

Présentation du projet de recherche : What's  
cooking ? Use recipe ingredients to categorize the  
cuisine

Ayman HAMZAOUÏ & Julien DELAUNAY

23 mars 2019

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Les données</b>	<b>3</b>
<b>3</b>	<b>Les méthodes utilisées</b>	<b>4</b>
3.1	TF-IDF . . . . .	4
3.2	Methode 2 . . . . .	4
3.3	Methode 3 . . . . .	4
3.4	Methode 4 . . . . .	4
3.5	Methode 5 . . . . .	4
3.6	Methode 6 . . . . .	4
<b>4</b>	<b>Les résultats</b>	<b>5</b>
<b>5</b>	<b>Conclusion</b>	<b>5</b>
<b>A</b>	<b>Choix du design</b>	<b>6</b>
<b>B</b>	<b>Utilisation de Git</b>	<b>6</b>

# 1 Introduction

Nous avons choisi la base de données *"What's cooking? Use recipe ingredients to categorize the cuisine"*<sup>1</sup> parmi les bases de données de Kaggle. Le domaine de cette base de données nous est apparu intéressant et nous avons remarqué que les données fournies semblent pouvoir être séparés et ainsi analysées avec les outils appris en cours.

## 2 Les données

Les données fournies par Kaggle sont sous la forme de deux documents json. Le fichier de test et le fichier d'entraînement. Pour le fichier d'entraînement, nous avons un id de la recette, le type de cuisine pour cette recette et les ingrédients nécessaires à sa fabrication.

Nous avons effectué des recherches sur les données fournies afin d'obtenir quelques informations. Nous avons ainsi découvert que les cuisines utilisent entre 1 et 65 ingrédients par recette. Les ingrédients les plus utilisés varient en fonction des cuisines, il est donc intéressant d'observer si un modèle peut prévoir en fonction des ingrédients à quelle cuisine la recette appartient.

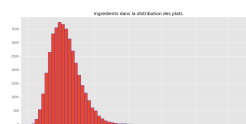


FIGURE 1 – Repartition des ingrédients

Afin de commencer l'analyse et le traitement des données nous avons décidé de modifier le format d'entrée et transformer la liste des ingrédients sous la forme d'un tableau, avec une recette par ligne, et un ingrédients par colonne.

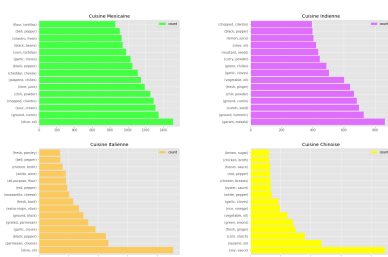


FIGURE 2 – Repartition des ingrédients par cuisine

---

1. <https://www.kaggle.com/c/whats-cooking-kernels-only>

## **3 Les méthodes utilisées**

### **3.1 TF-IDF**

### **3.2 Methode 2**

### **3.3 Methode 3**

### **3.4 Methode 4**

### **3.5 Methode 5**

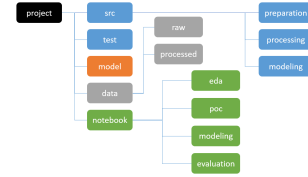
### **3.6 Methode 6**

4 Les résultats

5 Conclusion

## A Choix du design

Pour le design du code, nous avons utilisé ce format qui nous est inspiré de l'article écrit par Edward Ma<sup>2</sup>. Nous avons donc utilisé jupyter pour commencer à travailler sur les données afin de faciliter et rendre plus rapide le début du projet.



## B Utilisation de Git

Nous avons mis en place un github<sup>3</sup> afin de répartir le travail et indiquer le travail restant à faire. Pour cela, nous avons créé des *issues*<sup>4</sup> au fur et à mesure de l'avancée du projet.

FIGURE 3 – Design d'un projet de Data Science

---

2. <https://goo.gl/a8cffw>

3. [https://github.com/hamzaouiayman/IFT\\_712\\_projet\\_session.git](https://github.com/hamzaouiayman/IFT_712_projet_session.git)

4. Tâches à effectuer avant la fin du projet