

Healthy-Tomatoes



Master 2

ELBEZ Samuel 21200353 JACQUETTE Pierrick 21305551



Objectif



Peut-on déterminer le succès d'un film avant sa sortie au cinéma?



Technologie



Elasticsearch

kaggle



Python 3.5



réalisateur

genre

keywords

6 premiers acteurs

https://www.kaggle.com/tmdb/tmdb-movie-metadata/data

compagnies de production

vote_average

moyenne, médiane et l'écart-type pour calculer la note limite 6.2 entre succès et échec

Méthode

Naïve Bayes

TF-IDF

KNeighbors

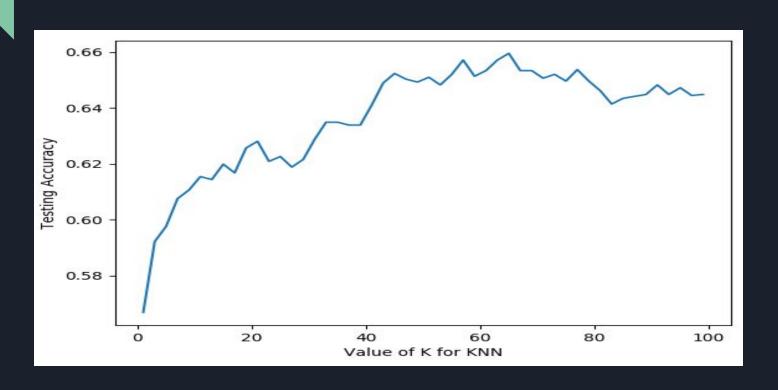
Validation croisée

Decision Tree

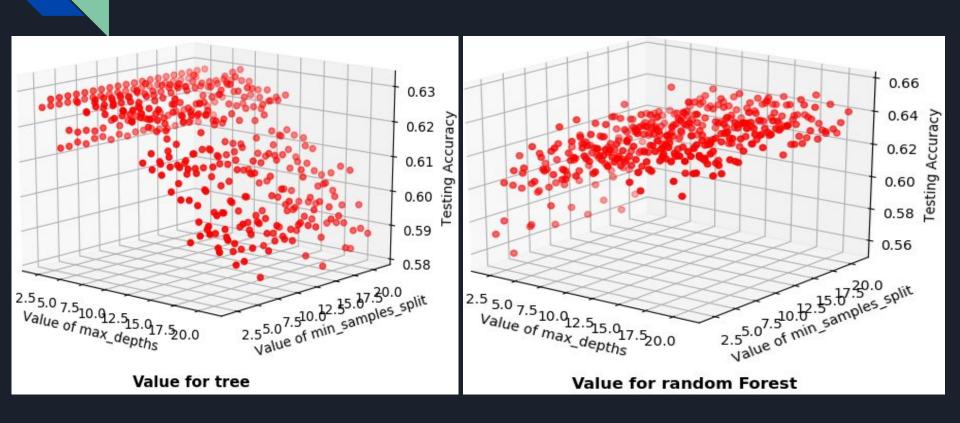
Random Forest



Validation croisée (1/2)



Validation croisée (2/2)



Résultat

- my naive Bayes: 63%
- naive Bayes: 65% (precision 65 % et recall 64%)
- k plus proches voisins: 61% (precision 62% et recall 61%)
- arbre de décision : 63% (precision 63% et recall 63%)
- random forest: 66% (precision 66% et recall 66%)

Explication



- Les données ne sont pas en nombre suffisant.
- On ne prend pas en compte la courbe des tendances, par exemple dans les années 90, le genre action est saturé et cela peut fausser les résultats des films à venir.
- Des champs que nous avons ignorés auraient pu être utilisés.

Difficulté



Le sujet : Où commencer? Où vas-t-on?

• ElasticSearch : liste de taille 10 & objet json



Extension



- Considérer tous les champs fournit en appliquant les bons poids.
- Appliquer un poids à certains champs pour influencer les résultats mais il est difficile de déterminer qui est quoi : certains acteurs sont parfois des réalisateurs.
- Une interface graphique ou textuelle permettant à l'utilisateur de rentrer un nouveau film avec tous les champs nécessaire.

Conclusion

Pour conclure, nous voulions rappeler que ce projet a été enrichissant pour nous.

On comprend mieux comment certains outils de la vie quotidienne fonctionnent. On est capable de prévoir pour pouvoir décider en fonction de l'expérience.

Merci a vous

Avez-vous des questions?