

Compte Rendu DS Machine learning

Sujet (Même sujet qu'en Data Mining): Le datamining pour la prévision de succès des films

NB :

- Le dataset a été traité et nettoyé dans le projet de DataMining ;
- Ici, on va s'intéresser au succès commercial (Revenue) et au succès critique (Vote)
- On fait le test sur 20% du dataset dans tout le projet

Tableau récapitulatif des modèles

Succès/modèles	Arbre de décision (%)	Forêt aléatoire (%)	Forêt aléatoire + Cross validation avec K = 44 et 200 arbres	Forêt aléatoire + Cross validation avec K = 28 et 200 arbres	Classifieur Bayésien (%)	K plus proches Voisins (%)
Succès critique	68	73	//	71.98	50	60
Succès commercial	82	87	87.14	86.73	81	83

On remarque ici que le meilleur modèle pour le succès critique est **Forêt aléatoire** et pour le succès commercial c'est **Forêt aléatoire (200 arbres) avec cross validation K=44**.

Pour le déploiement

Application de prédiction de succès commercial et de succès critique des films

Voir **app.py** pour le backend (en flask) et **frontend.html** pour le frontend