IAS

Prédiction de la satisfaction lors d'un vol

SARAH BOUNDAOUI & AHMED BAAROUN

Table des matières

Présentation du dataset

- Ensemble trouvé sur Kaggle de source malheureusement inconnue car ayant apparemment été remanié.
- Il se présente en format CSV, déjà scindé en deux fichiers "train" et "test".
- Chaque ligne correspond aux caractéristiques de voyage d'un passager ainsi que sa satisfaction (ou pas !).

Etapes réalisées (problème)

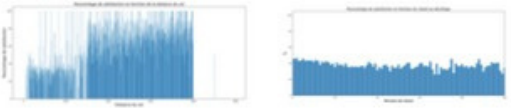
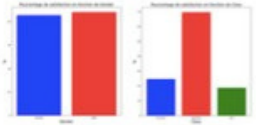
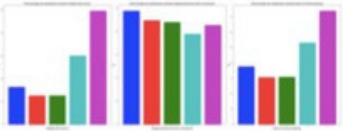
Chargement et exploration surfacique des données.

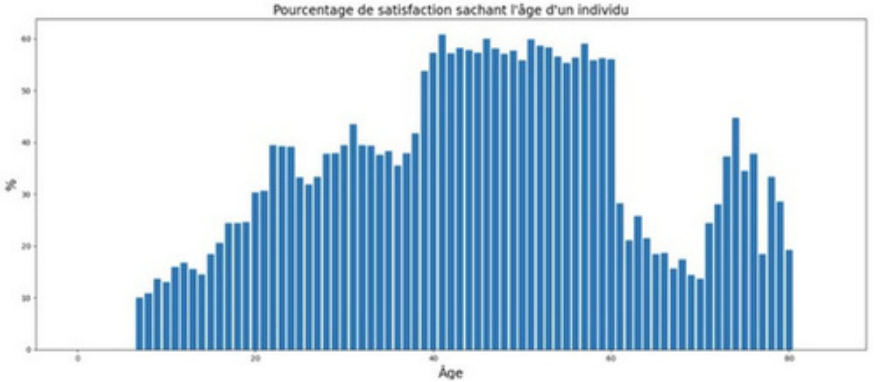
Recherche de liens et préparation à la classification.

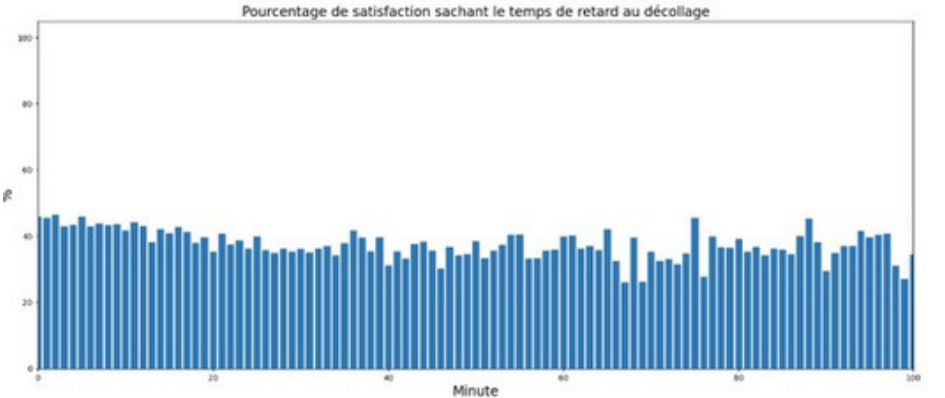
Application de :

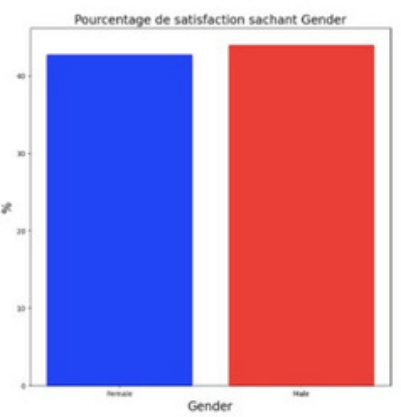
- Decision Tree Classifier
- Random Forest Classifier

Exploration des données

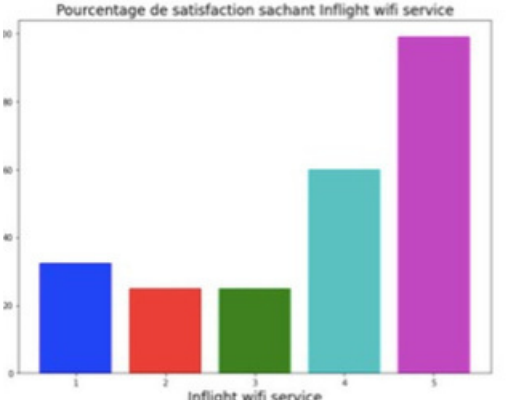
	Analyses Réalisées	Exemples de visualisations effectuées
Données "irrégulières"	Diagrammes à bâtons montrant les tranches de données en fonction du pourcentage de satisfaction	
Données catégoriques nominales	Diagrammes à bâtons de couleur pour chaque catégorie permettant d'avoir une comparaison entre les catégories en fonction du pourcentage de satisfaction	
Données catégoriques numériques	IDEM aux Données catégoriques nominales	

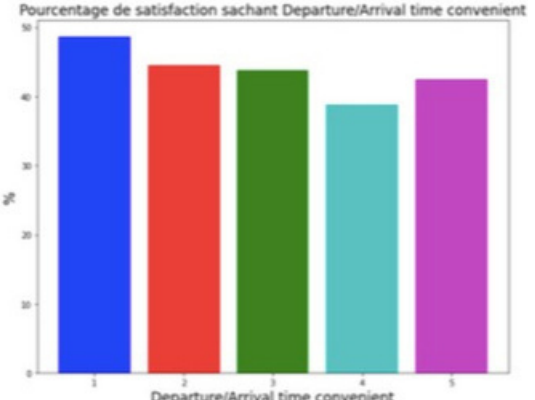

Pourcentage de satisfaction sachant l'âge d'un individu
%
Âge


Pourcentage de satisfaction sachant le temps de retard au décollage
%
Minute


Pourcentage de satisfaction sachant Gender
%
Female
Male
Gender

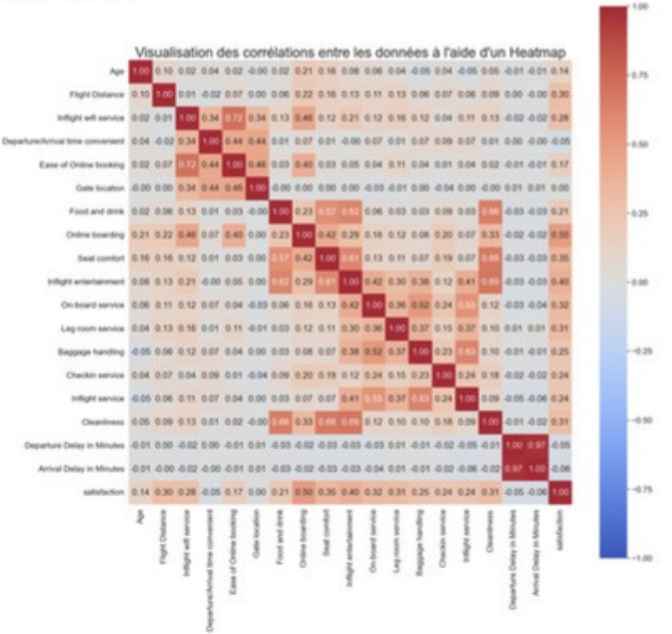

Pourcentage de satisfaction sachant Customer Type
%
Loyal Customer
disloyal Customer
Customer Type


Pourcentage de satisfaction sachant Inflight wifi service
Inflight wifi service


Pourcentage de satisfaction sachant Departure/Arrival time convenient
%
Departure/Arrival time convenient

Analyse par le biais de la corrélation

- Certaines features ne sont en rien corrélées avec la satisfaction, comme le montre la Heatmap ci-dessous :



Préparation des données

Binarisation (Numérisation) des données catégoriques nominales après la suppression de données jugées inutiles

Application de Decision Tree Classifier

Données à classifier (satisfaction ou non) donc utilisation d'un classificateur assez intuitif qu'est Decision Tree (Classifier)

Random Forest

Essai d'un autre modèle, qu'est une multitude d'arbres de décision, ce qui peut être intéressant de voir pour notre dataset

Optimisation de chaque modèle

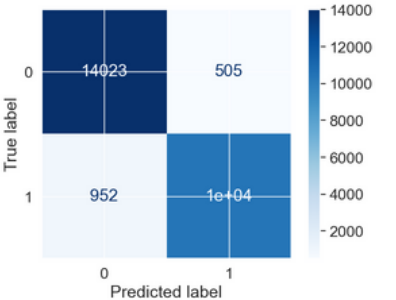
Première application des modèles de façon intuitive puis réglage des hyperparamètres avec des estimateurs plus rigoureux

Comparaison de précision

Comparaison des deux modèles avec des courbes ROC (Receiver Operating Characteristic) avec une AUC (Area Under the Curve) permettant d'évaluer la performance de chacun et de les comparer naturellement

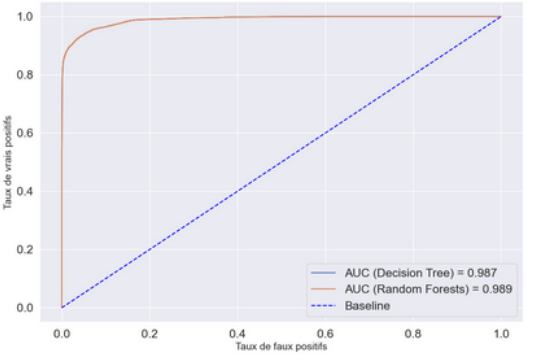
Courbe d'apprentissage de Decision Tree Classifier

Matrice de confusion de Decision Tree Classifier



Courbe d'apprentissage de Random Forest Classifier

Courbe ROC de comparaison entre les deux modèles



Pour conclure

Comparaison indiquant des performances assez similaires avec un léger avantage pour Random Forest

un écart et des performances sûrement plus élevées avec plus de données

L'intuition ne constitue pas forcément une vérité !