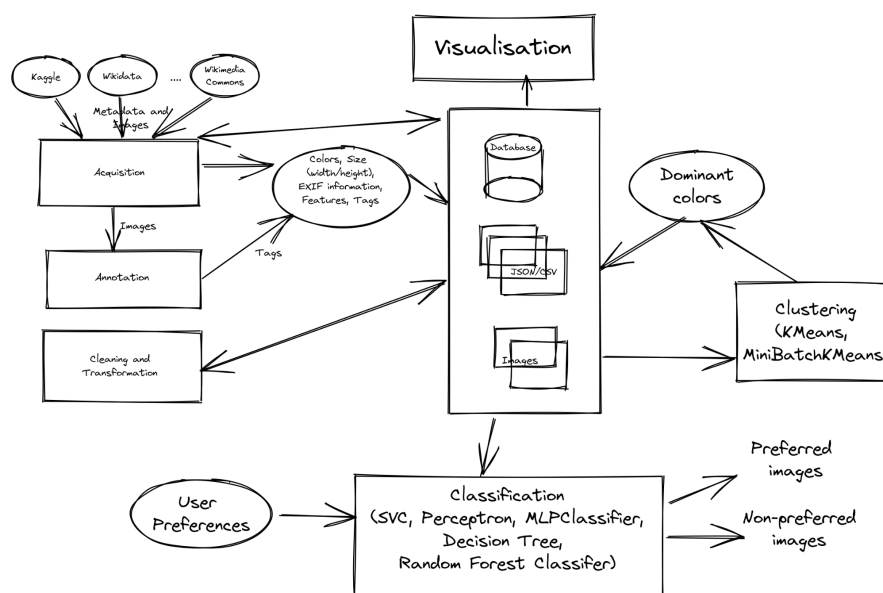


Data Mining - Recommandations d'images

Le but de votre projet	1
Sources des données(images) et de leurs licences	2
Taille des données	2
Informations stockées pour chaque image	2
Informations concernant les préférences de l'utilisateur	3
La visualisation des données	3
Les modèles d'exploration de données et/ou d'apprentissage machine que vous avez utilisés avec les métriques obtenues	4
L'auto-évaluation de votre travail	4
Remarques concernant les séances pratiques, les exercices et les possibilités d'amélioration	4
Conclusion	5

Le but de votre projet



Le projet a pour objectif de développer un système de recommandation d'images personnalisé en fonction des préférences de l'utilisateur.

Nous devons construire ce système en automatisant toutes les tâches liées à l'acquisition, l'annotation, l'analyse et la visualisation des données.

Les principales tâches du projet comprennent la collecte de données, l'étiquetage et l'annotation, l'analyse de données, la visualisation des données, la création d'un système de recommandation et les tests.

Sources des données(images) et de leurs licences

Les images sont récupérées via l'api de qwikidata. Les images disponibles sur WikiData sont libres de droit. Cette library permet d'envoyer des requêtes SQL, pour obtenir les données sous format json. On utilise ensuite request pour télécharger les images dans le répertoire ./images/ .

Nous téléchargeons 100 images de 4 thèmes différents : couleur, automobile de sport, décoration militaire et fruit.

Taille des données

Les images font environ 200 MB au total.

Précédemment, nous avions des images des images du thème tableau et drapeau, hors les drapeaux étaient généralement sous format .svg ce qui complexifie la lecture des couleurs et les tableaux avaient une taille beaucoup trop grande, les 100 images prenaient plus de 500MB.

Informations stockées pour chaque image

Afin de pouvoir ultérieurement annoter nos images, nous avons décidé de stocker les informations suivantes dans un fichier json:

Information	Type
name	string
format	string
width	int
height	int
date	date
make	string
model	string
orientation	string
couleur	string

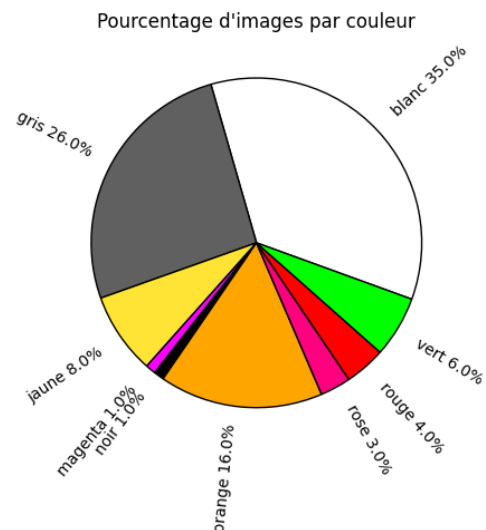
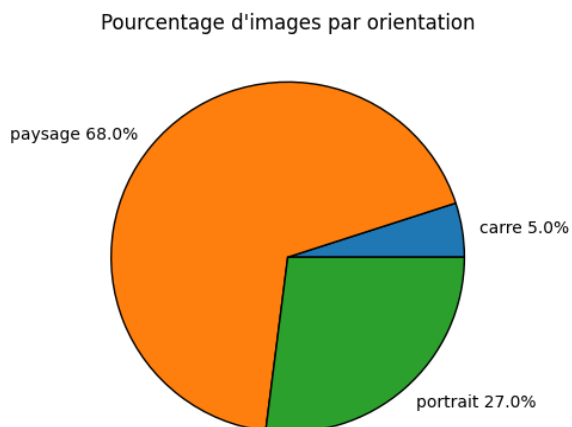
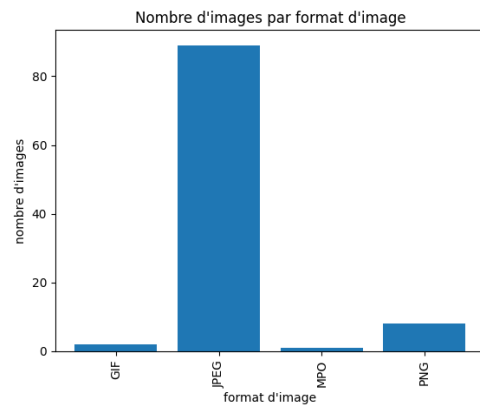
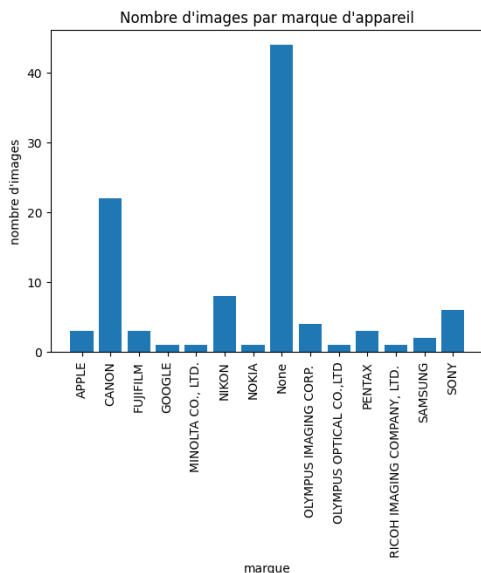
Informations concernant les préférences de l'utilisateur

Pour annoter nos images, nous avons finalement sélectionné 3 tags. Concernant les préférences de l'utilisateur en matière de tags, celui-ci peut choisir le nombre de critères qu'il souhaite observer, allant de 1 à 3 parmi la marque, l'orientation et la couleur.

Nous avons sauvegardé les préférences d'un profil dans un fichier JSON ce qui permet de les stocker de manière structurée et de les récupérer facilement pour personnaliser l'expérience utilisateur.

La visualisation des données

Pour savoir quel tag était le plus présent dans notre échantillon d'images, nous les avons affichés via des graphiques bâton et camembert. Cela nous permet de visualiser les tags les plus importants dans notre échantillon.



Les modèles d'exploration de données et/ou d'apprentissage machine que vous avez utilisés avec les métriques obtenues

Pour obtenir les tag favorites de l'utilisateur, nous avons parcourus les images favorites de l'utilisateur; obtenus au préalable en lui soumettant des images aléatoires. Pour cela, nous avons utilisé des tags remarquables apparaissent de nombreuses fois susceptible de permettre l'apparition d'une tendance. Nous avons comparé ces tags à ceux présents sur les images aimées et nous avons compté leur présence. Enfin, le tag pour chaque catégorie ayant le plus d'itération est choisi comme tag favori.

Après ce traitement, nous soumettons à l'utilisateur des images possédant les mêmes tags. Selon le choix de l'utilisateur, les images peuvent avoir 1,2 ou 3 tags favoris en commun avec les tags favoris. Et on affiche pour chaque images recommandées, les tags/couples de tags présents dans l'image qui on fait qu'elle a été choisie.

L'auto-évaluation de votre travail

Dans la partie classification de notre travail, nous avons rencontré des difficultés lors de l'implémentation du modèle DecisionTreeClassifier. Nous avons essayé plusieurs fois de l'implémenter en utilisant différentes techniques et paramètres, mais nous n'avons pas obtenu de résultats pertinents. C'est pourquoi nous avons décidé de passer à une approche de recherche manuelle pour trouver les meilleurs paramètres pour notre modèle.

La recherche manuelle nous a pris beaucoup de temps et d'efforts, mais elle s'est finalement avérée efficace. Grâce à cette approche, nous avons pu obtenir un modèle performant pour la classification de nos données.

Remarques concernant les séances pratiques, les exercices et les possibilités d'amélioration

Il y a beaucoup de travail à faire chez soi, et le nombre limité de séances ne nous permet pas de tout terminer pendant les TP. La partie sur le machine learning est la plus difficile et il est difficile d'obtenir de l'aide en dehors des heures de TP. Il serait peut-être utile d'ajouter plus de séances ou de consacrer plus de temps à l'explication de l'application du machine learning (comme l'utilisation de sklearn, par exemple). Les explications théoriques et mathématiques sont suffisantes, mais nous aurions besoin d'une aide supplémentaire pour la mise en pratique.

Conclusion

Dans l'ensemble, ce projet de recommandation d'images a été une expérience très enrichissante pour nous. Nous avons rencontré plusieurs difficultés tout au long du projet,

notamment dans la mise en place de l'algorithme de classification. Cependant, nous avons acquis des compétences pratiques en matière de collecte de données, d'annotation, d'analyse et de visualisation, ainsi que dans l'utilisation d'algorithmes d'apprentissage automatique pour créer un système de recommandation personnalisé.