❖ **Kaggle Challenge-**

Method 1- Multinomial Naive Bayes

- Ingredients are selected as the features
- Used **count vector** to make training data for the function of sklearn (Multinomial naive bayes)
- Removed extra spaces from the text
- The ingredient like baking powder get split into ‘baking ’ and ‘powder’ so i consider both as a different feature in space
- Created vocabulary from all the words in the ingredients of all the recipes of the training data

Method 2- Decision Tree

- A decision tree is a tree where each node represents a feature(attribute), each link(branch) represents a decision(rule) and each leaf represents an outcome.
- Created vocabulary from all the words in the ingredients of all the recipes of the training data
- Ingredients are selected as the features
- The ingredient like baking powder get split into ‘baking ’ and ‘powder’ so i consider both as a different feature in space
- Used count vector to make training data for the function of sklearn (Decision tree)

Accuracy in multinomial naive bayes- 72.3 (α = 0.5)
Accuracy in multinomial naive bayes- 71.9 (α = 1)

Analysis of all the methods used

- **Decision tree and count vectors**
  With this method i got the accuracy of 63.3, because of high variance and low bias.
- **K-NN and count vectors**
  With this method i got the accuracy of 62.22 , and it also takes 1 hour to give output
- **Decision tree + TF-IDF vectors**
  With this i got the accuracy of 62.8, tf-idf does not give good result on short text.
- **K-NN + TF-idf vectors**
  With this method i got the accuracy of 61.9, as tf-idf does not work well with short text