Projet Crypto Bot

Data Engineering Promo 2

Le projet a pour but de développer un système de recommandation pour l'achat et la vente de crypto-monnaies, en utilisant les données historiques et en temps réel de l'API Binance, une plateforme d'échange de crypto-monnaies.

Ce système sera soutenu par des modèles d'apprentissage automatique afin d'analyser les données et de prédire les mouvements futurs des prix. Le périmètre du projet s'étend à l'observation de trois devises spécifiques : Bitcoin (BTC), Cardano (ADA), et Ethereum (ETH),

Pour ce qui est du projet, la méthodologie que j’ai suivie et celle discutée avec mon groupe concernant le choix de la crypto monnaie le stockage des données.

Pour ce fait j’ai pu préparer une interface avec « streamlit » qui permet d’afficher non seulement l’historique mais aussi d’extraire les données de la crypto monnaie du jour courant en dépassant les limitations de l’API Binance. J’ai aussi suivi le modèle « CRISP-DM » pour la préparation et transformation des données mais aussi de lancer la partie machine learning sur l’ensemble des données d’entrainement et de test avec la Régression linéaire, pour enfin sortir une décision d’achat ou de vente.

Installation

Pour préparer le projet nous avons besoin d’installer les bibliothèques suivantes :

scipy: 1.2.1

numpy: 1.21.6

matplotlib: 3.2.2

statsmodels: 0.10.0

sklearn: 0.21.2

seaborn: 0.9.0

pip: 23.3.2

prophet: 1.1.5

pystan: 2.19.1.1

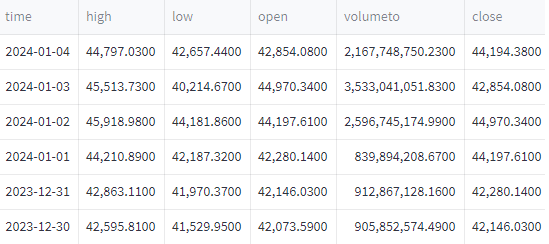
pandas: 1.3.0

streamlit: 1.11.0

1 . Récolte des données.

Récolte des données avec l’API Cryptocompare, j’ai lancé une fonction de récupération de données que j’ai pu manipuler afin d’extraire un dataset de 8 ans de n’importe qu’elle crypto monnaie.

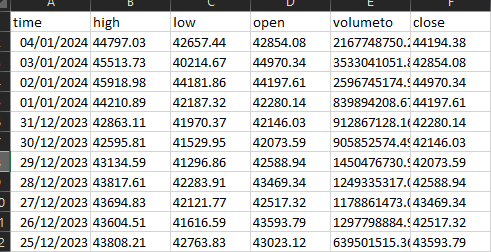
Développement d’un script de pré-processing pour réorganiser les données sortant de l’historique et streaming.



Ajout d’un lien pour télécharger les données en format CSV :



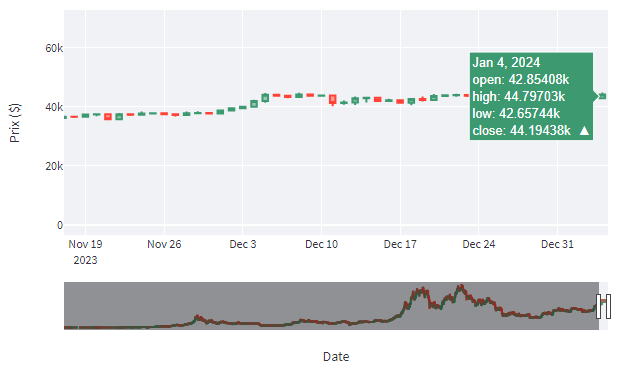
Aperçu des données téléchargées sous excel.



2. Visualisation des données.

Dashboard de l’évolution de Crypto monnaie choisit dans cet exemple ‘’BTC’’





3. Stockage des données au niveau d’Oracle.

Données historiques qui s’étale sur 7ans du « 1er janvier 2016 au 31 décembre 2022 » stocker au niveau de Oracle sous le nom « HIST\_DATA ».

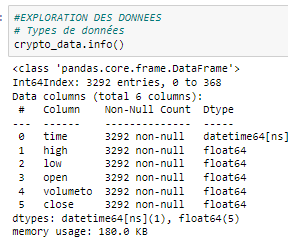
Colonnes : Time, Close, Open, High, Low, Volume

Données streaming qui début du 1er janvier 2023 à date stocker au niveau d’Oracle sous le nom « CRYPT\_DATA ». (fichier attaché : )

4. Exploration, préparation et transformation des données

Pour l’étape de consommation des données, avant d’appliquer l’algorithme machine learning, on est passé par l’étape de transformation et exploration des données pour bien structuré notre dataset depuis la source et vérifier la typologie des données et les valeurs nulles.

Exploration des données



Description des données

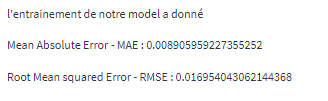


5. Machine learning

Après étape d’exploration et transformations, nous avons spliter les données en train 70% et test set 30% et normalisé nos données avec ‘’ MinMaxScaler’’ comme suit :



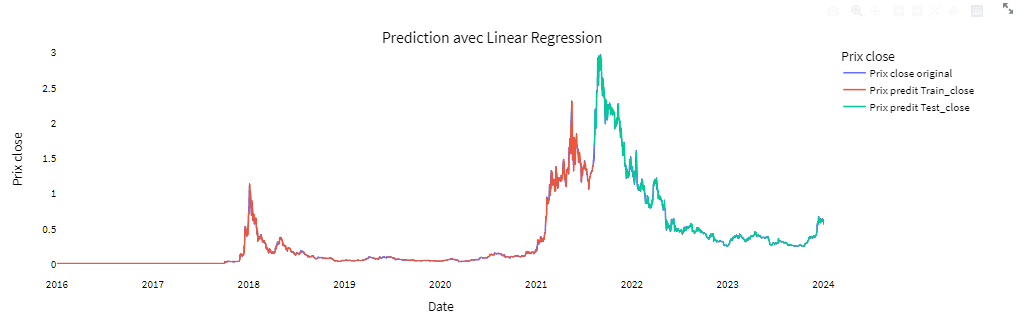
L’entrainement du modèle a donné :



L’évaluation du modèle donne un Square score de :



Prédiction avec l’algorithme Régression linéaire



Le résultat semble bien entrainé et donne de bon résultat.

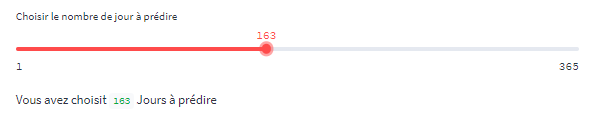
Nous avons par la suite lancer notre model sur les données de validation 

L’évaluation du set de validation donne :

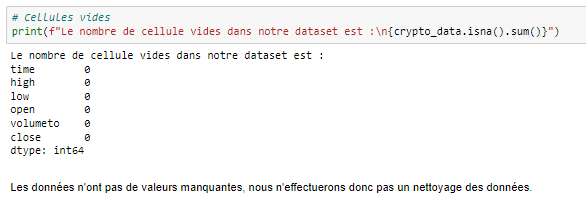


Résultat après apprentissage du model avec Prophet :

Nombre de jour à prédire (Prophet) :



Vérification des valeurs nulles



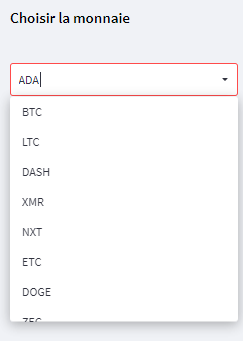
6. Mise en production

Les difficultés que j’ai rencontré lors de cette partie sont des limitations avec un aspect sécuritaire appliqué par Orange Maroc pour l’installation de quelques outils comme docker ou librairie python.

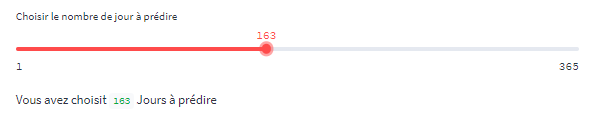
Mais j’ai tout de même pu faire une application web avec streamlit qui permet de faire la visualisation avec un dashboard un graphe d’évolution de la crypto monnaie choisie.

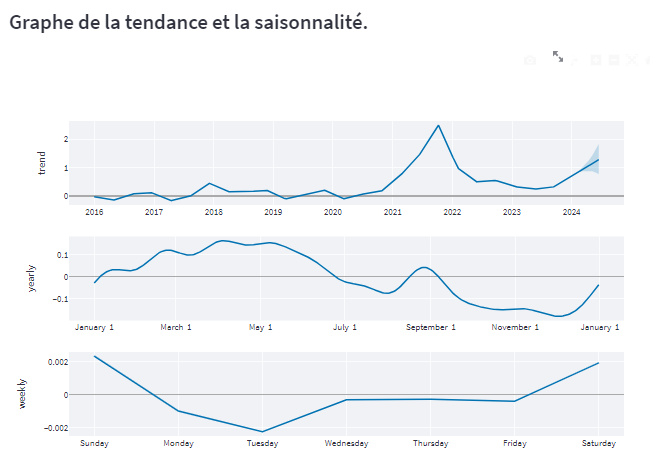
Pour lancer notre application, il faut lancer la commande suivante :



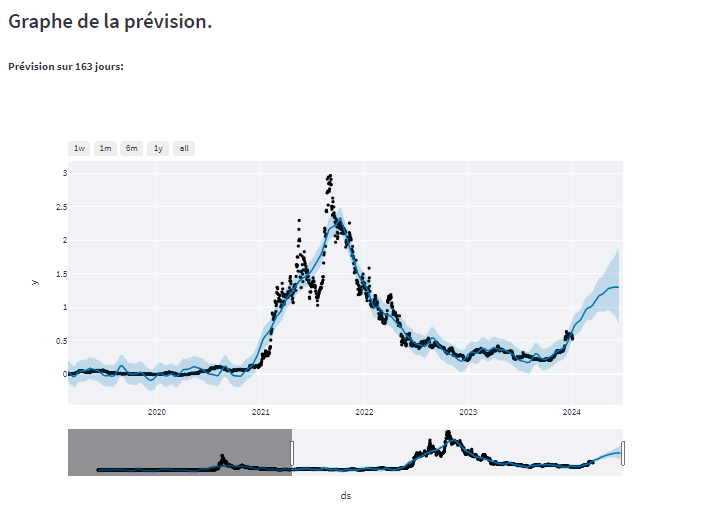


Résultat après apprentissage du model avec Prophet :

Nombre de jour à prédire (Prophet) :Graphe représentant la tendance et saisonnalité de la période choisit pour la prédiction, qui nous donne une idée sur la distribution des données.



Visualisation des données du dataset en point noir ainsi que les données prédites pour l’ensemble du train et test set sans oublier le nombre de jours prédits :

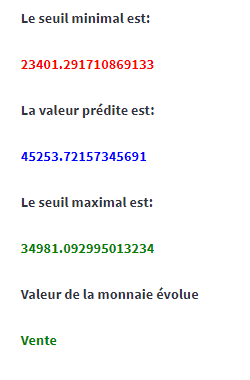
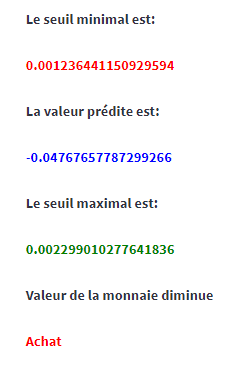


On passant par toutes les étapes déjà mentionnées depuis l’extraction à l’exploration la transformation et l’apprentissage du model ainsi que le calcul des différents scores pour l’étape d’évaluation du model et enfin retourné une décision d’achat.

Cette dernière suit la logique suivante :

Je lance l’entrainement et la prédiction d’une valeur future avec la librairie « Prophet », je prends le jour prédit par exemple « 08/01/24 » et je prends toute les valeurs dans l’historique avec une période de 7jours sur l’ensemble des données (tous les lundis ) ensuite je calcule la moyenne et l’écart type pour en créer un seuil min(mean-std) et max(mean+std), pour la décision d’achat je compare la ‘’ yhat : la valeur prédite‘’ avec min et max ;

Si la valeur prédite est supérieur au seuil min => ‘’Vente’’ si entre min et max :

* crypto stable ‘’Vente’’ sinon
* ‘’Achat’’.