Projet LU3IN026 – Data Science & IA
Analyses de données des applications sur le PlayStore

Achraf JDAY L3 ; Rida TALEB L3
Sorbonne Université

Introduction

Ce projet s'intéresse à l'analyse de données des applications mobile existantes sur le Google PlayStore. On s'intéresse:

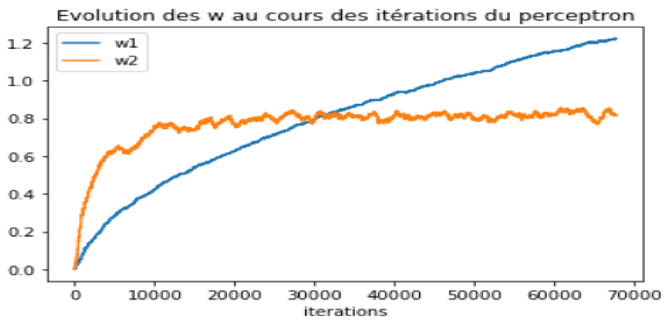
- À la classification supervise de la réussite de l'application (nombre d'installations) à partir de ses caractéristiques et notes.
- A la catégorisation non-supervise des reviews textuelles des utilisateurs individuels sur l'ensemble des applications payantes et gratuities.

Problématique 1

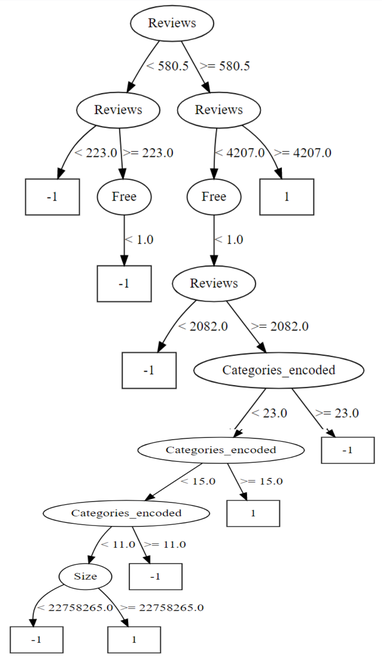
Prédiction si l'application réussira ou non. Au premier lieu l'application réussi si elle a un nombre d'installations très élevé (>100000), après nous modifions cette contrainte pour que ca soit un problème multiclasses dans les cas qui suivent.
Cette problèmatique est interessante surtout pour ceux qui souhaitent développer une app pour le PlayStore mais veulent vérifier si elle sera un succés avant en se basant sur ses caractéristiques ce qui permettra au développeur d'avoir un pre-insight sur le développement de son app et une longueur d'avance sur ses concurrents.

Modèles

Nous utilisons des modèles d'apprentissage supervisé que nous avons codé par nous meme, le Perceptron, le ADALINE Analytique, le MultiOOA et finalement les arbres de décisions numérique, tout en changont de paramètres dans plusieurs cas en split ainsi que en cross-validation strat après normalisation.



Classifieur	Apprentissage Accuracy	Tests Accuracy
Perceptron C1	66.27	67.29
Adaline C1	56.14	56.36
Perceptron C2	54.24	55.25
Adaline C2	45.49	45.89
Perceptron C3	41.07	41.23
Adaline C3	43.08	43.26

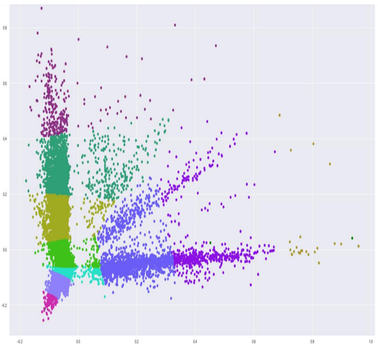


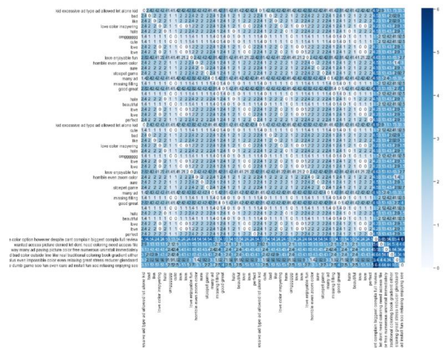
Resultats et Analyse

D'après les statistiques sur les modèles et les résultats nous pouvons clairement constater que le problème binaire marche le meilleur mais pour des raisons de réalisme nous allons regarder que le problème multiclasses et alors seulement le 2eme cas vu que d'après nos experimentations dans le 3eme cas le moins de dimensions/features on a le moins l'accuracy qu'on aura ce qui est logique.
Dans ce cas, le Perceptron est mieux que le Adaline dans les tests sur les données d'apprentissage ainsi que les données de tests, l'abre de décisions nous montre aussi des résultats logique (Exemple: réussite de l'app s'il y a un très grand nombre de reviews dans une catégorie populaire).
Résolution de problème réussi, nous avons réussi a prédire a 60% pret le succées d'une application mobile a partir de ses caractéristiques.

Problématique 2

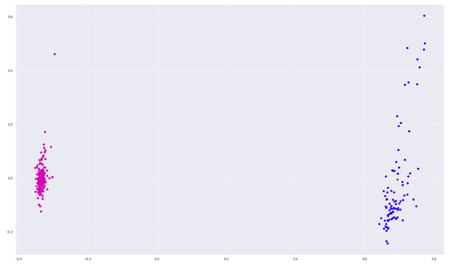
Analyse d'une base de données de reviews numérique sur les applications mobile du Google PlayStore ainsi qu'une analyse de sentiments et utilisation d'une base de reviews textuelle afin de prédire les clusters sur l'ensemble des applications gratuites et payantes.
Nous avons choisit cette problèmatique en apprentissage non supervisé parce que nous trouvons qu'il sera intéressant de voir la répartition des avis textuelles individuelles sur l'ensemble des données ainsi que sur les free apps et les paid apps séparément pour trouver des tendances même psychologique dans la manière dont les utilisateurs évaluent les produits qu'ils utilisent.

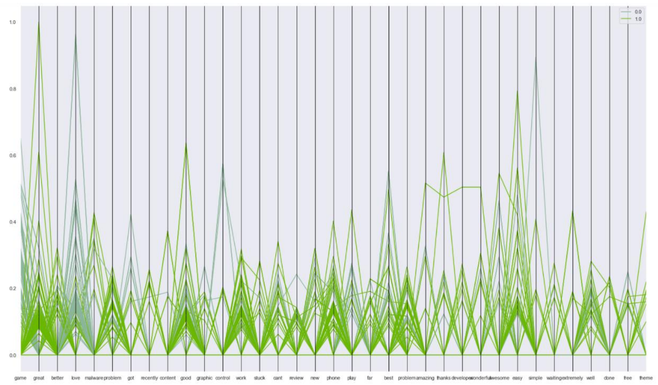


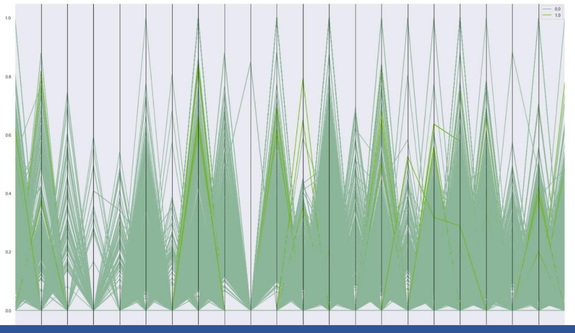


Modèles

Nous utilisons des modèles d'apprentissage non supervisé que nous avons codé par nous même, l'algorithme des K-Moyennes, nous avons aussi implémenté des clusterings heatmaps hiérarchiques comme modèles et des algorithmes statistiques notamment le TF-IDF pour adapter la base de données, tout en changont de paramètres dans plusieurs cas.







Resultats et Analyse

En regardant les résultats, nous remarquons instantanément que les applications gratuites sont critiqué beaucoup plus sévèrement que les applications payantes.
Une autre manière de résoudre ce problème est de faire d'abord de trouver le k de kmeans en suivant Dunn ou Elbow Method, ensuite on lance le kmeans pour obtenir les centroides et les utilisé pour créer des clustering hiérarchique dendrogramme.
Résolution de problématique réussit. Effectivement, la manière dont les utilisateurs critiquent les apps varie largement des apps gratuites au apps payantes.