LU3IN026 — Projet

Analyse de données de Google PlayStore Apps

Ung Thierry

Introduction

Ce projet s’intéresse à l’analyse de données de Google PlayStore Apps obtenues à partir de la base de donnée de Kaggle.

Ce projet s’intéresse à l’analyse de données de Google PlayStore Apps obtenues à partir de la base de donnée de Kaggle.

On s’intéresse à la :

* classification supervisée des avis laissés par les utilisateurs

. classification non-supervisée de facteur qui permette de prédire la popularité d’une d’application mobile.

On s’intéresse à la :

* classification supervisée des avis laissés pa les utilisateurs
* classification non-supervisée de facteur qui permette de prédire la popularité d’une d’application mobile.

La base de donnée Google PlayStore Apps disponibles contient plusieurs informations, qu’on a extrait :

. Le nombre de téléchargement

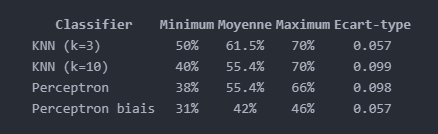
. La note de l’application

. Le nombre d’avis

. La taille d’une application

. Le type d’application (gratuit/payant)

Bases de données



On commence les problèmes de classification par un cas simple : étant données les avis des utilisateurs, peut-on savoir s’ils ont tendances à laisser plus d’avis négatifs que positifs ? En ne gardant que les applications avec au moins 1 avis et des avis inférieurs à 50000,

afin de limiter notre modèle, les accuracies obtenues par différentes méthodes ont été :

On a décidé de déterminer les facteurs qui permette de prédire la popularité d’une application mobile. On considère qu’une application est populaire quand : la note est supérieur ou égal à 4 et que le nombre de téléchargement est supérieur ou égal à 1 millions. On prédit le nombre de téléchargements en fonction des notes et des avis. On obtient le graphe suivant (chaque couleur représente un cluster et chaque croix, un centroide de clusters):

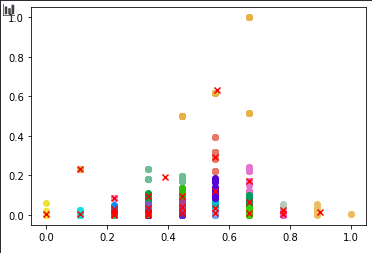
Le meilleur résultat possible est eut avec le classifier KNN. Néanmoins, les fluctuations constatées sur le classifier KNN sont trop élévée, on ne peut pas utiliser ces données comme référence. Donc, avec les résultats qu’on a obtenus, on ne peut émettre de conclusion.

On a décidé de déterminer les facteurs qui permette de prédire la popularité d’une applications mobile. On considère qu’une application est populaire quand : la note est supérieur ou égal à 4 et que le nombre de téléchargement est supérieur ou égal à 1 millions. On prédit le nombre de téléchargements en fonction des notes et des avis. On obtient le graphe suivant (chaque couleur représente un cluster et chaque croix, un centroide de clusters):

Classification des

avis

Classification non-supervisée



On effectue une validation croisée, pour obtenir les résultats suivants :

Le meilleure résultat avec le classifier KNN.

fluctuations constatée

KNN sont trop élevée,

utiliser ces données

Avec les résultats qu’

ne peut pas émettre possible est eut

Néanmoins, les

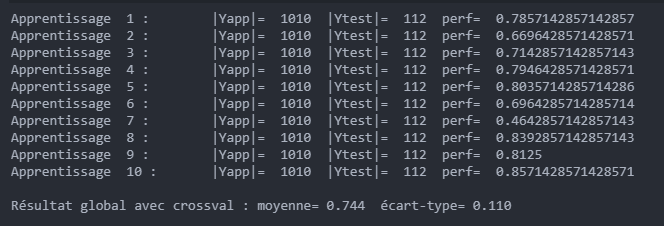
sur le classifier

on ne peut pas

comme référence.

on a obtenu, on

de conclusion.



Il est donc possible de le nombre de téléchargement d’une application populaire 3 fois sur 4.