**2. Data cleaning: clean the blank and duplicated rows of the dataset**

**4. One hot encoding: merge all the extracted words into a comprehensive food ingredient list and convert it to a one hot-encoded vector**

**3. Ingredient Extraction: Extract food ingredient terms from each sentence of the Cleaned Ingredient lists**

1. **Data Loading: load the ‘Food Ingredients and Recipes Dataset with images’ from Kaggle**