Licence 3 D.A.N.T.

U.E. Data Sciences

Prédire le cours du Bitcoin

Gabriel LEFFAD

Robin SAN-VICENTE

Sommaire

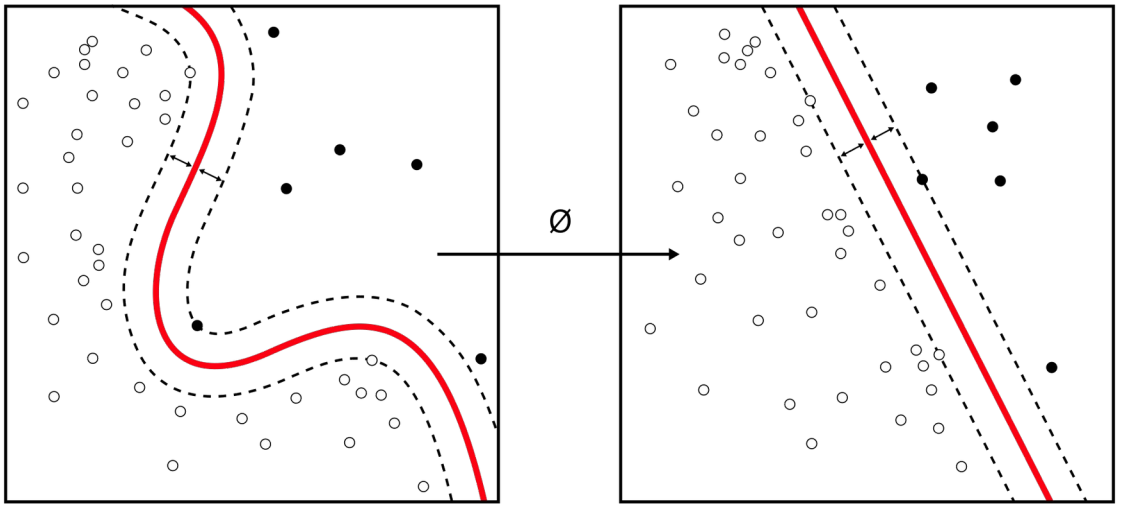
- Problématique et objectif
- Deep learning ?
- Réseau de neurones
- Technologies
- Création de l’environnement
- Les données
- Le model
- L’analyse des données et du model
- Conclusion
- Démonstration

Problématique et objectif

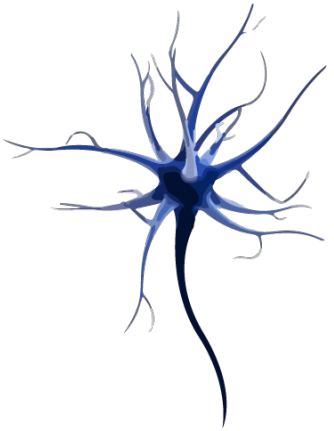
- Cryptomonnaie dont l'idée a été présenté en novembre 2008 par un groupe de personne sous le pseudonyme « Satochi Nakamoto ».
- Fourni des services de monnaies tout à fait légitime (d'après le Séna des USA).
- Il s'agit aujourd'hui bien plus d'une valeur refuge, plus que d'un moyen de paiement, on pourrais parler d'or numérique. Son avantage sur l'or est d'offre un moyen de paiement.
- Cette valeur est souvent jugé par les traders comme « radio-active », car extrêmement volatile et récente.
- On a pu observer un impacte directe de la crise du Corona virus de 2020, sur la valeur du bitcoin (-50 %, suivi de +100 % en moins de 3 mois).
- Pouvons-nous utiliser la science des données afin de prédire le cours du Bitcoin ?
- Notre objectif sera de tenter d'obtenir une prédiction sur le cours du Bitcoin.

Deep learning ?

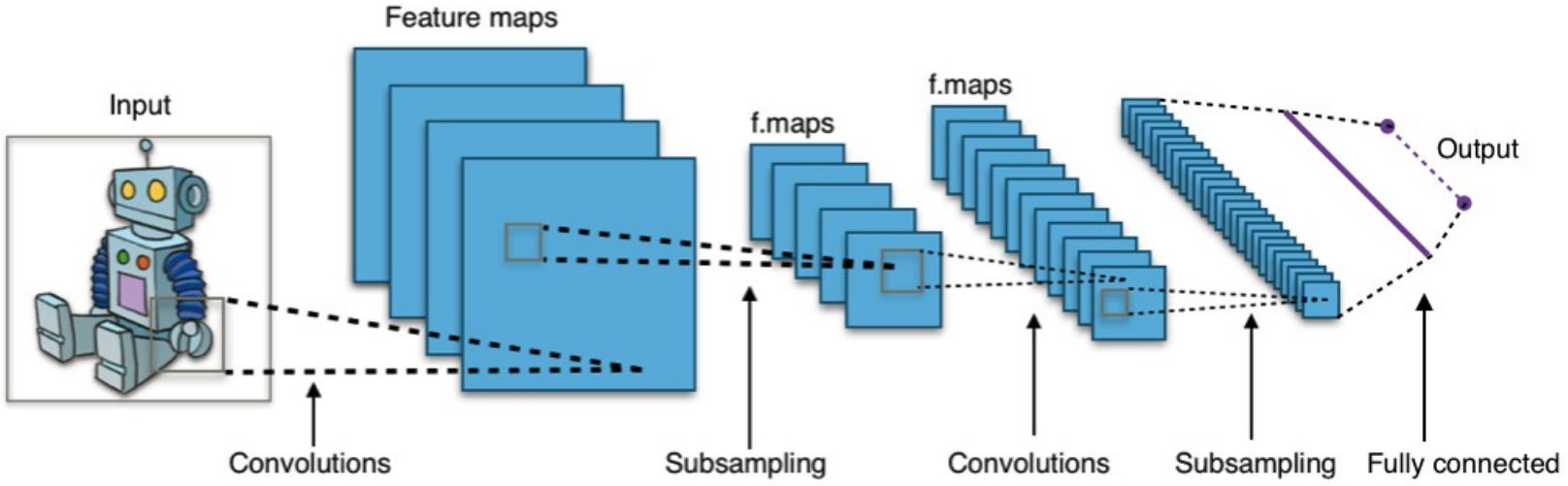
- Deep structured learning, hierarchical learning.
- Ensemble de méthodes d’apprentissage automatiques.
- Le but est de modéliser des données avec un haut niveau d’abstraction.
- Architecture articulé en différentes transformations non linéaires.
- Souvent utilisé dans l’imagerie ou le traitement du signal.



Réseau de neurones



- Inspiré du fonctionnement des neurones biologiques.
- Approche statistique.
- Méthode d’apprentissage de type probabiliste.
- Mécanisme perceptif indépendant de l’implémenteur.



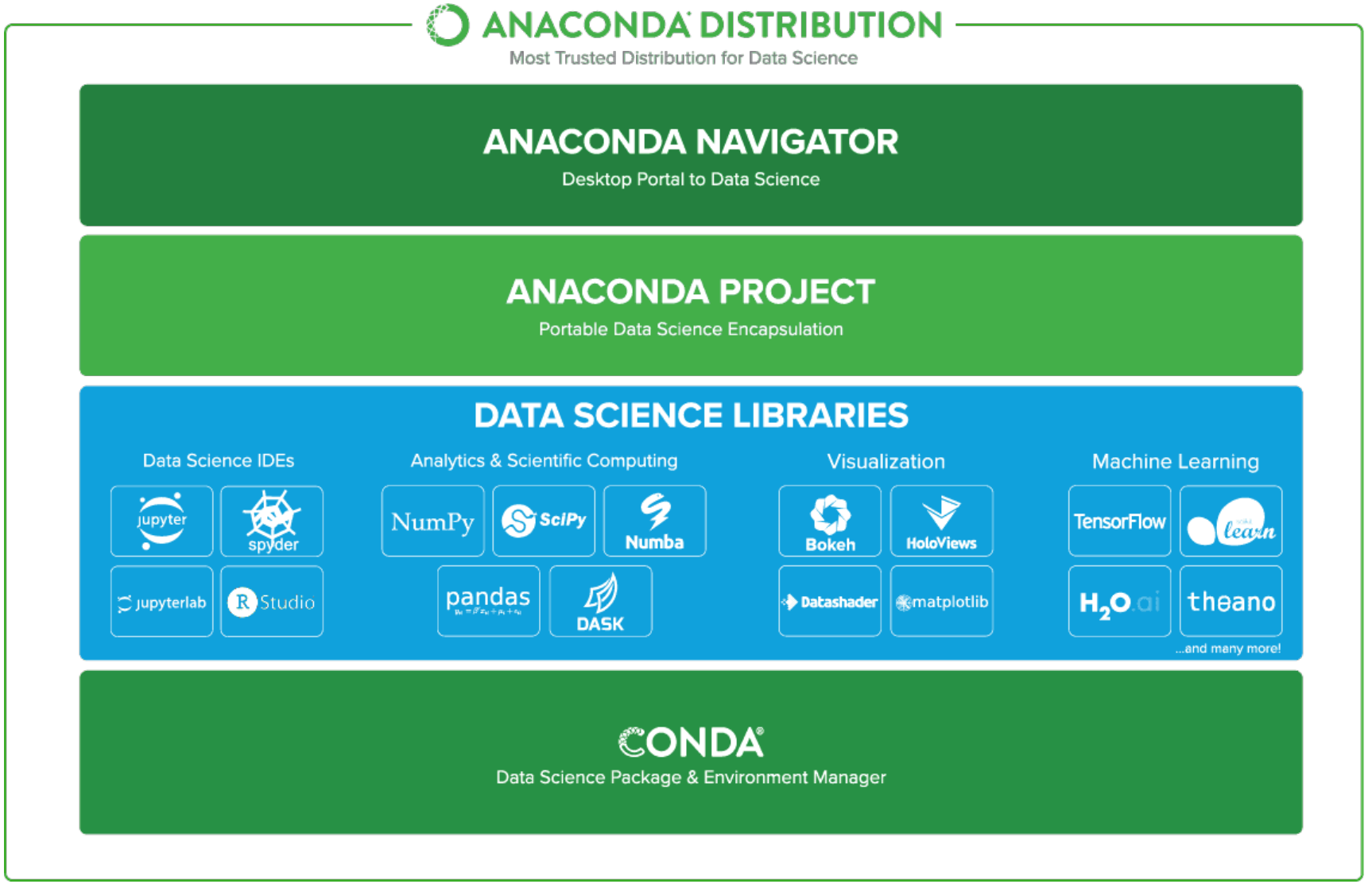
Technologies

- Language de programmation : Python 3.
- Distribution : Anaconda.
- Librairie : Pandas, Numpy, Matplotlib, scikit-learn, keras.
- Gestionnaire de version : Git.
- IDE : Vim, VSCode, Google Colab.





Création de l'environnement



Les données

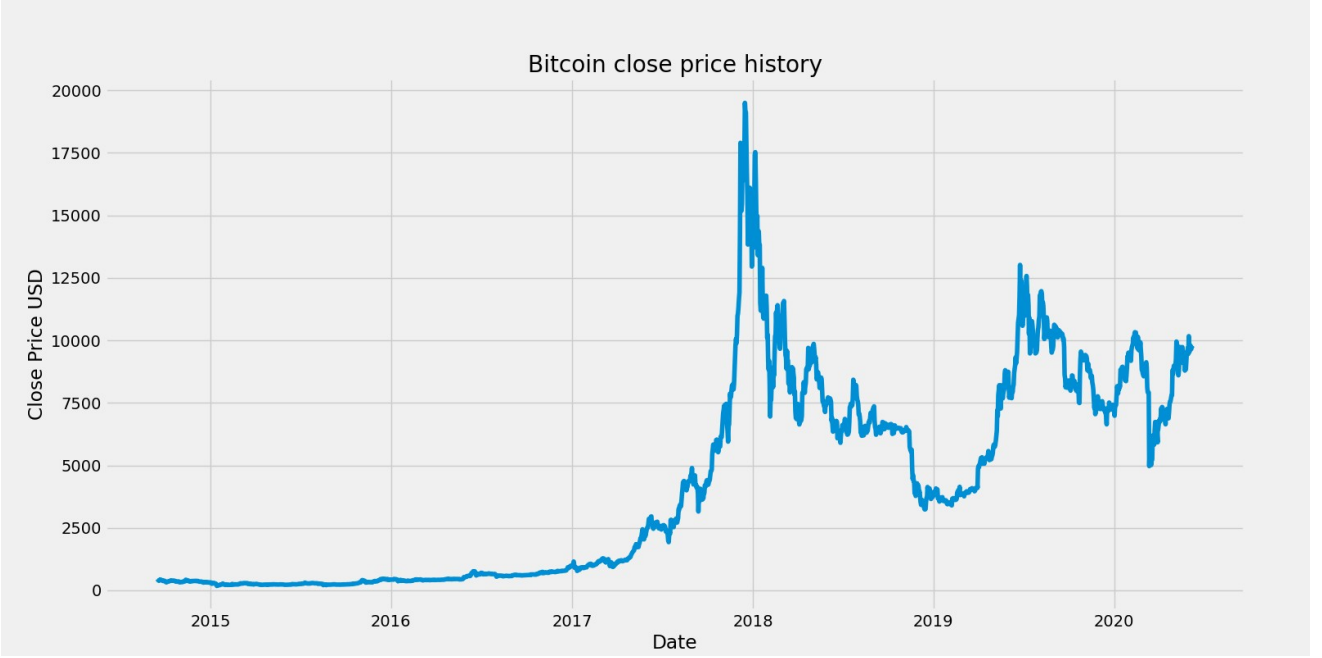
Création du dataset

- Utilisation de la libraire « pandas_datareader »
- Récupération des données du site yahoo finance.
- Récupération de 5 ans de données avec une date par ligne.
- Export en CSV possible mais moins efficace que la manipulation directe du dataframe.
- Manipulation du dataframe afin de choisir la colonne qui nous interesse : Prix à la fermeture du marché journalier.

YAHOO!

Visualisation des données

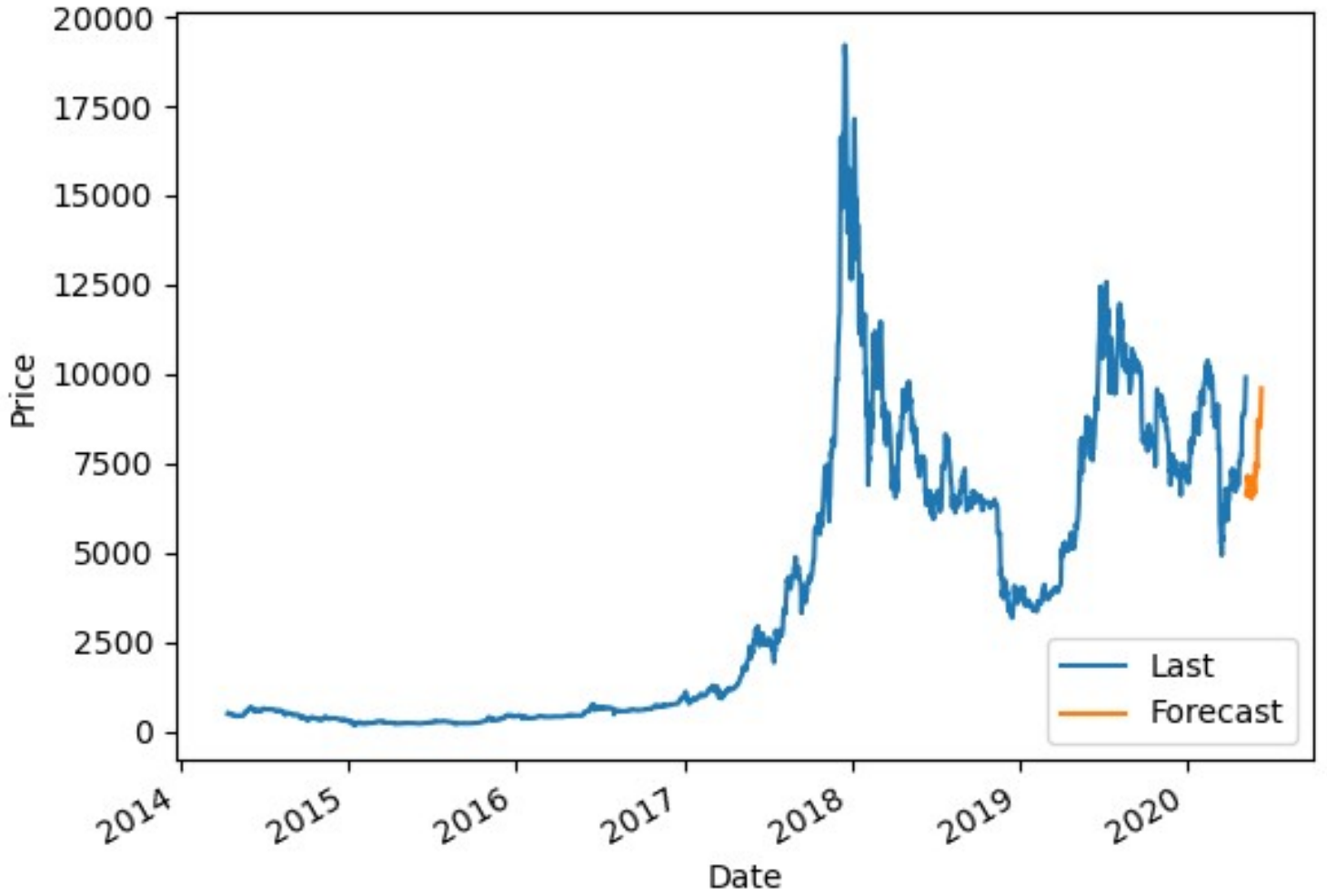
- Utilisation de matplotlib afin de visualiser les données.
- Choix du format.
- Attribution des axes X et Y respectifs.
- Sélection de la colonne du dataset à représenter.
- Attribution des noms pour chaque axe.
- Sauvegarde de la courbe.



Le model

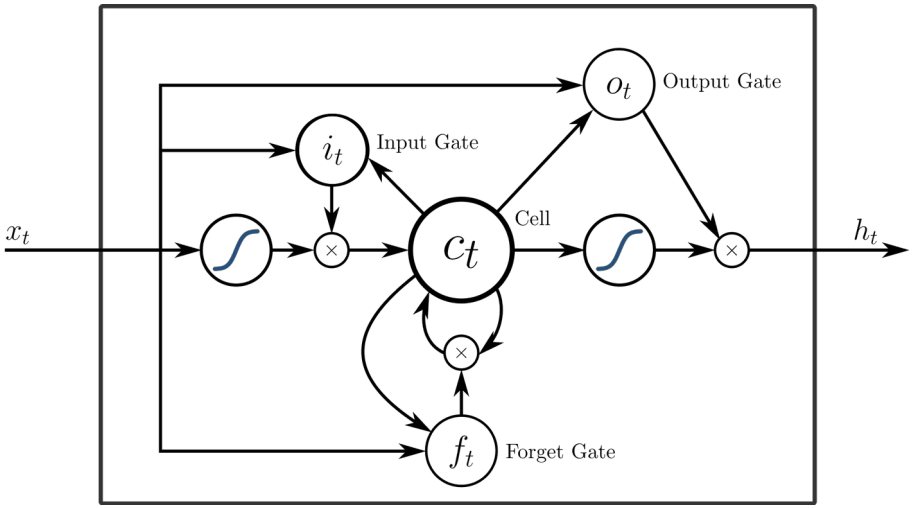
Choix du model

- Algorithme de régression linéaire :



Choix du model

- Nous avons opté pour un model « long short-term memory » (LSTM) de la librairie « keras ».
- Utilise un réseau de neurones récurrent (RNN).
- Il possède des connections de feedback contrairement au réseau « feedforward ».
- Il peu process des sets entier de données contrairement à d'autre réseaux de neurones qui ne traitent qu'un unique point de donné (tel qu'une image).



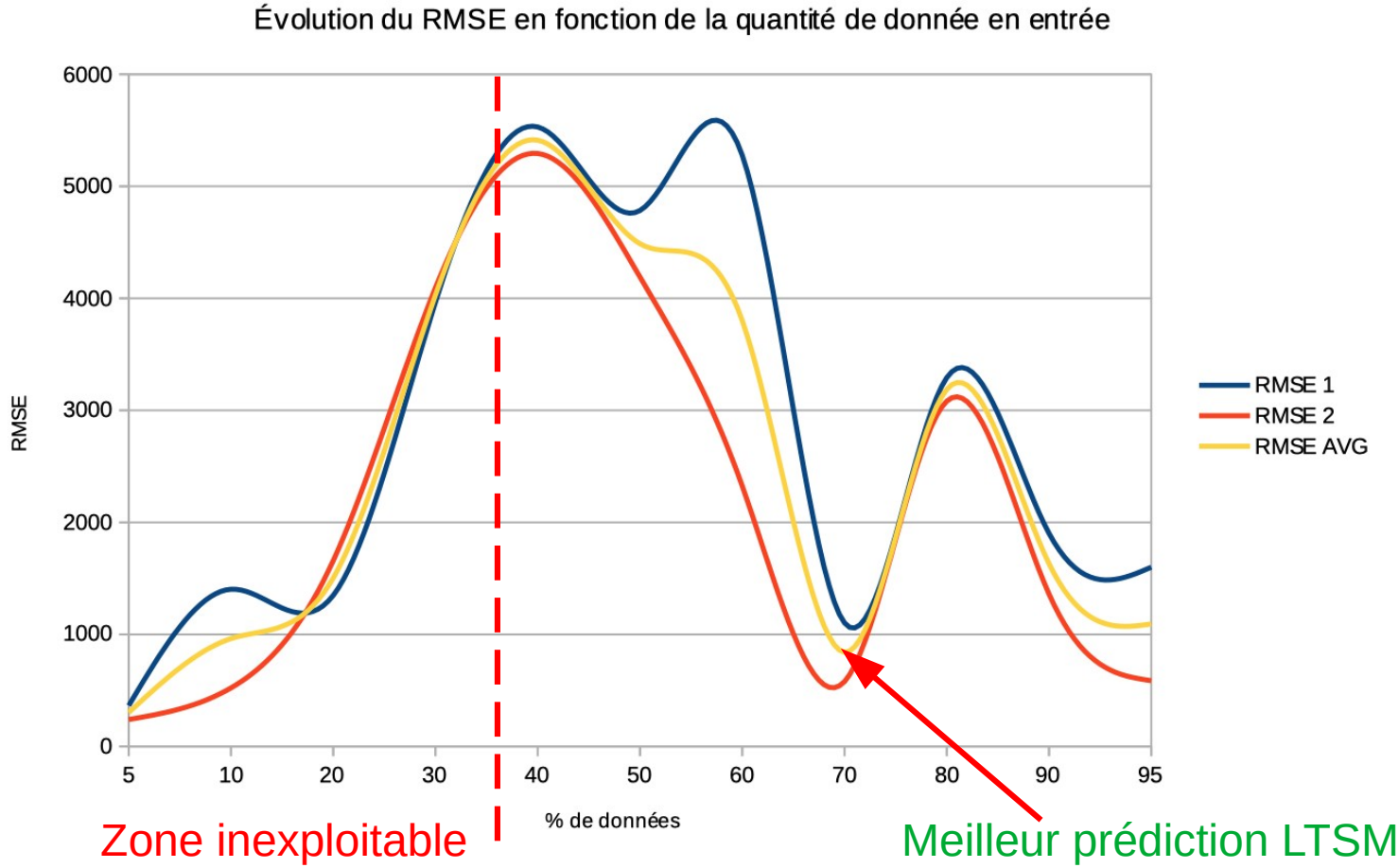
Préparation et utilisation du model

- Scaling des données en 0 et 1.
- Création du dataset d'entrainement en tant que portion du dataset d'origin.
- Utilisation de tableau numpy.
- Choix du mode Séquentiel pour le model.
- Présentation du dataset à l'input du model.
- Récupération du dataset en sortie du model.
- Etablissement de l'erreur à l'écart type (Root mean square error) et affichage de l'erreur.

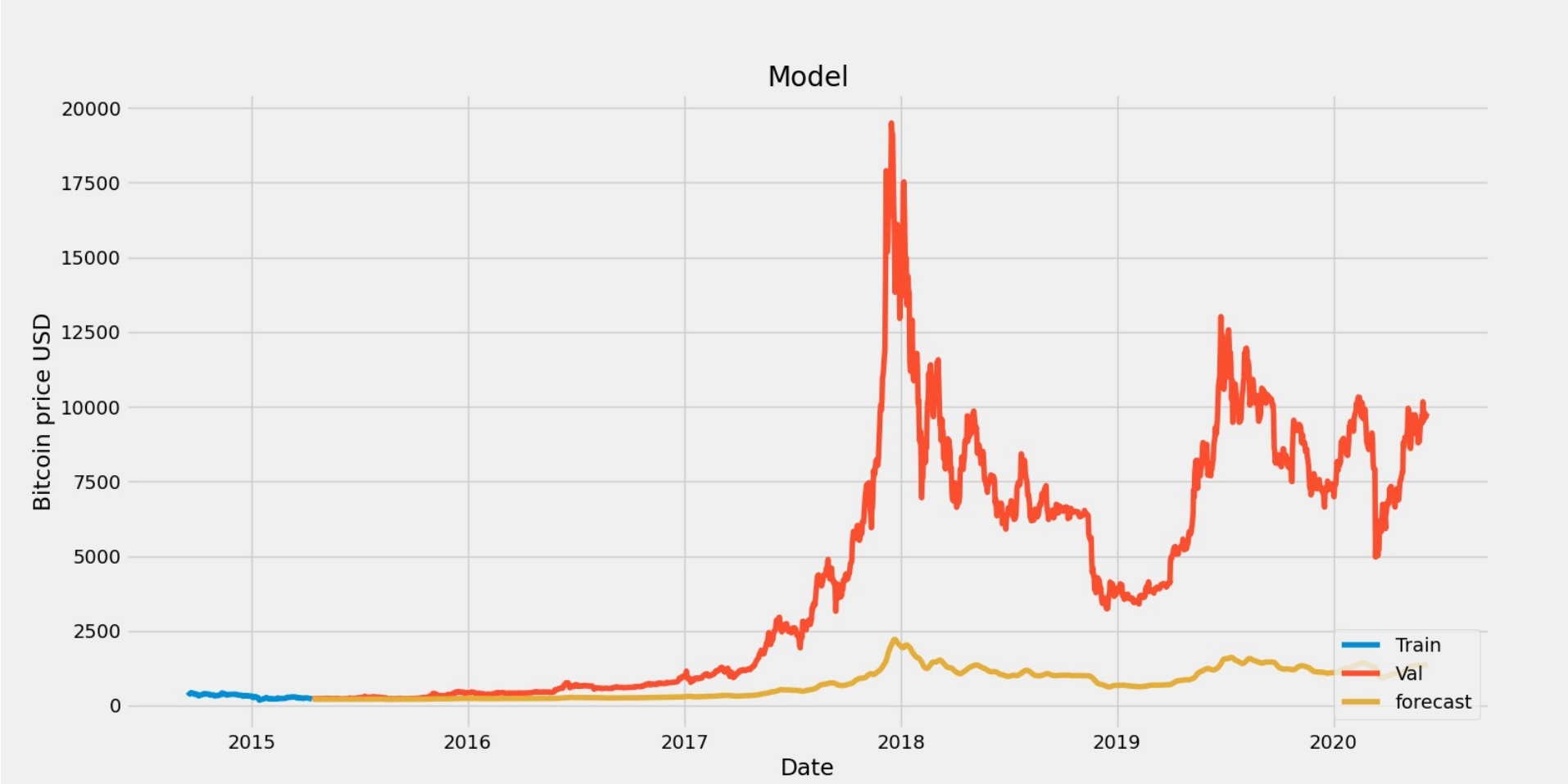
Analyse

Évolution de l’écart type du model LTSM

- Evolution du taux d’erreur du model LTSM en fonction de la quantité de donnée d’entrainement :

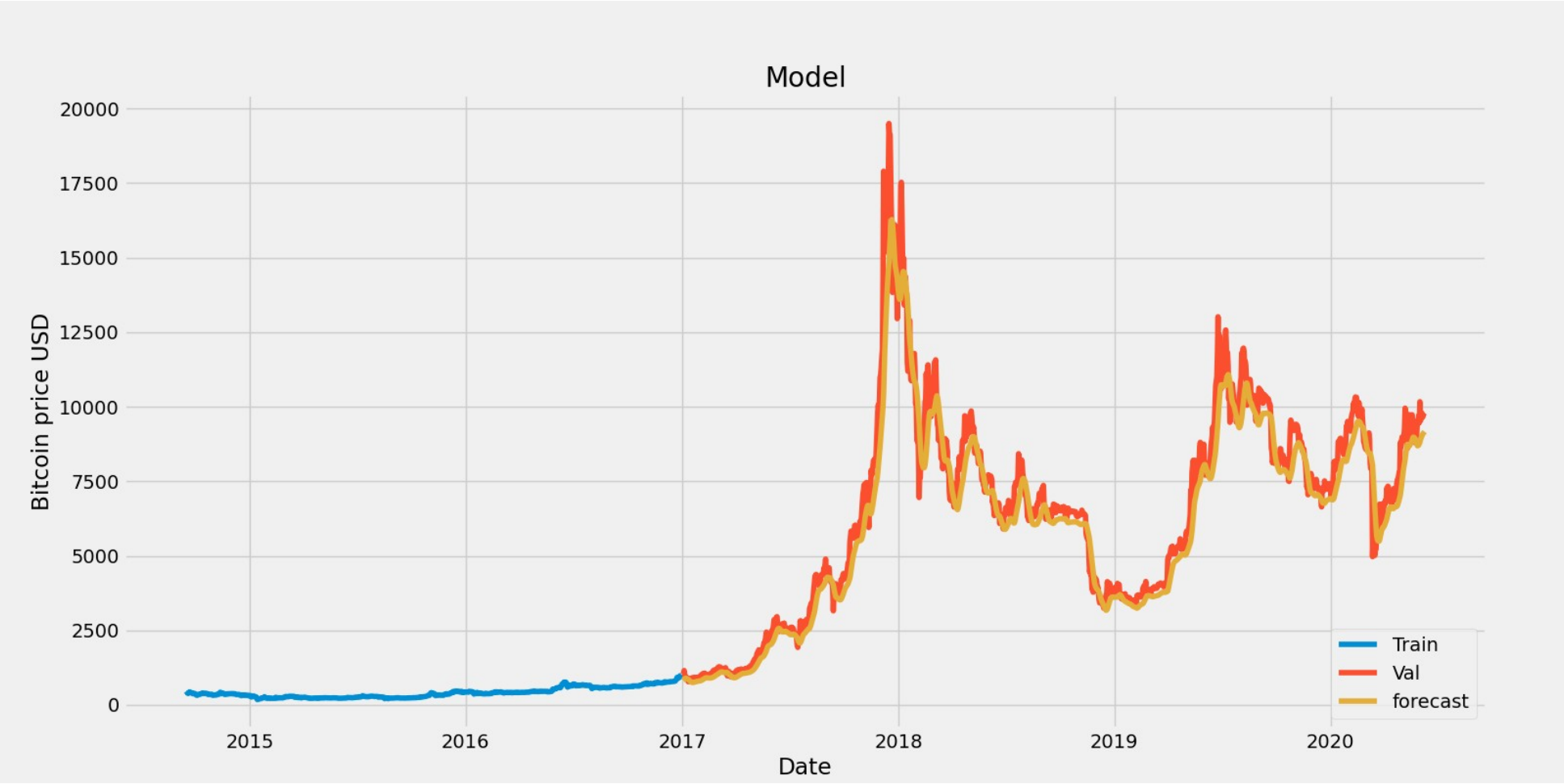


Zone inexploitable : traduction visuelle.



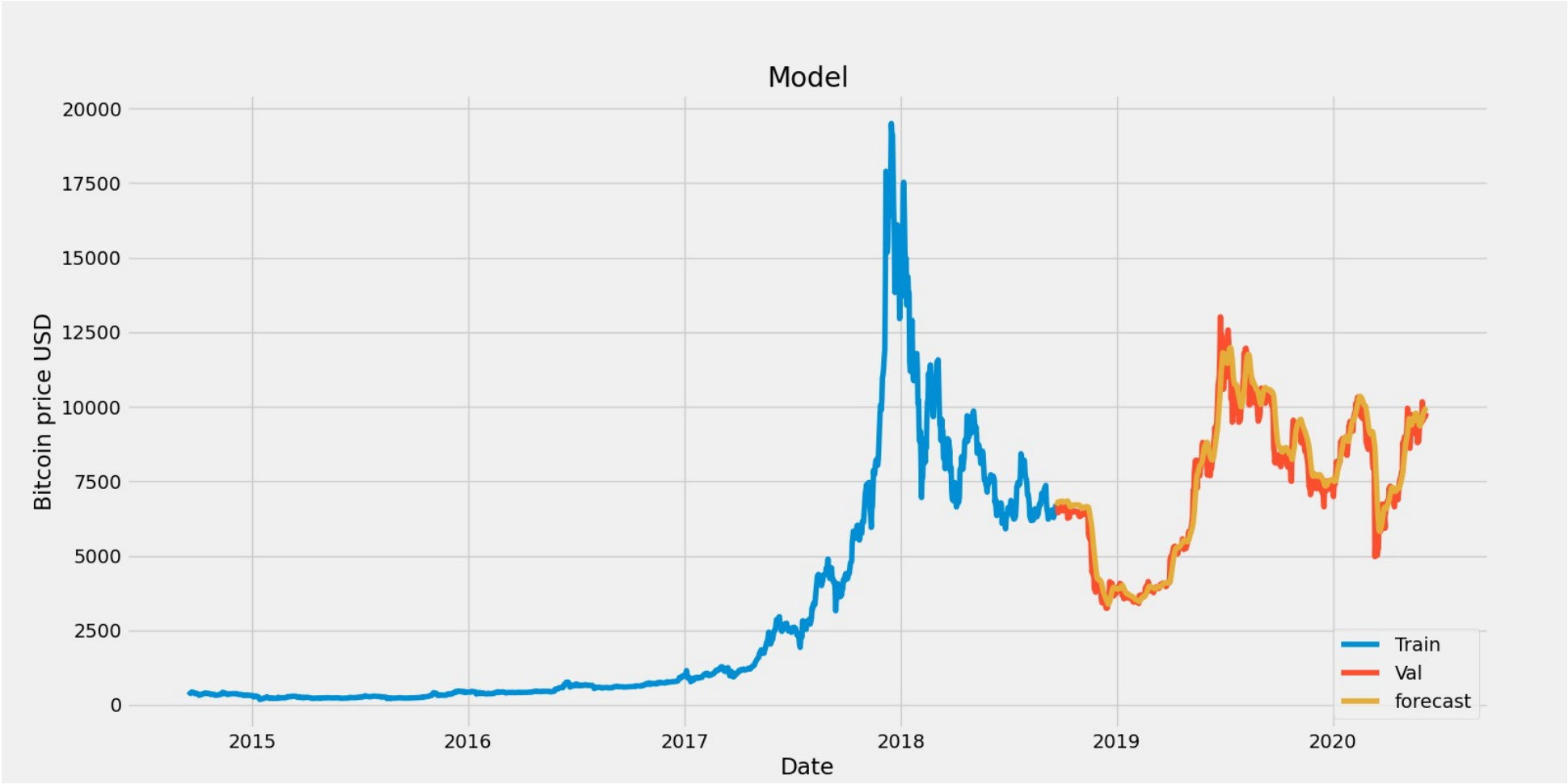
Test réalisé avec un entrainement sur 10 % des données.

Quantité de donnée acceptable



Test réalisé avec un entrainement sur 40 % des données.

Meilleure prédiction



Test réalisé avec un entrainement sur 70 % des données.

Conclusion

Conclusion

- Intérêt réel du machine learning pour la prédiction des cours boursiers.
- Utiliser un model c'est bien, tester sa cohérence c'est mieux.
- Possibilité d'incohérence du résultat du model en fonction du bruit en entrée, ou d'incompatibilité avec le besoin métier.
- Model LSTM particulièrement efficace pour effectuer une prédiction à court terme du court.
- Possibilité d'améliorer la prédiction en jouant sur le bruit de donnée en entrée et la répétition des entrainements.
- Possibilité d'utiliser notre programme pour évaluer un investissement sur le très court terme.

Démonstration