

# **Exploitation des biais de choix alimentaires pour recommander des recettes plus saines**

---

Membres  
KRISNI Almehdi  
ARICHANDRA Santhos

# Plan de la présentation

1. Introduction au papier
2. Récupération des données
  - a. Le site Allrecipes
  - b. Des recettes problématiques
3. Premier aperçu des données
  - a. Structure
  - b. Visualisation
4. Mise en place de l'expérience utilisateur
5. Analyse des résultats
  - a. Première étude
  - b. Deuxième étude
6. Dans la tête de l'utilisateur
  - a. Différents modèles, différents attributs, différents résultats
  - b. Le meilleur résultat pour la meilleure prédiction
7. Conclusion

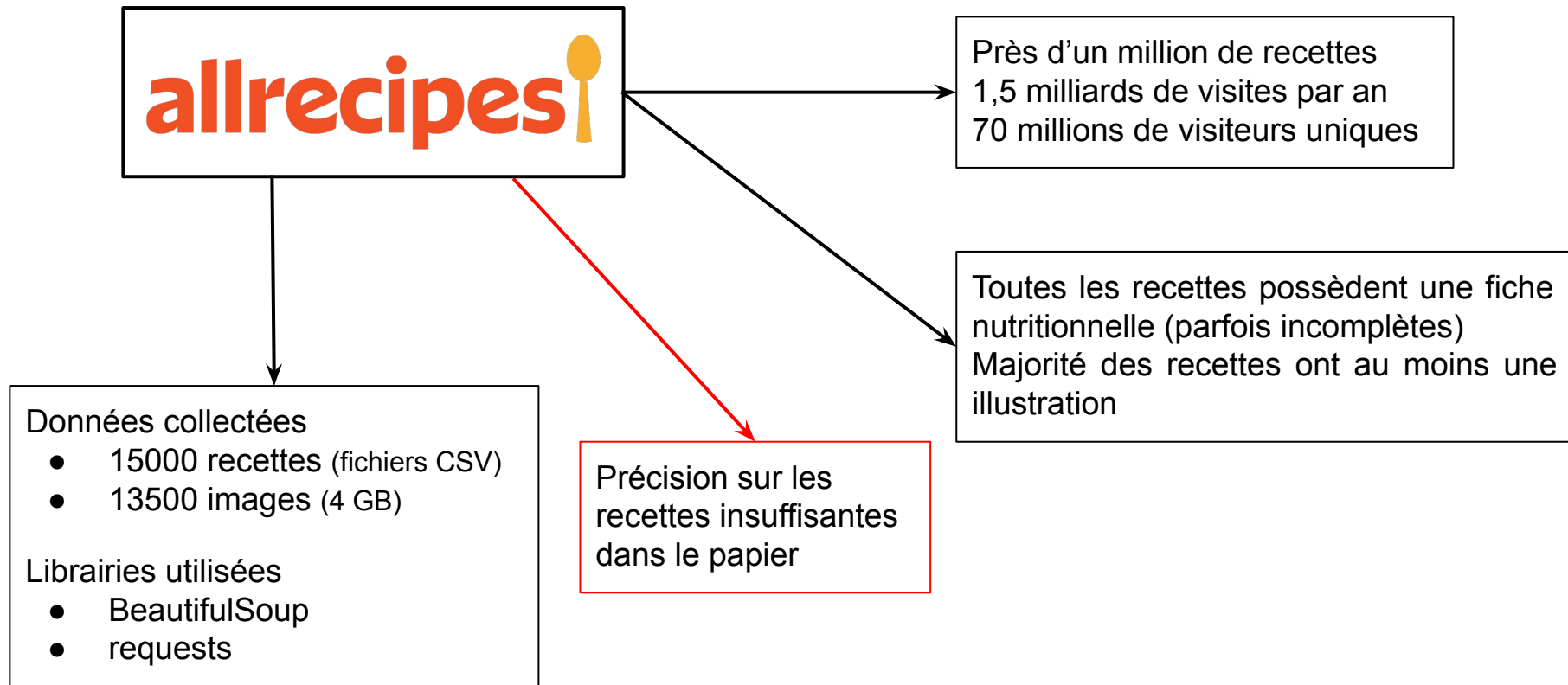


# Introduction au papier

- Système de recommandations modernes proposent le plus souvent des recettes peu saines
- Incorporer le caractère sain des recettes dans les systèmes de recommandation
- Élément contextuel majeur souvent ignoré, l'utilisateur, ses connaissances et ses préférences
- Tendance à favoriser des recettes moins saines et importance des sentiments
- En raison du train de vie quotidien infernal, on dispose d'un temps limité pour se décider
- Préférence pour les recettes appétissantes et souvent contenant plus de gras
- On cherche à prédire les choix des utilisateurs à travers les études d'expériences

# Récupération des données

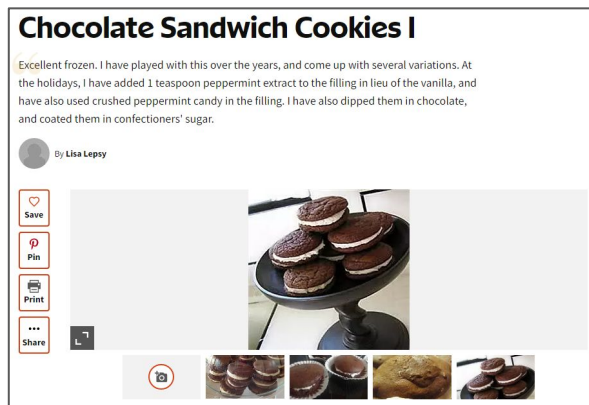
Le site Allrecipes



# Récupération des données

## Des recettes problématiques

- Illustration manquante ou parfois cachée
- Certains identifiants n'existent pas
- Valeurs parfois invraisemblables
- 1 portion = **100 grammes ?**



Exemple de recette utilisable



Exemples de recettes problématiques

# Premier aperçu des données

## Structure

### Données des recettes :

- fichiers CSV contenant 100 recettes chacun
- valeurs nutritionnelles (calories, gras, ...)
- adresse web de l'image

Concaténation de listes de valeurs pour former le dataframe global

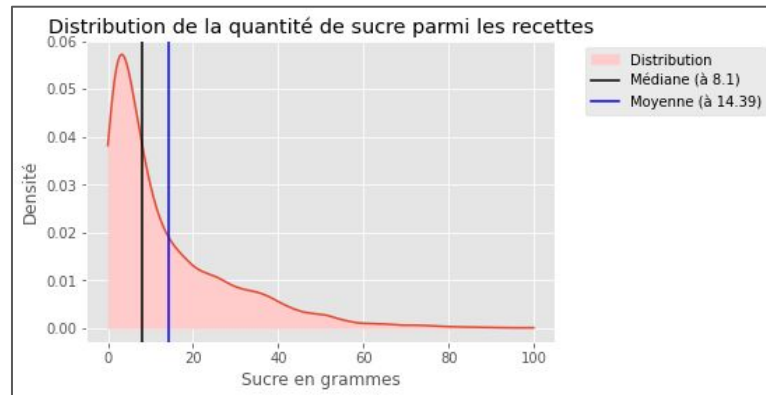
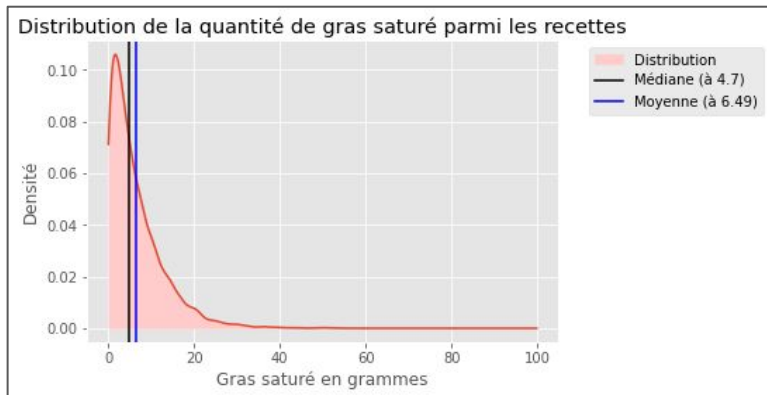
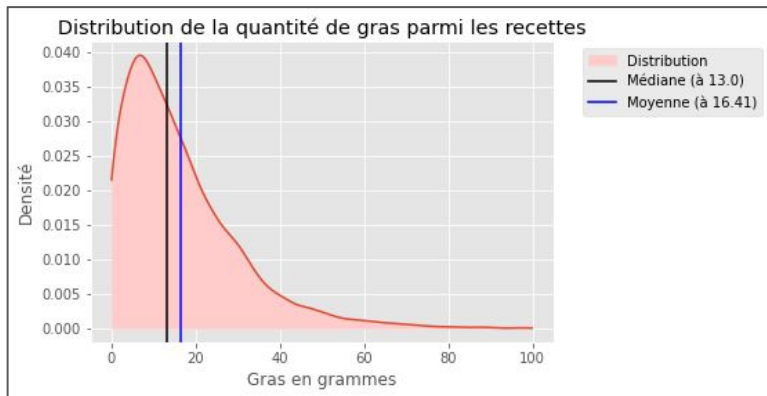
### Données des images :

- répertoire d'images téléchargées
- features (entropie, luminosité, ...)
- fichiers CSV contenant les features de 100 images chacun



# Premier aperçu des données

## Visualisation

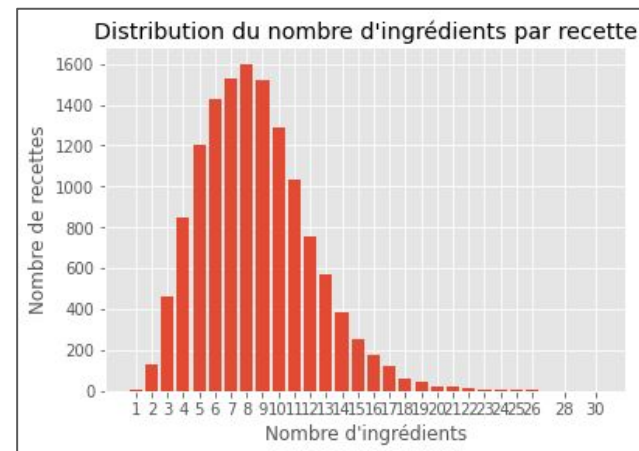
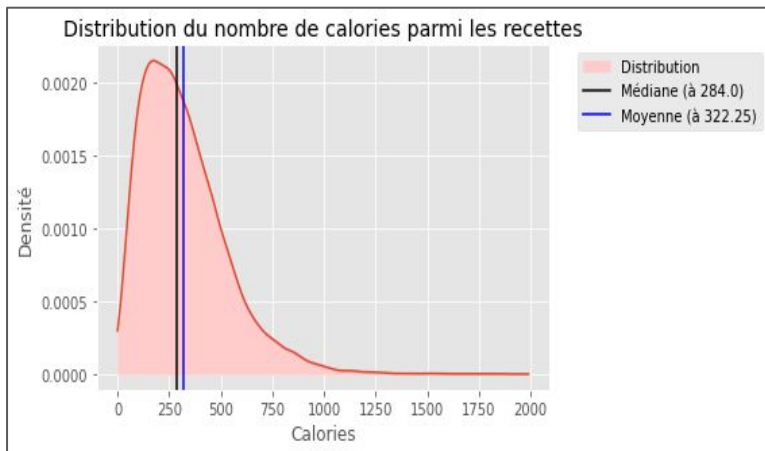
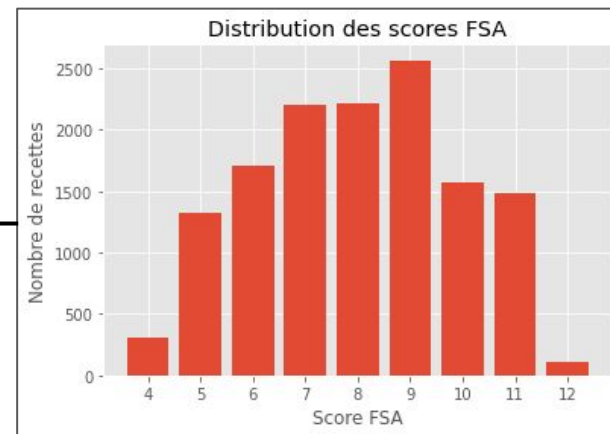


# Premier aperçu des données

## Visualisation

### Score FSA

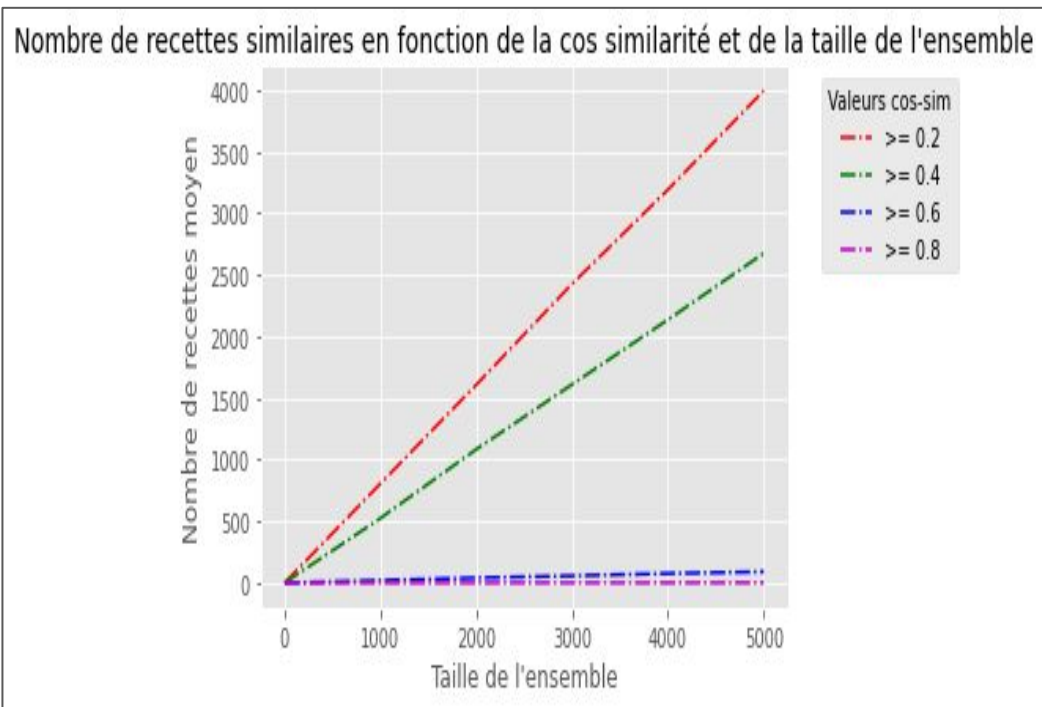
- Mesure de la “healthiness” d’une recette
- Une des limites du papier
- Manque d’information sur Internet
- Calcul réalisé approximatif





# Premier aperçu des données

## Visualisation



$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

Similarité calculée à partir :

- titre
- valeurs nutritionnelles
- avis utilisateurs
- ingrédients

En moyenne, 30 recettes similaires dans la base de données par recette

# Mise en place de l'expérience utilisateur



A

ou



B

Recherche de paires de recettes A et B fortement similaires



Récupération des images, des titres et des ingrédients



## Question

Entre la recette A et B, laquelle contient le plus de gras ?



### Première étude

Utilisateur se base uniquement sur les images ou les titres



A

ou



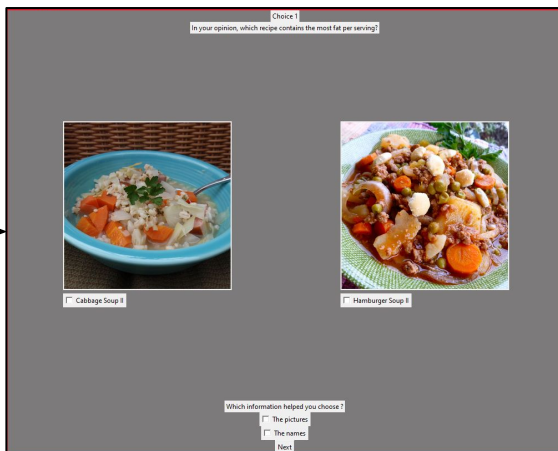
B



### Deuxième étude

Utilisateur se base uniquement sur les images, les titres ou les ingrédients

# Mise en place de l'expérience utilisateur



```
PS D:\Almehd\Documents\UPMC\WU_DACS2\RIITAL\ProjetR15> .\userExperience.py
Welcome to the User Experience Sequence Interface.

Please select a username :
(You may press Enter without giving a name in case you prefer to remain anonymous)
Thanks

Would you like to have the recipes' ingredient lists shown ?
(0 - No, 1 - Yes)
0

Please select a model from the existing ones :
> 1
> 10
> 11
> 12
> 13
> 14
> 15
> 16
> 17
> 18
> 19
> 2
> 20
> 3
> 4
> 5
> 6
> 7
> 8
> 9

Please enter a model ID :
1

The user experience will soon begin. A window should appear on your screen.
```

Les résultats sont par  
la suite analysés  
pour le Machine  
Learning

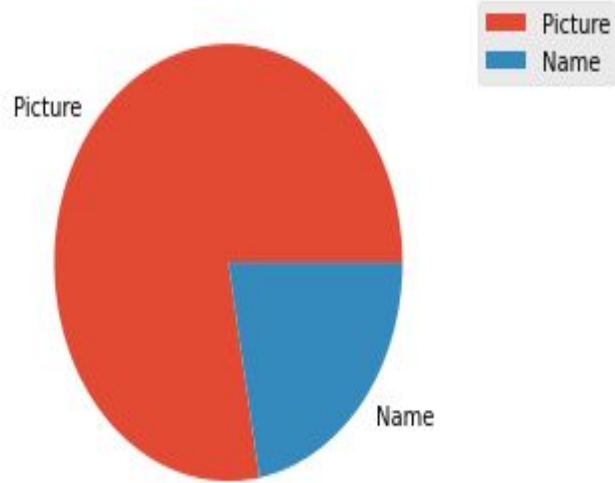
question_id	recipe_A_id	recipe_B_id	recipe_choice	information_choice	expected_answer
0	24078	22921	1	2	1
1	19937	20356	1	1	2
2	15510	24605	1	1	2
3	23430	13495	2	1	2
4	15136	17869	1	2	1
5	12953	17828	1	2	2
6	15538	22592	2	2	2
7	20900	20701	1	1	2
8	24472	14994	1	1	2
9	19836	17866	1	2	1

# Analyse des résultats

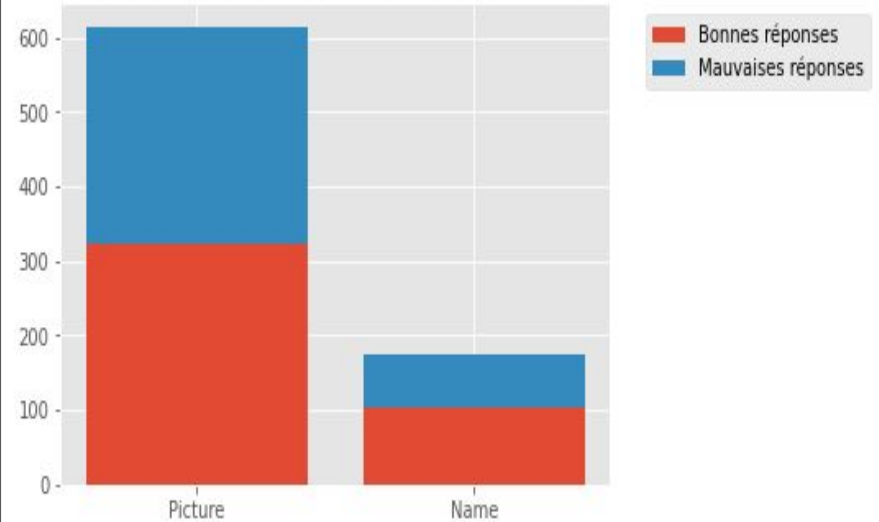
Première étude

**Précision - 54%**

Répartition du choix d'information (première étude)



Nombre de bonnes et mauvaises réponses en fonction du choix de l'information (première étude)



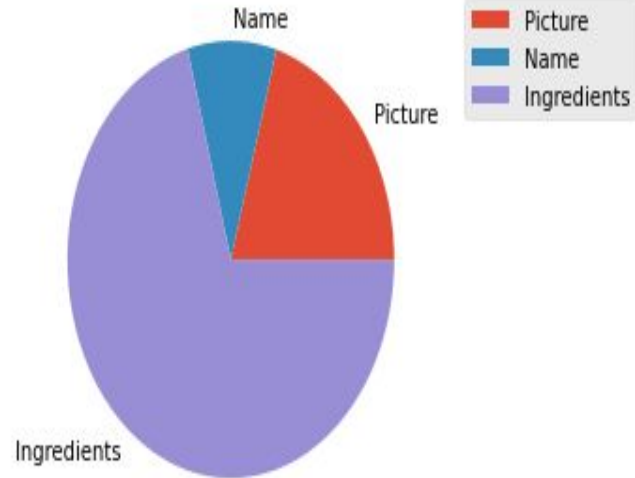
Résultats en accord avec le papier - La photo est l'élément de décision utilisé dans la majorité des cas

# Analyse des résultats

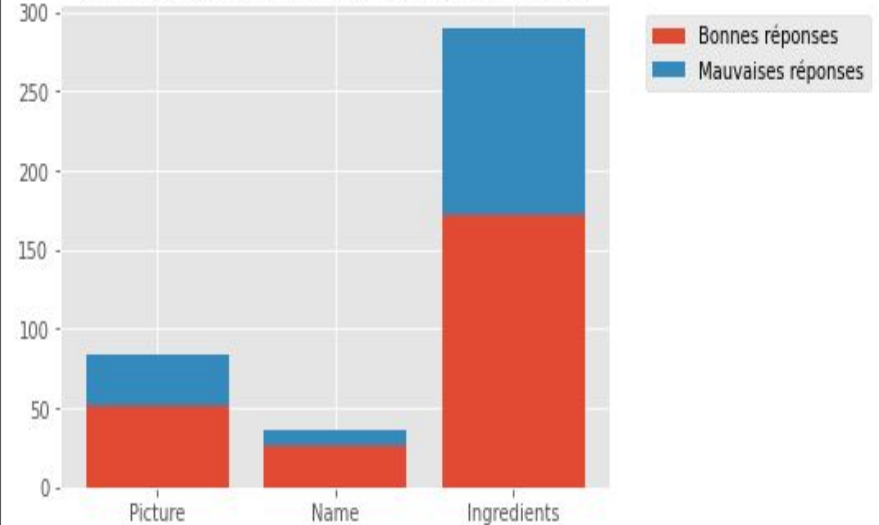
## Deuxième étude

Précision - 61%

Répartition du choix d'information (deuxième étude)



Nombre de bonnes et mauvaises réponses en fonction du choix de l'information (deuxième étude)



Résultats en accord avec le papier - La liste d'ingrédients est l'élément de décision utilisé dans la majorité des cas

# Dans la tête de l'utilisateur

Différents modèles, différents attributs, différents résultats

Feature Set	Rand.For.	Logistic	Naive Bay.	Nombre Feat.
Titre	56 %	53 %	41 %	54
Image	59 %	55 %	53 %	10
Ingrédients	58 %	53 %	53 %	54
Nutrition	57 %	52 %	52 %	12
Total	34 %	34 %	33 %	130

- Modèles entraînés sur les résultats de la première étude
- On retrouve les images comme élément de décision majeur

Résultats sur les données de la première étude

# Dans la tête de l'utilisateur

Différents modèles, différents attributs, différents résultats

Feature Set	Rand.For.	Logistic	Naive Bay.	Nombre Feat.
Titre	50 %	52 %	44 %	54
Image	60 %	50 %	48 %	10
Ingrédients	66 %	69 %	50 %	54
Nutrition	62 %	57 %	58 %	12
Total	35 %	35 %	31 %	130

- Modèles entraînés sur les résultats de la deuxième étude
- On retrouve les ingrédients comme élément de décision majeur

Résultats sur les données de la deuxième étude

# Dans la tête de l'utilisateur

Le meilleur résultat pour la meilleure prédiction

- Meilleur classifieur = Régression Logistique (ingrédients, entraîné sur la deuxième étude)
- On s'en sert pour prédire les choix de la majorité des utilisateurs
- D'après les résultats, on peut garantir 3 recettes sur 5





# Dans la tête de l'utilisateur

Le meilleur résultat pour la meilleure prédiction

- Meilleur classifieur = Régression Logistique (ingrédients, entraîné sur la deuxième étude)
- On s'en sert pour prédire les choix de la majorité des utilisateurs
- D'après les résultats, on peut garantir 3 recettes sur 5



# Conclusion

- Reproduction du papier simple au niveau informatique (code, affichages, ...)
- Manque de précision par rapport aux données utilisées
- Extraction de features pour les images pas détaillée (entropie utilisée, ...)
- Ressources inaccessibles ou trop anciennes (FSA score, ...)
- Moins d'utilisateurs mais plus de données par utilisateurs (manque de diversité)
- Préférences sont fortement liées à la première impression
- Résultat peut être considéré comme décevant
- Choix de la recette relève plus de la chance ...

The image features three concentric circles centered on a black background. The circles are rendered in a dark gray color, creating a subtle gradient effect. The text "The End" is written in a white, elegant script font, positioned horizontally across the middle of the circles.

*The End*