MANIPAL INSTITUTE OF TECHNOLOGY



(A constituent Institute of Manipal University)

MANIPAL - 576 104, KARNATAKA, INDIA

MARCH 2017

**Recipe Cuisine Classification on the basis of Ingredients**

*By*

**Ramya Janakiraman**

**140911422**

**Anisha Mathur**

**140911468**

*Under the Guidance of*

*Ms. Anju R.*

*Ms. Girija Attigeri*

*Mr. Arjun CV*

1. **INTRODUCTION**

When recipes are shared, two factors need to be observed: user’s preference and what genre of recipe it is. A lot of importance is given to user’s preference, with little importance to the latter. When designing an online recipe sharing website, an automatic method to classify recipes would save a lot of manual labor. The easiest classifier for a recipe would be cuisine.

1. **PROBLEM DEFINITION**

From a given set of recipes, based on the ingredients, we try to predict its cuisine. Since the recipes are all represented in the form of documents, we will have to use practices from text mining to attempt to predict and classify.

1. **OBJECTIVE**

Classifying recipes into cuisines based on ingredients.

1. **SCOPE / IMPORTANCE OF PROJECT**

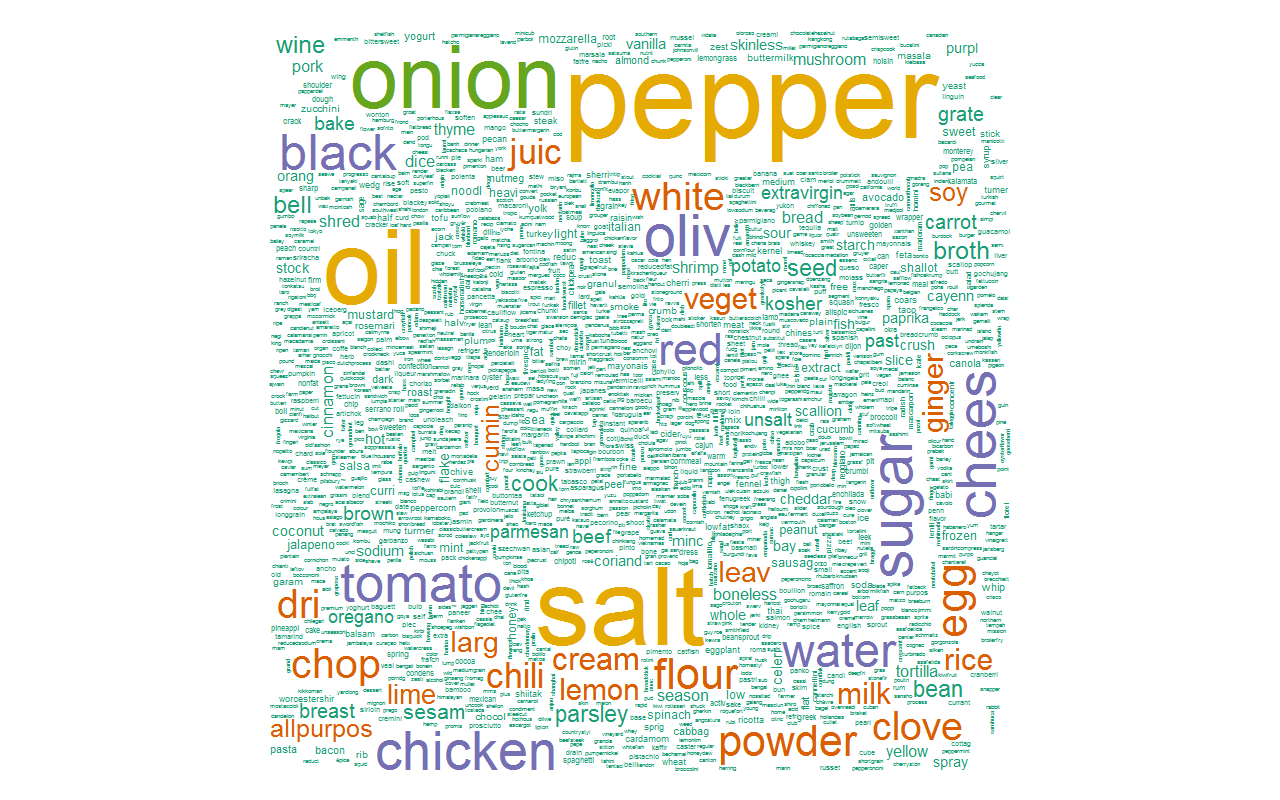
With the digital boom, looking up recipes online has become a lot easier than calling up your grandmother. Many techniques have come up with the goal of easing out the provision of recipe sharing services. For example, there are new data mining techniques to represent recipes. A classifier for the type of recipe would help too. This would also help us provide insights about the similarity between different cuisines worldwide that would in turn help us trace different cultures and their roots and origins.

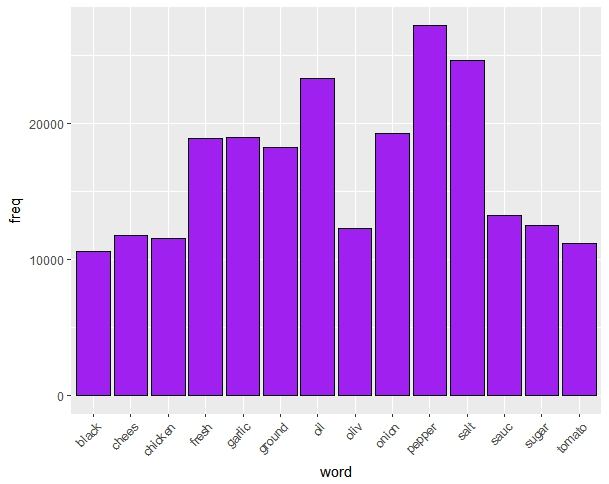
1. **METHODOLOGY**

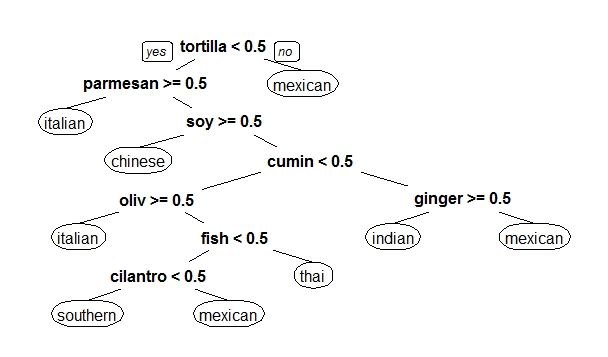
The paper, which we have referred to, uses Support Vector Machines (SVM) to classify the recipes based on their ingredients. We have used two algorithms: Decision tree and Random Forest and compared. We check the accuracy of each model thus trained, and find out the best method for this task.

1. **RESULTS**

We find that using Random Forest, we get an accuracy of 0.73. Decision Tree gives an accuracy of 0.41.







**Random forest statistics:**

Overall Statistics

Accuracy : 0.7368

95% CI : (0.7269, 0.7464)

No Information Rate : 0.1972

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.7029

Mcnemar's Test P-Value : NA

1. **CONCLUSION**

Out of the two methods we have used to classify, Random Forests gave the best result(73%). However this is still not reliable. We need to work on it and build a stronger model to get better results.

1. **REFERENCES**

* A comparative study of decision tree ID3 and C4.5
* Implementation of decision tree algorithm C4.5
* Automatic recipe cuisine classification by ingredients
* Kaggle
* Analytics Vidya
* R documentation