

ONLINE NEWS POPULARITY

Vincent Bonnet & Théophile Bornon

Problématique

Le dataset :

Cet ensemble de données résume un ensemble hétérogène de caractéristiques sur les articles publiés par Mashable sur une période de deux ans.

Plus d'info sur les variables [ici](#)

Le contexte:

L'objectif est de prédire la popularité des News à partir d'un ensemble de données, c'est-à-dire montrer la corrélation entre la popularité d'un article en fonction de ses caractéristique.

Réflexions

Après une compréhension des différentes variables et sur leur contenu, nous avons cherché à avoir un esprit critique sur le choix des variables de l'auteur, afin de déterminer quelles sont celles qui avaient un impact sur la prédiction de la variable cible.

Variables créées

La variable cible pour déterminer la popularité des News est le nombre de partages. Pour visualiser les données puis prédire leur niveau de popularité on a choisi de créer un label de popularité correspondant allant de très faible à exceptionnelle (7 catégories au total)

Résultats

Le problème initial est de savoir s'il est possible de prédire efficacement la popularité à partir de certaines données. Au cours de cette démonstration nous avons cherché à montrer cette corrélation dans le contexte de l'étude.

A travers quelques visualisations puis des algorithmes de prédiction nous avons soutenue cette hypothèse:

- Il y a bien une relation entre les caractéristique d'une News et sa popularité
- Une prédiction efficace est alors possible avec 50% de précision avec notre label. (Des résultats plus précis peuvent obtenue avec des optimisations des modèles testés).