

RAPPORT PROJET IAS

Sarah Boundaoui.

&

Ahmed Baaroun.

INTRODUCTION :

Ce projet est un projet à but “ouvert” qui consiste à choisir un Dataset libre sur lequel nous allons devoir choisir un objectif, raffiné le data approfondir la maitrise et découvrir des nouveaux algorithmes et accomplir l’objectif via un modèle vu en cours.

PRESENTATION DU DATASET :

Nous avons choisi un Dataset qui contient les informations d'un survey donnant le niveau de satisfaction de passager après un vol d’avion.

Lien vers le Dataset : <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

Le Dataset comporte environ 130 000 lignes et 25 labels, il comporte 2 fichiers .csv, l'un pour l'entraînement des données et l'autre pour le test de donnée sur un ordre de 80/20 au sujet des passagers d’un vol.

Voici ci-dessous certaine des colonnes présentent dans notre dataset :

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	...	5	4	3	4	4
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	...	1	1	5	3	1
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	...	5	4	3	4	4
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	...	2	2	5	3	1
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	...	3	3	4	4	3
...
103899	103899	94171	Female	disloyal Customer	23	Business travel	Eco	192	2	1	...	2	3	1	4	2
103900	103900	73097	Male	Loyal Customer	49	Business travel	Business	2347	4	4	...	5	5	5	5	5
103901	103901	68825	Male	disloyal Customer	30	Business travel	Business	1995	1	1	...	4	3	2	4	5
103902	103902	54173	Female	disloyal Customer	22	Business travel	Eco	1000	1	1	...	1	4	5	1	5
103903	103903	62567	Male	Loyal Customer	27	Business travel	Business	1723	1	3	...	1	1	1	4	4

TABLEAU DES LABELS :

<u>Labels</u>	<u>Attributs</u>
Sexe	Sexe des passagers (femme, homme)
Type de client	Le type de client (client fidèle, client déloyal))
Âge	L'âge réel des passagers
Type de voyage	But du vol des passagers (voyage personnel, voyage d'affaires)
Classe	Classe de voyage dans l'avion des passagers (Business, Eco, Eco Plus)
Distance de vol	La distance de vol de ce voyage
Service wifi à bord	Niveau de satisfaction du service wifi à bord (0:Sans objet, 1-5)
Heure de départ / arrivée pratique	Niveau de satisfaction de l'heure de départ / arrivée pratique
Facilité de réservation en ligne	Niveau de satisfaction de la réservation en ligne
Emplacement de la porte	Niveau de satisfaction de l'emplacement de la porte d'entrée
Nourriture et boissons	Niveau de satisfaction de la nourriture et des boissons
Embarquement en ligne	Niveau de satisfaction de l'embarquement en ligne
Confort du siège	Niveau de satisfaction du confort du siège
Divertissement à bord	niveau de satisfaction du divertissement à bord
Service à bord	Niveau de satisfaction du service à bord
Service d'espace pour les jambes	Niveau de satisfaction du service d'étagage pour les jambes
Check in des bagages	Niveau de satisfaction du Check in des bagages
Service d'enregistrement	Niveau de satisfaction du service d'enregistrement
Service en vol	Niveau de satisfaction du service en vol
Propreté	Niveau de satisfaction de la propreté
Retard de départ en minutes	minutes retardées au moment du départ
Retard d'arrivée en minutes	Minutes retardées à l'arrivée
Satisfaction	Niveau de satisfaction de la compagnie (satisfaction, neutre ou insatisfaction)

OBJECTIF :

Notre objectif va être de prédire la satisfaction d'un passager grâce à nos données en utilisant une classification (mais pas que) en commençant par explorer les données et les nettoyer, un Processing et une visualisation. Notre métrique de performance va être la différence entre la satisfaction prédit et la satisfaction réelle.