

BILLET Marianne

CRESPIN Mathis

TRUONG VINH TONG Guillaume

Projet de DataMining



1 : Image provenant du projet

Table des matières

Introduction	3
Sources des images	3
Taille des images.....	3
Informations stockées pour les images	4
Informations préférences de l'utilisateur.....	4
Modèles d'exploration de données	4
Auto-évaluation du travail.....	5
Remarques séances/exercices	6
Conclusion	6

Introduction

Le but de notre projet est de proposer à notre utilisateur un ensemble d'images qui correspondent à ce qu'il aime. Pour ce faire, nous lui proposons un ensemble aléatoire de dix images, afin de découvrir ses préférences. Ensuite, nous procédons à un filtrage de notre base de données afin qu'il se retrouve avec principalement des images qu'il pourrait aimer, ainsi que d'autres qui n'ont pas encore de résultat. Notre projet peut être résumé par la figure 2 suivante :

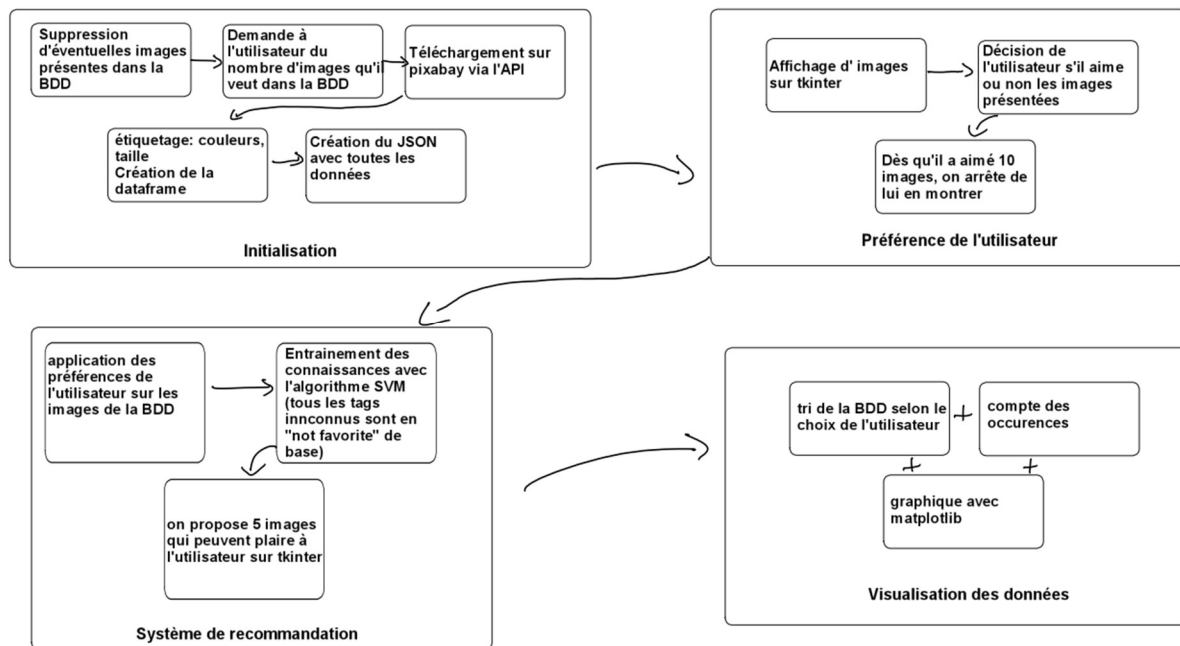


Figure 2: Résumé du projet

Sources des images

Nos images viennent directement de pixabay.com. Nous les avons intégrés dans notre projet en utilisant l'API proposée par ce site, et en filtrant par la recherche « Animaux », afin de n'avoir que des photos de ce style. Nous aurions également pu ne pas filtrer, mais comme notre base de données est assez petite (minimum 100 images), il nous fallait des points communs entre les images pour pouvoir organiser un système de recommandation.

Taille des images

La taille de nos images est assez importante, car nous avons privilégié une bonne qualité afin que l'utilisateur puisse indiquer les images qu'il aime réellement, indépendamment du bruit de l'image.

Informations stockées pour les images

Nous avons dû faire un choix sur les informations que nous allions stockées pour les images.

En effet, le site pixabay.com nous donnait déjà une grande partie d'informations, notamment l'id de l'image, sa taille, son nombre de « likes » et surtout 3 tags par photos. Nous avons décidé de garder les tags, l'id afin de pouvoir construire notre base de données sur cette base, l'URL de l'image, sa longueur et sa largeur, sa taille, son nombre de likes, l'utilisateur qui a publié l'image, le format, et ses 4 couleurs principales.

Informations préférences de l'utilisateur

Pour ce qui est des préférences de l'utilisateur, nous les découvrons en lui proposant des images choisies au hasard parmi notre base de données. L'utilisateur doit alors déclarer s'il les aime ou non, ce qui nous donne une base de ce que l'utilisateur aime, notamment des connaissances sur quelques tags et sur les couleurs qu'il préfère.

Modèles d'exploration de données

Dans notre projet, nous avons utilisé un modèle de recommandation basé sur l'apprentissage machine, en particulier un classificateur SVM (Support Vector Machines). Ce choix s'est avéré judicieux pour notre tâche de recommandation d'images, car les SVM sont efficaces pour la classification binaire et peuvent être adaptés à des problèmes de recommandation en considérant les préférences de l'utilisateur comme des classes à prédire. Nous avons entraîné le SVM sur les tags et les couleurs principales.

Nous avons utilisé des mesures de précision pour évaluer la performance de notre modèle de recommandation basé sur le classificateur SVM. La précision représente la proportion d'images recommandées qui ont été correctement appréciées par l'utilisateur parmi toutes les images recommandées. Pour calculer la précision, nous avons comparé les recommandations du modèle aux réponses réelles des utilisateurs sur un ensemble de données de test. Notre performance est en moyenne de 0.9.

Auto-évaluation du travail

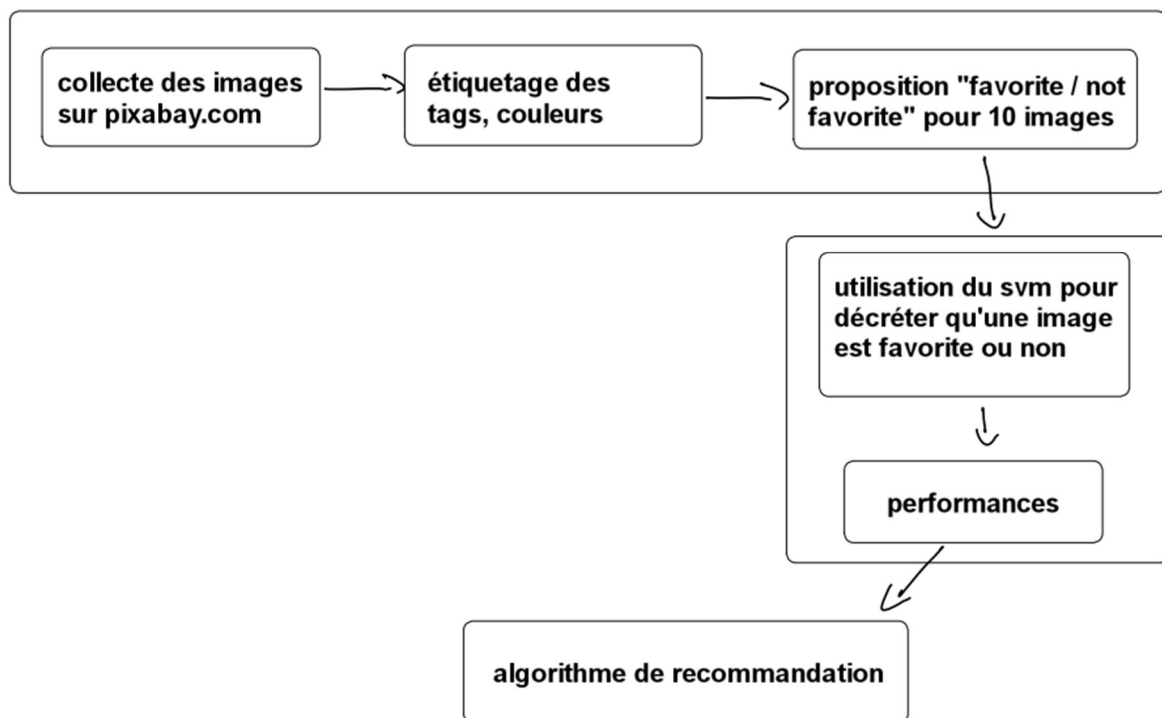


Figure 3 : Résumé du travail effectué

La Figure 2 résume le travail effectué pour notre projet, expliqué avant.

Les points forts de notre travail sont les suivants :

Collecte des images : En utilisant l'API de Pixabay, nous avons pu intégrer une variété d'images correspondant au thème des animaux, offrant ainsi une diversité visuelle à nos utilisateurs.

Fonctionnalité de base mise en place : Nous avons réussi à créer un système de recommandation qui propose des images à l'utilisateur en fonction de ses préférences déclarées.

Collecte et stockage d'informations pertinentes : Nous avons pris soin de collecter et de stocker des informations importantes sur les images, telles que les tags, les likes, les dimensions, et les couleurs principales. Cela nous permet d'avoir une base solide pour recommander des images pertinentes à l'utilisateur.

Nous voyons cependant quelques axes d'amélioration, telles que :

Amélioration de l'algorithme de recommandation : L'intégration d'un clustering des images en complément de la classification SVM représenterait une avancée significative pour notre système de recommandation d'images. En regroupant les

images selon leur similarité, le clustering offrirait une diversité accrue dans les recommandations, tout en permettant la découverte de nouvelles préférences utilisateur. Cette combinaison permettrait une personnalisation plus fine des recommandations en exploitant à la fois les caractéristiques individuelles des images et les tendances globales au sein de la base de données. En résultat, notre système offrirait une expérience utilisateur enrichie, plus diversifiée et mieux adaptée aux besoins et aux goûts spécifiques de chaque utilisateur.

Expansion de la base de données d'images : Avec seulement 100 images dans notre base de données, la variété des recommandations peut être limitée. L'expansion de notre base de données avec plus d'images diverses pourrait améliorer la qualité et la pertinence des recommandations.

Remarques séances/exercices

Les exercices ont été appréciés car pertinents pour montrer l'étendue du datamining. Nous avons cependant relevé qu'il était dommage que les exercices s'apparentent plutôt à de l'analyse de code qu'à de la création de notre propre code.

Pour ce qui est des séances, l'environnement bruyant était particulièrement déconcentrant, il nous arrivait à avoir du mal à travailler correctement.

Conclusion

En conclusion, notre projet de recommandation d'images basé sur les préférences des utilisateurs est sur la bonne voie, offrant une première version fonctionnelle du système. Toutefois, des améliorations significatives restent à apporter pour optimiser la précision et la pertinence des recommandations. En poursuivant notre travail et en tenant compte des axes d'amélioration identifiés, nous visons à fournir une expérience utilisateur encore plus enrichie et personnalisée.

Bibliographie

Les codes donnés par M. Jonh Samuel ont été notre seule source d'inspiration.