Sorbonne Université

Rapport de projet

BIUM Projet

Zhirui QI
Zhihao ZHU
Master 1 Semestre 2

Février 2021 — Mai 2021

Index

Table des matières

1 Introduction

Le but de notre projet :

trouver la relation entre l'industrie du streaming en direct et l'industrie vidéoludique

Concrètement, il faut trouver quel type d'entreprise vidéoludique promeut le développement de l'industrie du streaming en direct. Au début, on voudrait dire qu'on a divisé les entreprises vidéoludique qui développe des jeux vidéos de 4 catégories de 4 catégories :

- **fix** ça veut dire les jeux vidéos sur console ou des équippements fixés et les entreprises développe juste cette catégorie de jeux vidéos
- **fixplus** la plupart des jeux vidéos développé par cet entreprise sont de catégorie fix
- **mobile** ça veut dire les jeux vidéos sur portable ou des équippements mobile et les entreprises développe juste cette catégorie de jeux vidéos
- **mobileplus** la plupart des jeux vidéos développé par cet entreprise sont de catégorie mobile

On analyse à partir de la volume de transaction d'actions pour chaque entreprise, et bien la revenue pour que ces deux données a une influence sur la situation actuelle du streaming en direct. On a choisi la volume de transaction des actions comme la métrique pour mésurer les conditions d'affaires des entreprises car c'est une métrique plus balancé et plus concrète. D'ailleurs, on fait des petites recherche additionnelles pour la relation entre les l'industre relationnel à l'industrie vidéoludique comme **NVIDIA** (entreprise graphisme) et **Gamestop** (Plate-forme et vendeur de jeu) et entreprise vidéoludique.

2 Sources de données et ETL

- Quand on a décidé le topic de notre projet, on voudrait chercher les données dans les datasets en ligne recommandé par le sujet de projet. Mais on trouve que ça ne suiffit pas en fait, donc par conséquent pour les annuaires des jeux vidéos, on achète le droit d'accès à un grand dataset statista .
- D'ailleurs,pour obtenir les données sur la transaction d'actions pour les entreprises vidéoludi--ques, on trouve un site inversting qui contient tout.
- On a ainsi fait *une capture Web* sur site **Steam** qui est le plus grand vendeur des jeux vidéos en ligne et on obtient un grand tableau sur prèsque tous les jeux vidéos avec son ventes etc.
- Enfin, on cherche les rapports des analyses des donnée en quelque site chinois pour compléter et enrichir notre dataset. analysys et iresearch

C'est tout pour notre sources de données.

Et puis, on fait une nettoie sur les données juste avant les prétraitements de façon **Prepare, Jointure et Group** pour notre données pour confirmer la forme de notre donnée et assurer l'utilisation des démarches suivantes.

3 Modélisation

3.1 Schéma conceptuel en étoile

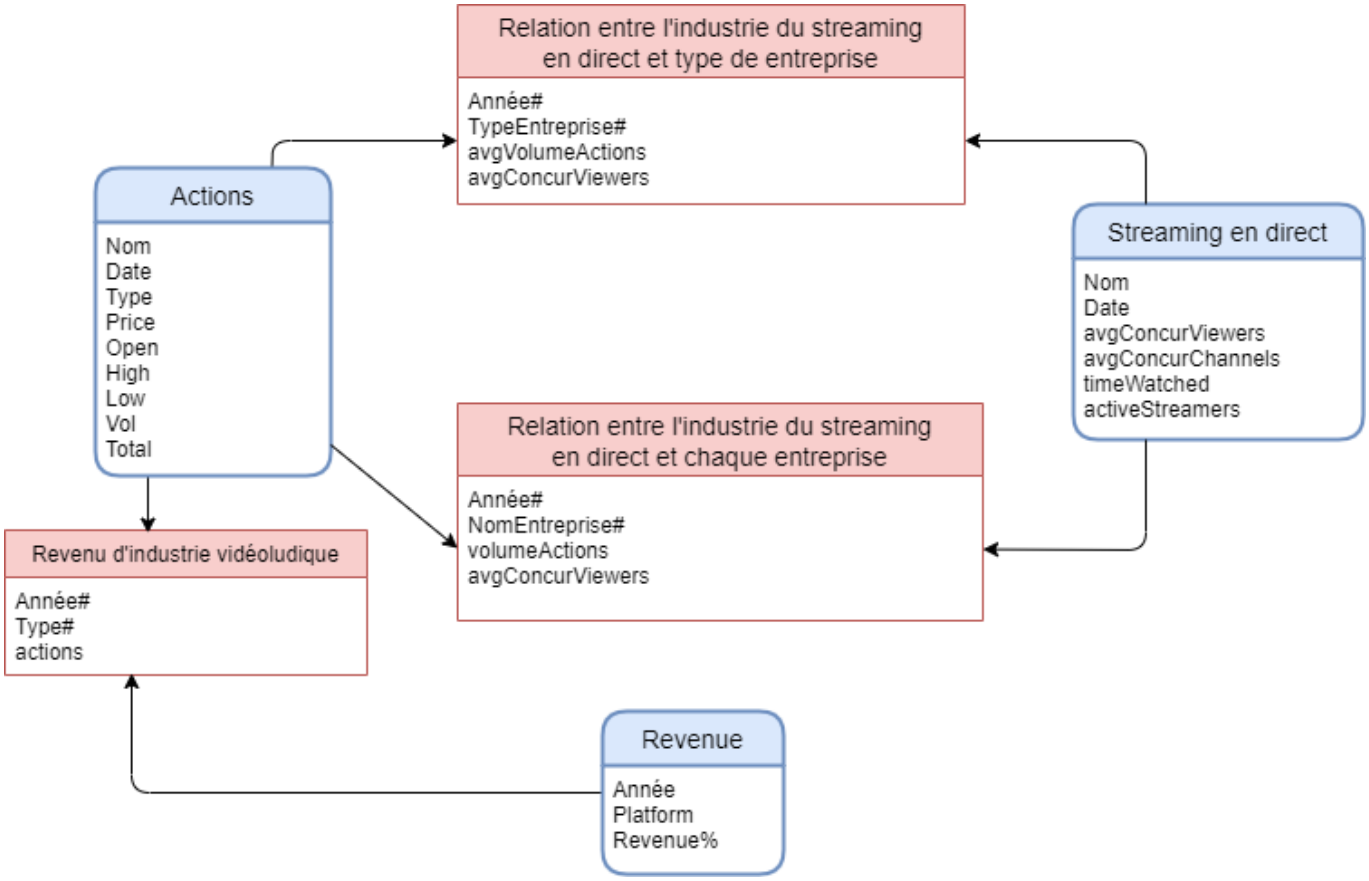

FIGURE 1 – Schéma en étoile

3.2 Schéma logique dénormalisé

Fait :

- Relation entre l'industrie du streaming en direct et type de entreprise
  (Année#, TypeEntreprise#, avgVolumeActions, avgConcurViewers)
- Relation entre l'industrie du streaming en direct et chaque entreprise
  (Année#, NomEntreprise#, VolumeActions, avgConcurViewers)
- Revenu d'industrie vidéoludique
  (Année#, Type#, Actions)

Dimension :

- Actions
  (nom, Date, Type, Price, Open, High, Low, Vol, Total)

– Revenu
(<u>Année, Platform</u>, Revenue%)
– Streaming en direct
(<u>Nom, Date</u>, avgConcurViewers, avgConcurChannels, timeWatched, activeStreamers)

3.3 Implémentation

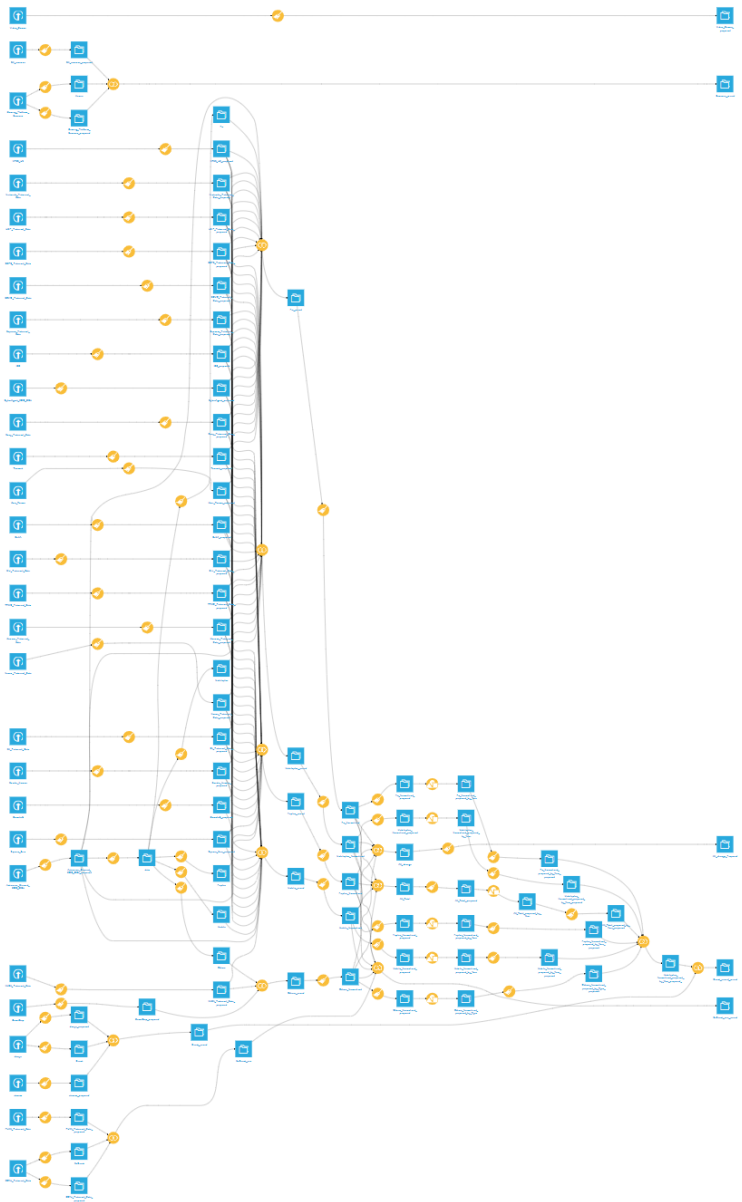

FIGURE 2 – Dataiku

Explication

Les données d'actions sont de forme séparé, donc il y a beaucoups de fichiers dans notre DataBase de *Dataiku*. On fait **filtrage, jointure et group** pour obtenir les données on souhaite.

Après, on traite les `Actions` par 2 méthodes :

1. pour `Relation entre l'industrie du streaming en direct et type de entreprise`, on regroupe par `Type` et fait jointure avec `Streaming en direct`
2. pour `Relation entre l'industrie du streaming en direct et chaque entreprise`, on fait jointure avec `Streaming en direct`

4 Analyses plus poussées

Dans cette partie, on fait totalement 4 analyses plus poussées pour approfondir notre recherche.

4.1 Clustering : kmeans

on fait une recherche supplémentaire : on essaye de cluster les actions en 5 clusters par la méthode `k-means`.

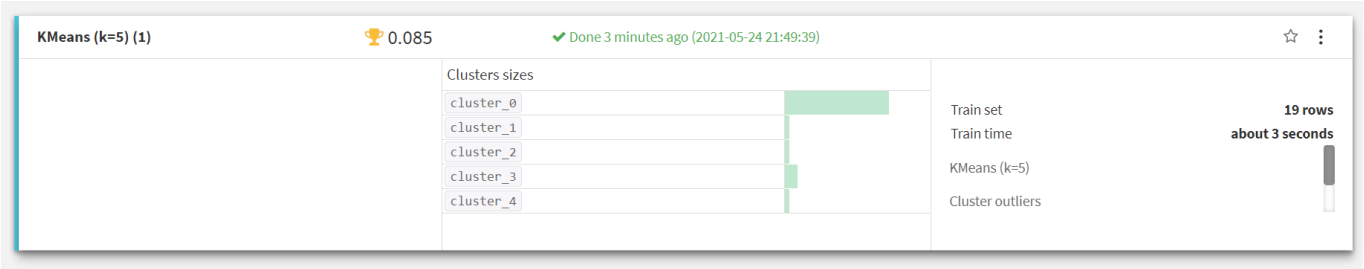

FIGURE 3 – k-means

on peut observer que les données se ressemblent dans cluster 0. On peut dire que les actions d'industrie vidéoludique sont similaire.

4.2 Random Forest et XGBoost

On fait une recherche supplémentaire : on analyse quelle est l'élément qui joue un rôle plus important pour augmentation et diminution d'actions par la méthode `Random Forest` et `XGBoost` de prediction type `Regression`.

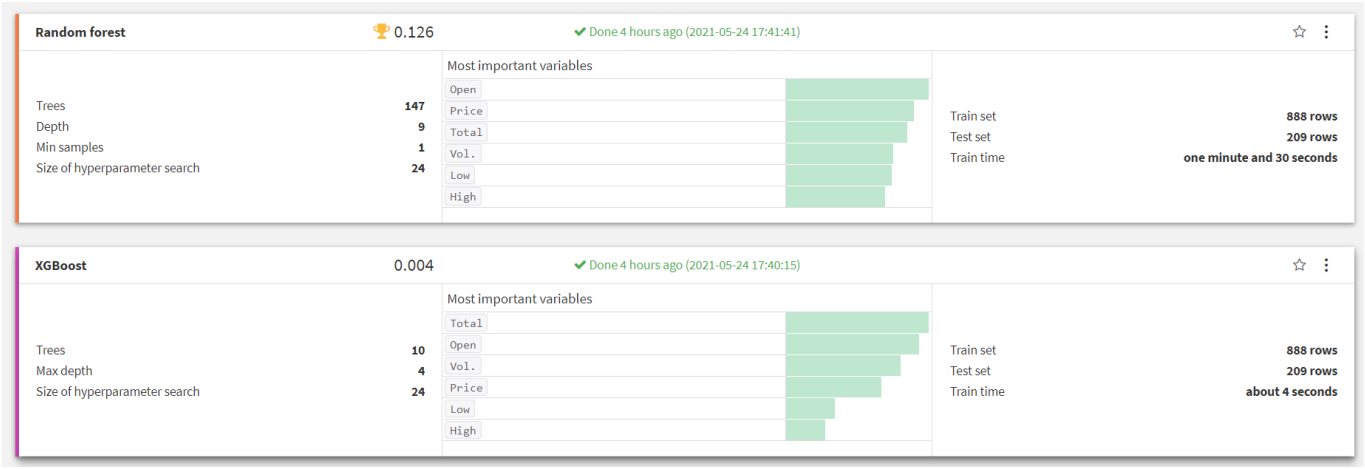

FIGURE 4 – CyberAgent

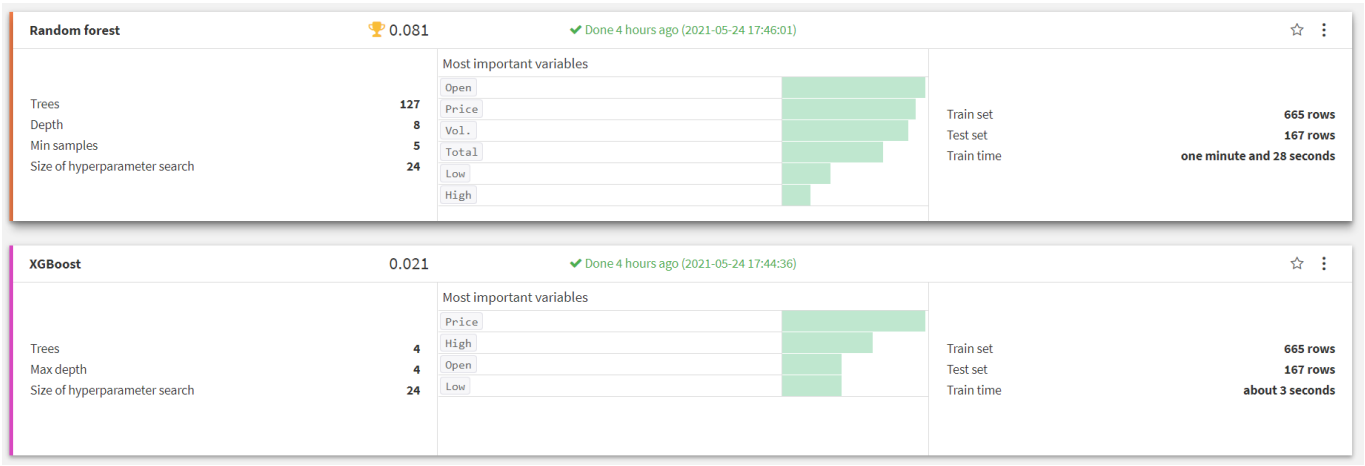

FIGURE 5 – Activision Blizzard

On peut dire que le *Open* est le plus important.

4.3 Analyse principal 1

On étudie quelle catégorie d'entreprise vidéoludique influence le plus pour l'industrie du streaming en direct par la méthode `Random Forest` de prediction type `Regression`. Du coup, parmi les aspects des données du industrie du streaming en direct, on trouve que le nombre d'audience fidèle est la critère cruciale. Donc, on cherche la relation entre celle-ci avec la catégorie d'entreprise vidéoludique.

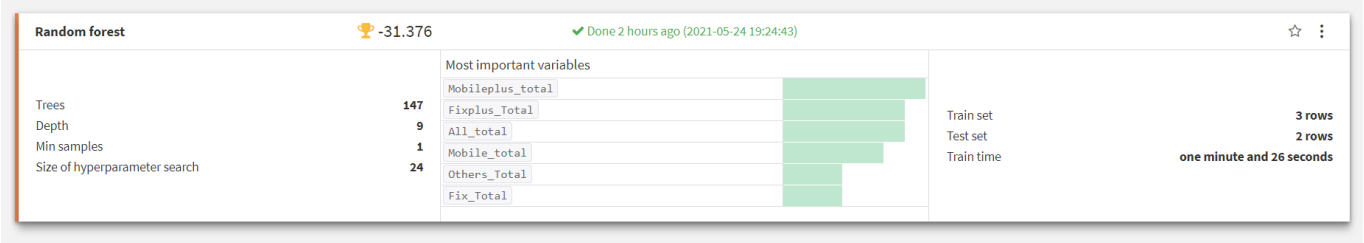

FIGURE 6 – DouYu

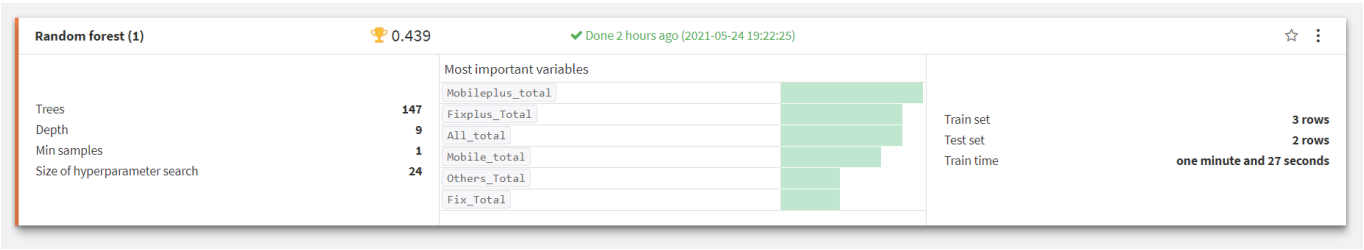

FIGURE 7 – Twitch

On peut dire que la catégorie *Mobileplus* influence le plus.

4.4 Analyse principal 2

On étudie quelle entreprise vidéoludique influence le plus pour l'actions d'industrie du streaming en direct par la méthode `Random Forest` de prediction type `Regression`.

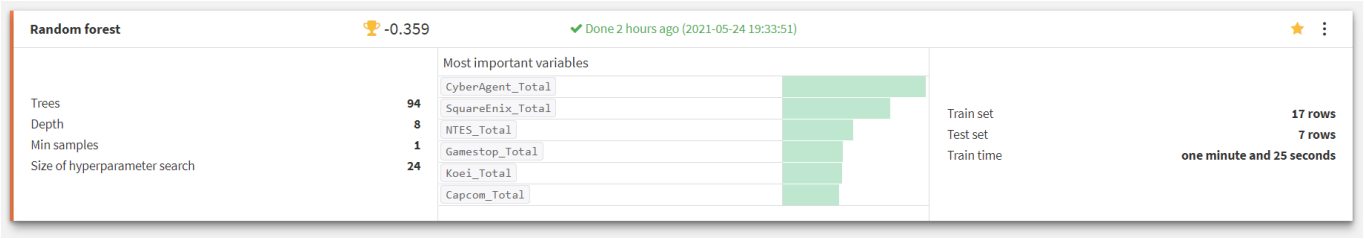

FIGURE 8 – DouYu

On peut dire que l'entreprise *CyberAgent* influence le plus.