-1/2 cups Rice Vermicelli Noodles (Thin),1 On... | [rice vermicelli noodles, onion, carrots, gree... | 20 |
| 3 | 4 | Gongura Chicken Curry Recipe - Andhra Style Go... | Gongura Chicken Curry Recipe - Andhra Style Go... | 500 grams Chicken,2 Onion - chopped,1 Tomato -... | [chicken, onion, tomato, green chillies, ginge... | 15 |
| 4 | 5 | आंध्रा स्टाइल आलम पचड़ी रेसिपी - Adrak Chutney ... | Andhra Style Alam Pachadi Recipe - Adrak Chutn... | 1 बड़ा चमच चना दाल,1 बड़ा चमच सफ़ेद उरद दाल,2... | [chana dal, white, red chillies, coriander see... | 10 |

Now we need to save this preprocessed file for future use in prediction

```
In [ ]: df_filtered.to_csv('df_filtered.csv')
```

```
In [ ]: df_filtered.head()
```

```
Out[ ]:
```

| | Srno | RecipeName | TranslatedRecipeName | Ingredients | TranslatedIngredients | Prep |
|---|------|--|---|--|---|-----------------|
| 0 | 1 | Masala Karela Recipe | Masala Karela Recipe | 6 Karela (Bitter Gourd/ Pavakkai) - deseeded, S... | gourd,salt,onion,flour,turmeric powder,red chi... | |
| 1 | 2 | टमाटर पुलियोगरे रेसिपी - Spicy Tomato Rice (Re... | Spicy Tomato Rice (Recipe) | 2-1/2 कप चावल - पका ले,3 टमाटर,3 छोटा चमच्च बी... | cooked,tomatoes,belle,salt,chickpea lenti... | rice |
| 2 | 3 | Ragi Semiya Upma Recipe - Ragi Millet Vermicel... | Ragi Semiya Upma Recipe - Ragi Millet Vermicel... | 1-1/2 cups Rice Vermicelli Noodles (Thin),1 On... | noodles,onion,carrots,green pe... | rice vermicelli |
| 3 | 4 | Gongura Chicken Curry Recipe - Andhra Style Go... | Gongura Chicken Curry Recipe - Andhra Style Go... | 500 grams Chicken,2 Onion - chopped,1 Tomato -... | chicken,onion,tomato,green chillies,ginger,gar... | |
| 4 | 5 | आंध्रा स्टाइल आलम पचड़ी रेसिपी - Adrak Chutney ... | Andhra Style Alam Pachadi Recipe - Adrak Chutn... | 1 बड़ा चमच्च चना दाल,1 बड़ा चमच्च सफ़ेद उरद दाल,2... | chana dal,white,red chillies,coriander seeds,g... | |