What's Cooking?

Aditya Gouroju

October 29 2020

1 Introduction

Classification of the cuisines based upon the ingredients used in the cuisine

2 Data Acquisition and Preprocessing

2.1 Data sources

Both Training and Test Datasets are provided in the competition in ".json" format.An Additional Data Description text file is also provided for the Datasets.

2.2 Data Cleaning

The provided dataset in .json is converted into to Dataframe using the pandas library. The DataFrame consisted of 3 columns in which two of the columns are "Id" and the target feature "Cuisine" and the third column is tuples of ingredients used in a cuisine

Both the training and Test datasets are combined as it is more effective to preprocess them at a time rather than preprocessing them individually.There are no Columns with "Null" data .

I still have to explore the dataset properly.This is just a basic attempt.

2.3 Feature Selection

Using some basic functions the total ingredients can be found. One Hot encoding the datasets with the ingredients and the cuisines we obtain a huge dataset.All the columns are used as features

3 Methodology

Neural networks are used for this classification of the cuisines as neural networks are more efficient than basic classification models

Using tensorflow and keras a basic model is made with SGD optimizer and Sparse Categorical Cross Entropy loss function with accuracy as its metric as the evaluation is based upon classification accuracy

4 Results & Conclusion

By using this model i was able to get a training a accuracy of 93% and a validation accuracy of 80%