

Rapport de projet DATA MINING

Joseph Pouradier-duteil & Pierre-Louis TELEP

Sommaire

- [Rapport de projet DATA MINING](#)
 - [Sommaire](#)
 - [1. Introduction](#)
 - [2. Présentation du projet](#)
 - [3. Présentation des données](#) **PL**
 - [4. Analyse des données](#) **JO**
 - [5. Prédiction](#)
 - [6. Auto-évaluation](#)
 - [7. Remarques](#)
 - [8. Conclusion](#)

1. Introduction

Ce rapport a pour but de présenter le projet de Data Mining. Le projet a été réalisé par **Joseph POURADIER-DUTEIL** et **Pierre-Louis TELEP**.

Le but de ce projet est de réaliser un système de recommandation. Un système de recommandation est un algorithme qui va analyser les données des images proposées à l'utilisateur. En fonction de si l'utilisateur aime ou pas les images l'algorithme va pouvoir proposer de nouvelles images et prédire si l'utilisateur va aimer ou non cette image. Pour cela nous avons utilisés des images de **Motos,Voitures,Pokemons** et **Exoplanet**.

2. Présentation du projet

Ce projet a pour but de faire un système de recommandation. Pour cela nous avons utilisés des images de **motos,voitures,Pokémons** et **planètes**. Nous avons utilisés des images trouvées en ligne sur WikiData que nous avons ensuite analyser pour récupérer leurs **métadonnées**.

Pour la partie de recommandation nous avons utilisé **SKLearn** et les **DecisionTree** pour analyser les choix de l'utilisateur en fonction des images proposées et prédire des images qu'il aime ou pas. Nous avons fait un deuxieme modèle de recommandation mais basé sur des préférences utilisateur aléatoire.

3. Présentation des données PL

Nous sommes allés sur wikidata pour récupérer les images de **motos**, **voitures**, **Pokémons** et **planètes**. Pour télécharger les images nous avons utilisé un csv contenant les Urls des images.

Pour ensuite télécharger les images nous avons utilisé un script **python** qui va parcourir le csv et générer un dictionnaire contenant le nom de l'image, le dossier où enregistrer l'image, le lien pour télécharger l'image ainsi que certaines métadonnées. Ce même **script** va ensuite télécharger chaque image en utilisant **requests** avec des **get** et les enregistrer dans un dossier.

Une fois que toutes les images étaient téléchargées nous les avons utilisées avec un **script** python pour les mettre les photos au même format. Pour cela nous avons utilisé **PIL** et **numpy** pour les redimensionner au format 16:9, les normaliser et les convertir en **RGB**.

4. Analyse des données JO

Meta data Dataframe gestion des couleurs

Après avoir récupéré toutes les images, nous avons récupéré leurs métadonnées.

Nous avons utilisé les données exif des images tel que sa taille, son orientation quand elle était disponible et le format.

5. Prédiction

6. Auto-évaluation

ff

7. Remarques

ff

8. Conclusion

ff