

Anatomie d'une salade

La cuisine est plus qu'un moyen de satisfaire son appétit, elle permet de manger mieux et avec plus de plaisir. Les salades sont des plats classiques et déclinées dans toutes les cultures. Aimant cuisiner, j'ai choisi d'analyser ce type de plat qui semble universel et pourtant si simple.

La cuisine est l'art de transformer divers produits bruts en un seul plat qui les rassemble. À leur façon, les salades réalisent cette transition grâce à leur assemblage d'ingrédients et de saveurs. Mon travail consistera à étudier cette conversion.

Positionnement thématique (ÉTAPE 1) :

- INFORMATIQUE (*Informatique pratique*)
- INFORMATIQUE (*Informatique Théorique*)

Mots-clés (ÉTAPE 1) :

Mots-clés (en français)	Mots-clés (en anglais)
-------------------------	------------------------

Salade

Salad

Goût

Taste

Algorithmes d'apprentissage

Machine Learning

Base de donnée

Data base

Automatisation

Automation

Bibliographie commentée

Selon Paul Bocuse, un des plus grands chefs cuisiniers du XXe siècle, « Les deux secrets d'un succès : la qualité et la création » [6]. En effet, la créativité humaine est au cœur de la cuisine. Ainsi, pour étudier le fonctionnement culinaire, il faut se confronter à la multiplicité de possibilités qu'offre la création.

Deux méthodes semblent être possibles pour analyser ce type de situation : une étude chimique et biologique, qui examinerait les interactions moléculaires et la perception du goût, ou une étude fondée sur l'apprentissage à partir d'un ensemble de données. Deux conditions essentielles doivent être réunies pour cette seconde approche : l'analyse des données doit permettre de converger vers les habitudes culinaires réelles et l'échantillon étudié doit être suffisamment grand pour assurer cette convergence.

La loi des grands nombres, formulée par Jakob Bernoulli en 1713 [4], se traduit par le fait que, lorsque l'on augmente la taille des données, alors le résultat de leur étude converge vers une situation représentative du lot de données. Ce principe justifie bien que l'étude de données mène à s'approcher de la réalité. Ainsi, en se basant sur une large base de données, on peut étudier les caractéristiques générales d'un plat et donc faire apparaître sa structure.

La convergence de l'étude étant acquise, c'est la théorie des algorithmes d'apprentissage qui permet de réaliser ce travail d'étude de données et d'évaluation des caractéristiques du plat. À partir des données, comment faire apparaître la compatibilité des ingrédients entre eux ? Combien d'ingrédients sont nécessaires ou suffisants pour créer une salade ? Comment réussir à démontrer que certains ingrédients servent de bases et d'autres de garnitures ? Comment évaluer si une nouvelle recette est bien une salade ou non ? Voilà un ensemble des problèmes qui doivent être résolus pour obtenir une analyse fine de ces plats. Certains de ces questionnements sont directement résolus par de célèbres approches algorithmiques : l'algorithme des k plus proches voisins [1] pour évaluer si une nouvelle recette est une salade à partir de données ou encore l'algorithme de Lloyd [2] qui répond au problème des k moyennes pour réussir à cartographier les ensembles de recettes de salades et d'ingrédients.

Cependant, un lot de données important entraîne des difficultés techniques de stockage et d'exploitation. Les bases de données relationnelles sont une solution efficace pour travailler sous ces difficultés techniques. Pour travailler à l'intérieur de ces bases, ordonnées logiquement et imaginées par Edgar Frank Codd [3], il est inventé le langage Structured Query Language (SQL) par Donald Chamberlin et Raymond Boyce [5]. Ainsi, les algorithmes d'apprentissage seront en mesure d'accéder efficacement aux données.

C'est en alliant la théorie des algorithmes d'apprentissage et la technologie des bases de données relationnelles que l'on peut créer un modèle de recette de salade qui sera cohérent.

Problématique retenue

Mon sujet a pour but d'étudier les caractéristiques des salades afin d'élaborer un modèle capable de générer de nouvelles recettes de salades à partir d'ingrédients spécifiques.

Objectifs du TIPE du candidat

1. Réaliser différents graphes pour représenter les liens entre ingrédients et entre les différents types de recettes.
2. Déterminer les caractéristiques des recettes de salades
3. Créer un modèle pour générer de nouvelles recettes de salades cohérentes à partir d'ingrédients clefs.

Références bibliographiques (ÉTAPE 1)

- [1] EVELYN FIX AND J. L. HODGES, JR. : Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties : *International Statistical Review / Revue Internationale de Statistique*, vol. 57, no. 3, 1989, pp. 238–47. <https://doi.org/10.2307/1403797>
- [2] STUART LLOYD : Least squares quantization in PCM : *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129-137, March 1982, <https://doi.org/10.1109/TIT.1982.1056489>
- [3] EDGAR FRANK CODD : A Relational Model of Data for Large Shared Data Banks : *Communications of the ACM*, 1970, 13, 377-387. <https://doi.org/10.1145/362384.362685>
- [4] JAKOB BERNOULLI : *Ars Conjectandi* : 1713
- [5] CHAMBERLIN, DONALD D. AND RAYMOND F. BOYCE : SEQUEL: A structured English query language. : <https://doi.org/10.1145/800296.811515>
- [6] PAUL BOCUSE : <http://evene.lefigaro.fr/citations/paul-bocuse>