

Rapport projet Machine Learning

| Josse DE OLIVERA & Léo BARBIER

L'objectif de ce projet est de recommander des images en fonction des préférences de l'utilisateur grâce à un système de recommandation en python. Les tâches liées à l'acquisition, l'annotation, l'analyse et la visualisation des données sont automatisées.

Nous avons décidé de réaliser ce projet avec des images de casinos.

Le projet se déroule avec plusieurs étapes :

1. Collecte de données
2. Étiquetage et annotation
3. Analyses de données
4. Visualisation des données
5. Système de recommandation
6. Tests

1. Collecte de données :

Nous récupérons les images de casinos sur Wikidata, nous avons choisi de récupérer les 128 premières images sur les 424 disponibles pour réduire la taille de l'échantillon et réduire la taille de stockage requise pour les enregistrer.

Cela représente environ 420 Mo de données pour les photos des 128 casinos.

Ces images sont censées être libres de droits car sur les plateformes de Wikimedia dont Wikidata fait partie, les utilisateurs des services doivent mettre en ligne que des images libres de droits. Elles sont donc normalement sous la Creative Commons CC0 License.

Nous récupérons ensuite les métadonnées de chaque image en utilisant les informations Exif. Elles sont introduites dans un fichier JSON global par la suite en même temps que l'étiquetage et annotation.

2. Étiquetage et annotation :

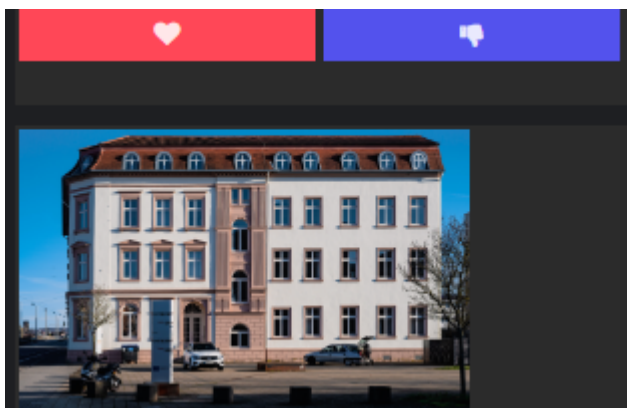
En plus des informations Exif des images nous avons décidé d'enregistrer les 3 couleurs prédominantes automatiquement en analysant l'image. Nous prenons aussi des tags qui sont entrés par nous via un script qui affiche l'image et attend l'entrée de l'utilisateur et stocke dans un fichier par image les tags.

Les informations Exif, les couleurs prédominantes et les tags forment ensemble un fichier JSON qui rassemble toutes les images.

Pour les 128 images nous avons donc un fichier JSON de 170Ko contenant les informations qui vont nous permettre d'analyser les données et qui vont nous permettre de recommander des images aux utilisateurs en fonction de leurs préférences.

3. Analyses de données :

Pour analyser les préférences des utilisateurs, nous affichons une image avec un bouton like et un bouton dislike.



Nous demandons à l'utilisateur son avis sur 10 images pour avoir assez d'informations pour pouvoir lui proposer d'autres images de casinos qui logiquement devrait aimer.

Grâce au choix de l'utilisateur on va connaître ses couleurs, ses tags, et ses Exif (à préciser lesquels on utilise) favoris

4. Visualisation des données
5. Système de recommandation
6. Tests

L'auto-évaluation de notre travail :

1. Collecte de données
 - a. Approches automatisées de la collecte de données
 - b. Utilisation d'images sous licence libre
 - c. Stockage et gestion des images et des métadonnées associées
2. Étiquetage et annotation
 - a. Approches automatisées de l'étiquetage
 - b. Stockage et gestion des étiquettes et des annotations des images
 - c. Utilisation d'algorithmes de classification et de regroupement
3. Analyses de données
 - a. Types d'analyses utilisées
 - b. Utilisation de Pandas et Scikit-learn
 - c. Utilisation d'algorithmes d'exploration de données
4. Visualisation des données
 - a. Types de techniques de visualisation utilisées
 - b. Utilisation de matplotlib
5. Système de recommandation
 - a. Stockage et gestion des préférences et du profil de l'utilisateur
 - b. Utilisation d'algorithmes de recommandation
6. Tests

- a. Présence de tests fonctionnels

- b. Présence de tests utilisateurs

7. Rapport

- a. Clarté de la présentation

- b. Présence d'une introduction et d'une conclusion claires, architecture des diagrammes, un résumé des différentes tâches réalisées et des limites

- c. Bibliographie

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

Conclusion :

Bibliographie :