



Master 1 Informatique et Ingénierie des Systèmes Complexes (IISC)

Université Cergy-Paris

Projet de synthèse

---

# Big Cooking Data

---

*Rapporteur : Arthur MIMOUNI*



*Equipes :*

Arthur MIMOUNI  
Mamadou Bella DIALLO  
Marouane RACHIDY  
Imane CHBIRA  
Matthieu SAUVAGEOT

*Tuteur technique :*

Pr. Dan VODISLAV

*Encadrant de gestion de projet :*

Pr. Tianxiao LIU

Rendu le  
13 février 2022

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Contexte du projet . . . . .	3
1.2	Mise en scénario . . . . .	3
1.3	Objectif du projet . . . . .	4
1.4	Organisation du rapport . . . . .	4
<b>2</b>	<b>Présentation et spécification du projet</b>	<b>5</b>
2.1	Fonctionnalités attendues . . . . .	5
2.2	Conception globale du projet . . . . .	6
2.2.1	Vue pour l'utilisateur . . . . .	6
2.2.2	Architecture technique . . . . .	6
2.3	Problématiques identifiées et solutions envisagées . . . . .	6
2.4	Environnement de travail . . . . .	6
<b>3</b>	<b>Chapitre technique : webscraping</b>	<b>7</b>
3.1	Analyse de la problématique . . . . .	7
3.2	Etat de l'art : études des solutions existantes . . . . .	7
3.3	Solution proposée et sa mise en œuvre . . . . .	7
<b>4</b>	<b>Chapitre technique data clustering</b>	<b>7</b>
4.1	Analyse de la problématique . . . . .	7
4.2	Etat de l'art : études des solutions existantes . . . . .	7
4.3	Solution proposée et sa mise en œuvre . . . . .	7
<b>5</b>	<b>Chapitre technique algorithme de suggestion</b>	<b>7</b>
5.1	Analyse de la problématique . . . . .	7
5.2	Etat de l'art : études des solutions existantes . . . . .	7
5.3	Solution proposée et sa mise en œuvre . . . . .	7
<b>6</b>	<b>Rendu final</b>	<b>8</b>
6.1	Interface utilisateur finale . . . . .	8
6.2	Tests utilisateur et certification . . . . .	8
6.3	Autres tests et certifications . . . . .	8
<b>7</b>	<b>Gestion de projet</b>	<b>9</b>
7.1	Méthode de gestion . . . . .	9
7.2	Répartition de tâches . . . . .	9
<b>8</b>	<b>Conclusion et perspectives</b>	<b>10</b>
8.1	Conclusion . . . . .	10
8.2	Perspectives . . . . .	10

## Table des figures

1	Mise en scénario du projet . . . . .	3
2	Diagramme de cas utilisation de l'application web . . . . .	5
3	Architecture base de données . . . . .	6

# Remerciements

# 1 Introduction

## 1.1 Contexte du projet

Avec la popularité croissante des sites Web au cours des dernières années, Internet est devenu une source accessible pour de grandes quantités d'informations et a ouvert les portes à un partage collaboratif à grande échelle notamment en ce qui concerne les recettes de cuisine. Ces recettes sont proposées aux utilisateurs à l'aide d'algorithmes de suggestions de contenu qui sont devenus indispensables sur la plupart des applications web. Leur but est de proposer une meilleure approche pour organiser le contenu de façon pertinente aussi bien pour l'utilisateur que pour la plateforme.

## 1.2 Mise en scénario

La mise en scénario que nous avons imaginé se déroule en plusieurs étapes. Pour se faire, l'utilisateur aura accès à une interface graphique depuis laquelle il pourra sélectionner les différentes fonctionnalités de notre application.

**Remarque :** Nous ne parlerons pas du processus de data-scraping et d'apprentissage non supervisé visant à remplir notre base de données car il ne s'agit pas d'une fonctionnalité de l'utilisateur. Pour plus de détails sur cette partie voir « lien ».

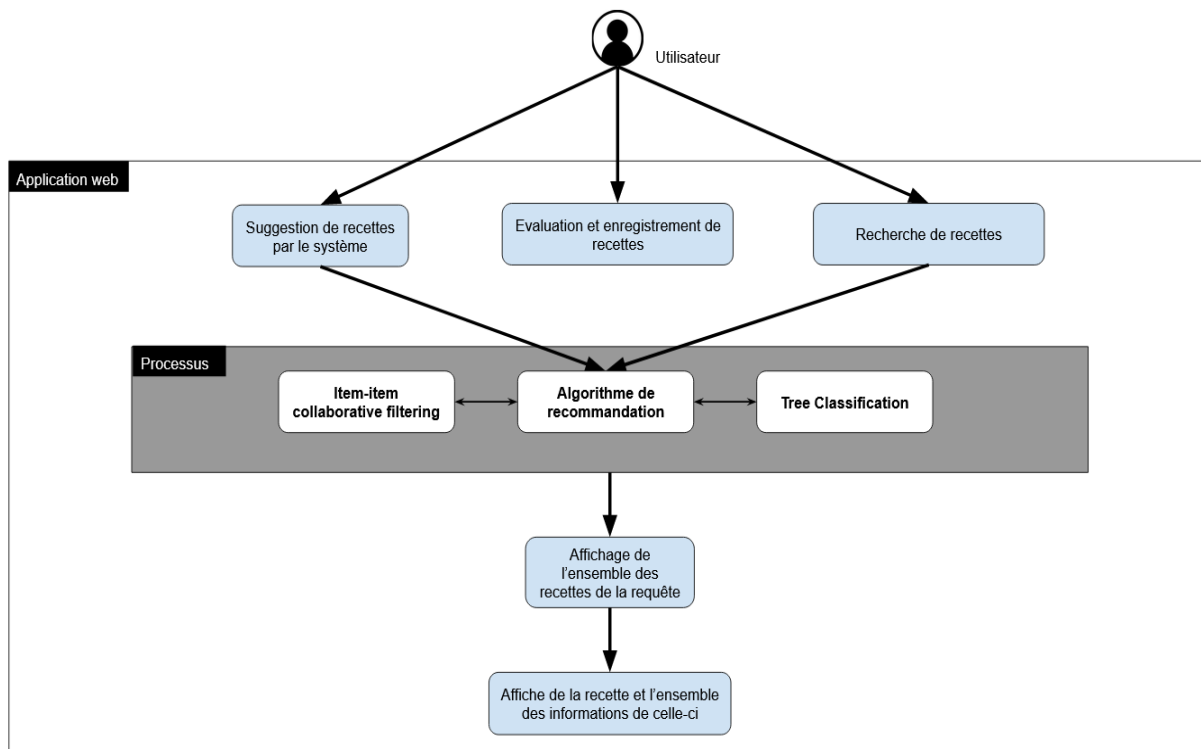


FIGURE 1 – Mise en scénario du projet

Tout d'abord, l'utilisateur a la possibilité d'enregistrer des recettes dans un carnet ainsi que d'évaluer celles-ci. Ces deux fonctionnalités sont indispensables car elles permettront de générer des recommandations selon les préférences de l'utilisateur. Ce type de recommandation est un algorithme de filtrage collaboratif article-article qui a été développé par Amazon en 1998 et qui a joué un grand rôle dans le succès d'Amazon. Lorsque l'utilisateur sera sur l'index de l'application web ou effectuera une recherche particulière (planification de repas, recherche basique), nous exécuterons un processus algorithmique visant à suggérer des recettes les plus pertinentes pour l'utilisateur. Ces recettes seront affichées et pourront être visualisées en détails en cliquant dessus.

Lorsque l'utilisateur procédera à une recherche de recettes par ingrédient, nous utiliserons un arbre de classification pour pouvoir trouver la meilleure catégorie de recette, c'est-à-dire la catégorie dont le score de pertinence par rapport aux ingrédients demandés est le plus élevé.

### 1.3 Objectif du projet

L'objectif de ce projet est de créer une application web proposant une multitude de recettes aux utilisateurs. L'intérêt principal de l'application est de pouvoir suggérer automatiquement des recettes aux utilisateurs selon leurs préférences.

La première étape que nous devons franchir est évidemment d'obtenir les données nécessaires au fonctionnement de cet algorithme de suggestion. Ce jeu de données devra être collecté, nettoyé et structuré de manière uniforme avant qu'une analyse puisse être effectuée. Cette collecte de recette sera faite sur le site web « Marmiton ».

Pour évaluer et classifier nos données, nous utiliserons la méthode de partitionnement de données : K-means clustering qui permettra de trouver des relations entre des recettes appartenant à différentes catégories de repas dans l'espace défini par les ingrédients les plus fréquents. Pour les recommandations de recettes, nous élaborerons un algorithme de type Item-item collaborative filtering ainsi que d'un arbre de classification.

### 1.4 Organisation du rapport

Nous nous consacrerons tout d'abord aux spécifications du projet, c'est-à-dire les fonctionnalités attendues ainsi qu'une description de la conception globale du projet. Cette dernière rendra compte des différentes vues (utilisateur et technique) du projet et pourront mettre en évidence les différentes problématiques qui seront traitées au cours des chapitres techniques. Enfin, une dernière section concernera l'environnement de travail au cours de laquelle sera dressée une liste exhaustive du matériel ainsi que des logiciels et outils utilisés.

Les problématiques soulevées précédemment dans le cahier des charges seront respectivement traitées dans les sections techniques. Ces sections, qui constituent les chapitres techniques auront pour sujet principaux : le Web Scraping (Extracting, Cleaning et Organizing data) des recettes sur le site web "Marmiton" ; l'apprentissage non supervisé (K-mean Clustering) et la classification de notre jeu de données ; l'élaboration de notre algorithme de type Item-item collaborative filtering et enfin la conception de notre arbre de classification pour la recommandation de recettes selon des ingrédients demandés.

Un chapitre concernant le rendu final figurera également sur ce rapport. Il aura pour finalité de présenter le résultat attendu du projet ainsi que des différents outils que l'utilisateur sera amené à manipuler pour faire fonctionner l'application web. Plusieurs tests seront mis en place pour témoigner du bon fonctionnement des différentes fonctionnalités.

Enfin, un chapitre sera consacré à la gestion du projet dans laquelle figureront les méthodes de travail employées ainsi que la répartition des tâches au cours du projet. Nous concluons ce rapport par une explication du déroulement du projet, du travail réalisé et des améliorations possibles à effectuer.

## 2 Présentation et spécification du projet

Ce chapitre présente le cahier des charges du projet et les spécifications techniques, il contient les principaux éléments nécessaires pour comprendre la conception technique et l'architecture du projet.

### 2.1 Fonctionnalités attendues

Notre application web permet à l'utilisateur d'effectuer des recherches de recettes de deux manières différentes :

- Recherche par mots-clefs
- Recherche par ingrédients (avec / sans ingrédients)

Le système devra générer un algorithme de suggestion propre à chaque moyen de recherche. Si la recherche du client est faite par mots-clefs alors on récupèrera les recettes dont le score de pertinence avec les mots-clefs est non nul puis nous trierons les résultats selon les scores des recettes. Dans le cas de la recherche par ingrédients, nous utiliserons un arbre de décision qui nous permettra de trouver les recettes similaires aux critères de l'utilisateur.

L'utilisateur peut aussi enregistrer, évaluer et commenter une recette. La note donnée à une recette est comprise entre 1 et 5 étoiles et le commentaire d'une recette ne peut être fait seulement si l'utilisateur a évalué la recette. L'ensemble des recettes pourront être affichés en détails afin de visualiser les caractéristiques des recettes (temps de cuisson, étapes de préparations etc...).

Lorsque l'utilisateur est sur l'index, le système générera un algorithme de type Item-item collaborative filtering permettant de proposer à l'utilisateur des recettes selon ce qu'il a visualisé, évalué et enregistré.

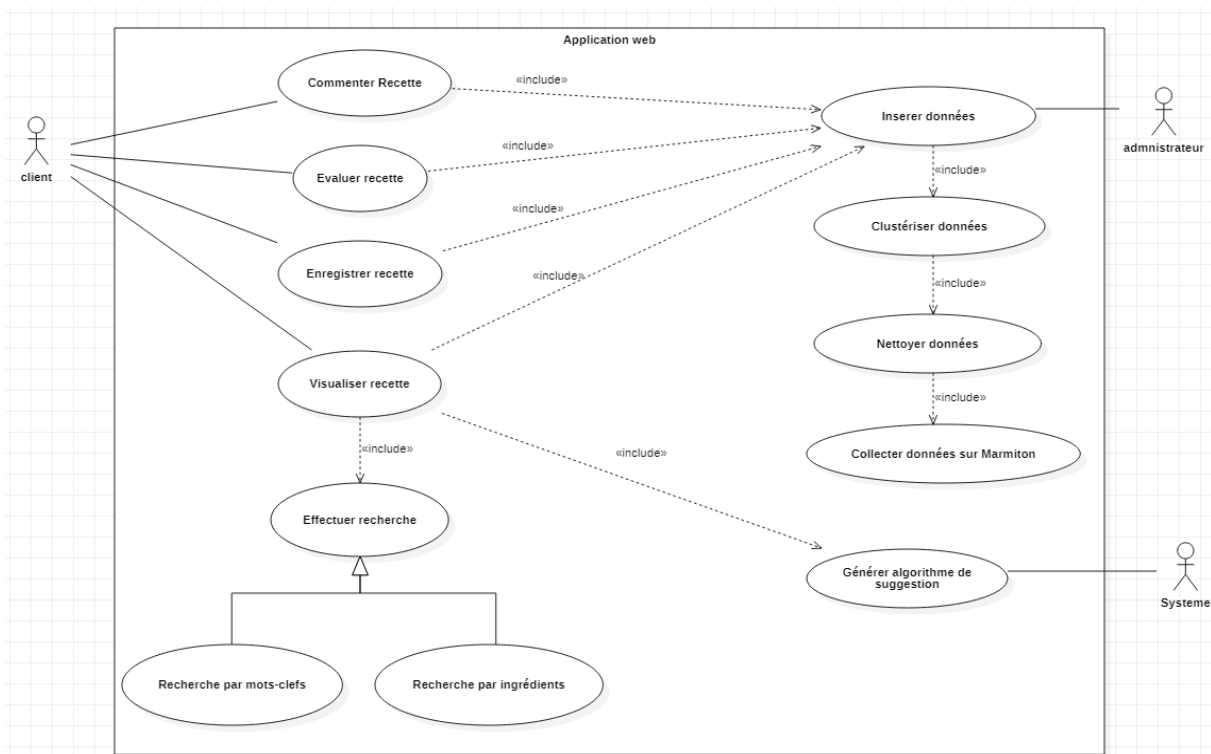


FIGURE 2 – Diagramme de cas utilisation de l'application web

## 2.2 Conception globale du projet

### 2.2.1 Vue pour l'utilisateur

### 2.2.2 Architecture technique

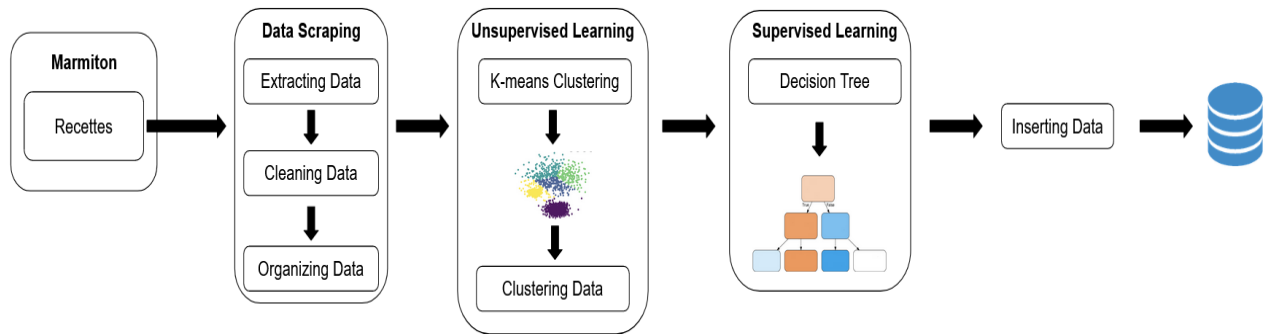


FIGURE 3 – Architecture base de données

## 2.3 Problématiques identifiées et solutions envisagées

## 2.4 Environnement de travail

### 3 Chapitre technique : webscraping

#### 3.1 Analyse de la problématique

#### 3.2 Etat de l'art : études des solutions existantes

#### 3.3 Solution proposée et sa mise en œuvre

### 4 Chapitre technique data clustering

#### 4.1 Analyse de la problématique

#### 4.2 Etat de l'art : études des solutions existantes

#### 4.3 Solution proposée et sa mise en œuvre

### 5 Chapitre technique algorithme de suggestion

#### 5.1 Analyse de la problématique

#### 5.2 Etat de l'art : études des solutions existantes

#### 5.3 Solution proposée et sa mise en œuvre



## **6 Rendu final**

### **6.1 Interface utilisateur finale**

### **6.2 Tests utilisateur et certification**

### **6.3 Autres tests et certifications**

## 7 Gestion de projet

### 7.1 Méthode de gestion

### 7.2 Répartition de tâches

## 8 Conclusion et perspectives

### 8.1 Conclusion

### 8.2 Perspectives

## Références

- [HLR02] L. M. Haas, E. T. Lin, and M. A. Roth. Data integration through database federation. *IBM Syst. J.*, 41(4) :578–596, 2002.
- [HP11] John L. Hennessy and David A. Patterson. *Computer Architecture, Fifth Edition : A Quantitative Approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 5th edition, 2011.