Note de synthèse

Bank Marketing

Quentin PICARD

Sommaire

[**Collecte des données**](#_4uccx8pu124x) **2**

[**Préparation des données**](#_x5b6c2m61dfs) **2**

[**Choix du modèle et des hyperparamètres**](#_l3lo4y7mwmr) **4**

[**Choix des features**](#_zbf0mngrlu2b) **8**

[**API**](#_2ibid6mer4qa) **9**

[**CI/CD**](#_cz2fks4bg0xj) **10**

# Collecte des données

J’ai effectué la recherche de données sur différents sites proposant des datasets.

J’ai choisi un dataset provenant de kaggle, à propos de campagne marketing bancaire.

# Préparation des données

Les données sont constituées de 16 features et d’un label booléen.  
Le dataset ne comporte pas de données manquantes.

Le dataset contient uniquement de données numériques et catégorielles, pas de texte.

Actions réalisées:

-Encodage numérique des catégories

-Test du OneHotEncoding

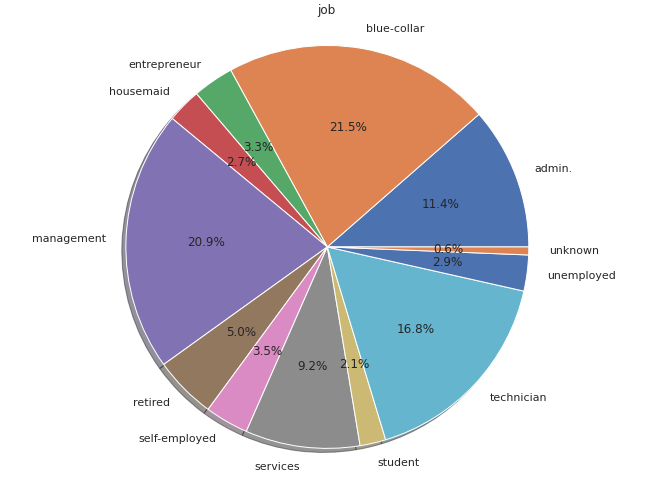
OneHotEncoding:

Permet de faire tourner certains algorithmes (SVM, SVC, LinearSVC, …) en encodant les catégories non ordinales.

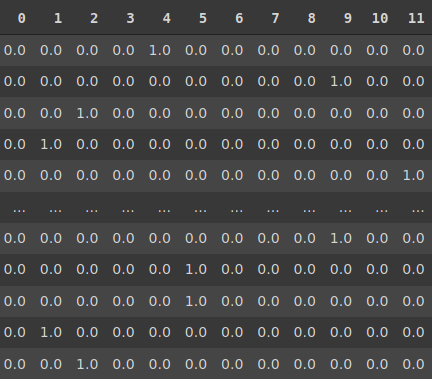
Le principe est de représenter une feature catégorielle en plusieurs colonnes booléenne.

Exemple de OneHotEncoding sur ma feature ‘job’:

Avant: (sur une colonne)



Après:



# Choix du modèle et des hyperparamètres

Après quelques tests sur différents algos, j’en ai retenu 3:

KNearestNeighbors

RandomForestClassifier

DecisionTreeClassifier

J’ai donc effectué une recherche d’hyperparamètres sur ses algos, respectivement:

Nombre de voisins (n\_neigbors)

Nombre d’estimateurs (n\_estimators)

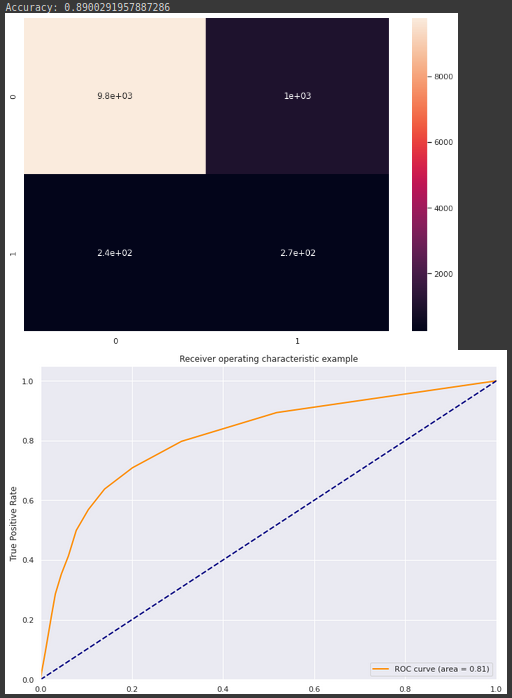
Profondeur maximale (max\_depth)

Résultat de la recherche d’hyperparamètres:

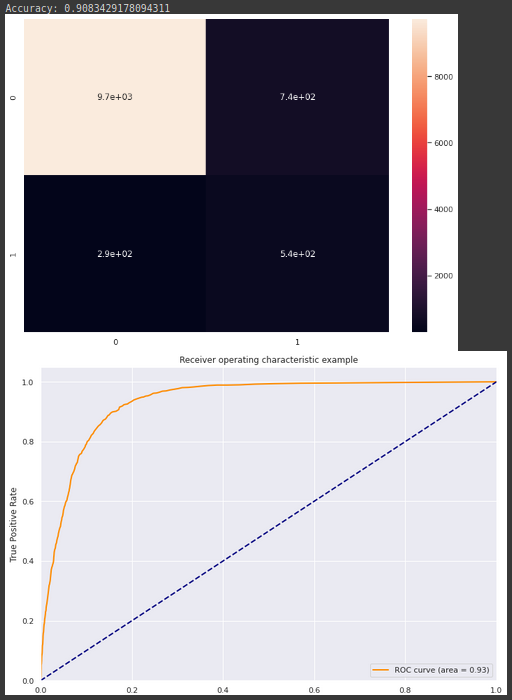
# 

Résultat des entrainements:

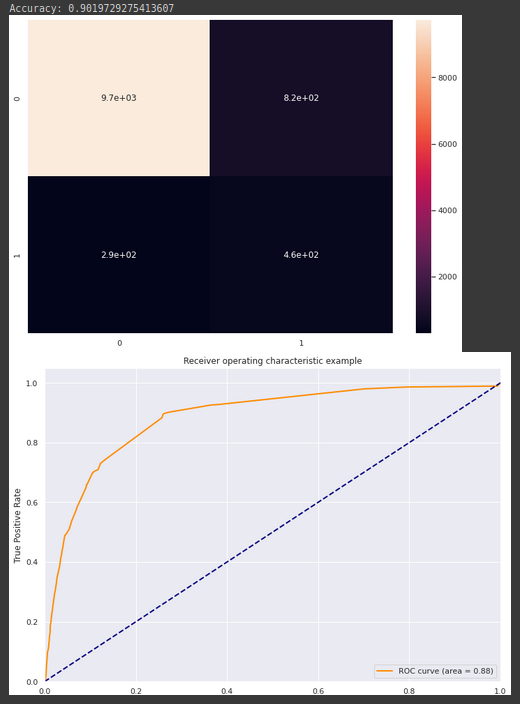
KNN: Accuracy: 0.89 AUC: 0.81



RF: Accuracy: 0.91 AUC: 0.93



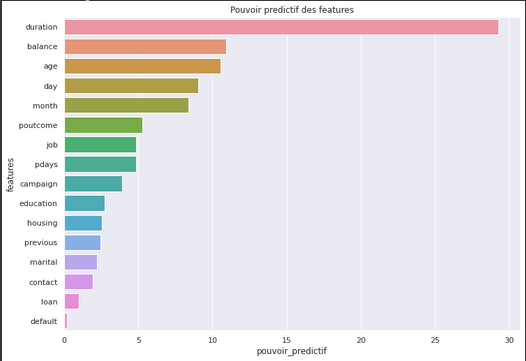
DT: Accuracy: 0.90 AUC: 0.88

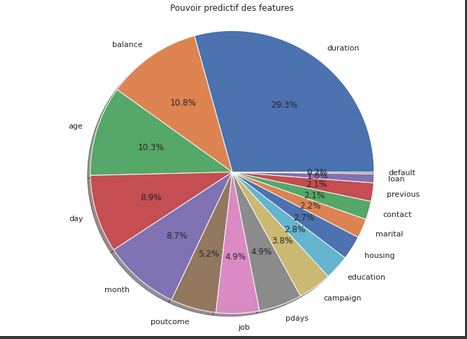


# 

# Choix des features

Pouvoir prédictif obtenue avec un RandomForest(le modèle retenu):





Top 3 des features:

Duration

Balance

Age

# API

Api réalisé avec fastapi.

# CI/CD

CI/CD mis en place via Google Cloud.