



Prédiction de la valeur de l'action Google dans le futur

Préparé et présenté par
Nidal – Zohra – Cyril & Théo



Datascientist

CONTEXTE

PRÉSENTATION DES DONNÉES

ANALYSE DES DONNÉES

DÉFINITION LSTM

PRÉPROCESSING

MODÈLE

PRÉDICTION

CONCLUSION

TRAVAIL D'ÉQUIPE



SOMMAIRE

Contexte



DataScientist chez Google

Nous sommes en charge du nouveau projet autour des TimeSeries basée sur les réseaux de neurones récurrents de type LSTM.

Notre projet consiste à proposer une solution qui va permettre de prédire la tendance de l'action Google dans le futur.

Et par la suite, former une nouvelle équipe de Datascientists juniors, sur les RNN et en particulier les LSTM.

Présentation des données



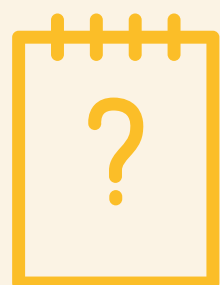
Dataset Train

Google_Stock_Price_Train
1258 rows × 6 columns



Date

2012 - 2016



0

Valeurs manquantes et
Duplicates



Dataset Test

Google_Stock_Price_Test
20 rows × 6 columns



Date

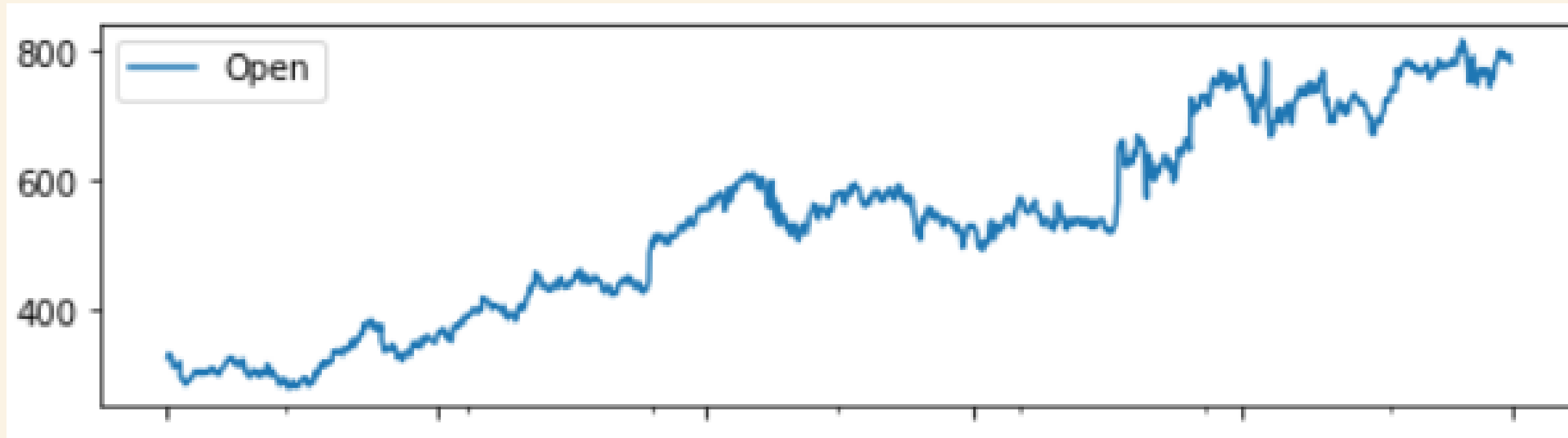
2017



0

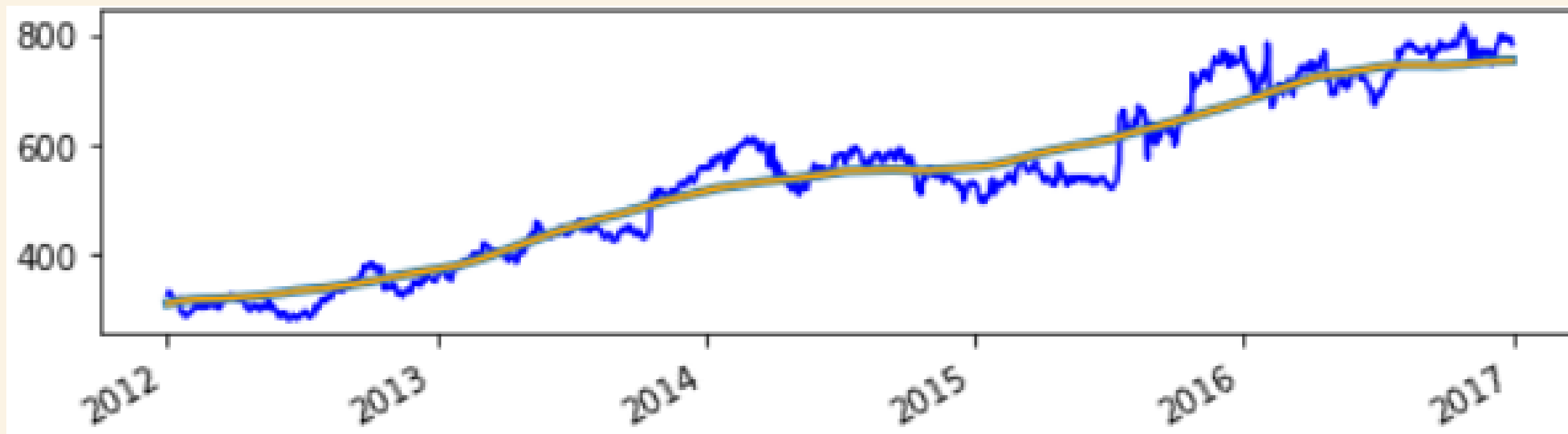
Valeurs manquantes et
Duplicates

Analyse des données



Observation :

Une certaine tendance qui augmente au cours du temps. Nous allons travailler sur la recherche de sa trend.



Observation :

Il y a deux allures différentes. Une croissance assez élevée et une période où les actions stagnent mais une croissance basse.

Preprocessing

Target

Prix d'ouverture de l'action : 'Open'

Scaling

Normalisation : MinMaxScaler

Training

Entraînement du modèle dans des données de 60 jours à la fois

Reshaping

Remodeler et mettre sous forme d'un tableau tridimensionnel



LSTM

Définition rapide :

LSTM (Long Short Term Memory) : est une architecture de réseau de neurones récurrents utilisé dans le domaine de l'apprentissage en profondeur

Une **unité** LSTM de base est composée d'une cellule, d'une porte d'entrée, d'une porte de sortie et d'une porte d'oubli.

Recherches faites en amont pour plus d'explication :

[*Lien_Veille_LSTM*](#)



Construire le modèle LSTM

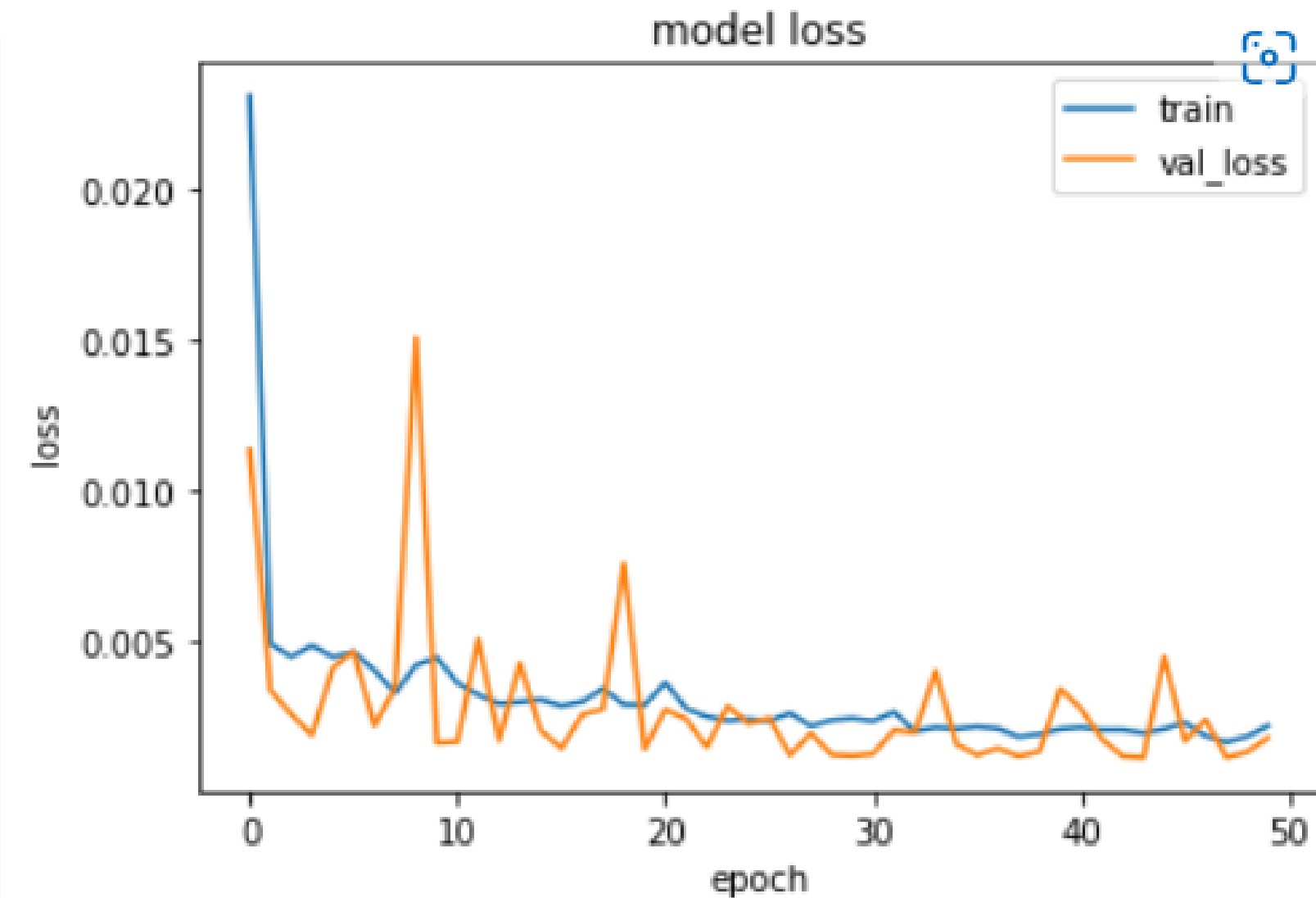
Couches nécessaires pour construire notre réseau neuronal récurrent :

- ▶ Nous utiliserons le modèle **Sequential** car il est préférable d'utiliser ce modèle lorsque chaque couche sera dotée d'un tenseur d'entrée et aura un tenseur de sortie.
- ▶ La première couche LSTM qui aura **60 unités**.
- ▶ La **return_sequences** est définie sur **True**, afin de renvoyer la séquence complète dans la sortie.
- ▶ Le **Dropout** ignore les unités (neurones) pendant l'entraînement de manière aléatoire.
- ▶ La dernière couche, **Dense**, nous affirme que les neurones de cette couche sont entièrement connectés aux neurones de la couche précédente.
- ▶ Le nombre **d'unités** dans la dernière couche doit être défini sur **1** nous prédisons une valeur unique

Ajustement du modèle

On compile le modèle en utilisant :

- ▶ Un **optimizer adam** pour l'optimisation du modèle.
- ▶ Une fonction de perte qui sera l'erreur quadratique moyenne, **mean_squared_error**.
- ▶ Une **époque** de **50** pour former ce modèle.
- ▶ Un **batch_size** de **32** qui indique le nombre d'échantillons d'entraînement



Puisque nous prédisons le cours de l'action en termes des 60 jours précédents, nous stockerons les prix des actions pour ces jours dans la variable. Pour prédire le cours de l'action pour le premier jour dans notre ensemble de données de test, nous prendrons les données des 60 derniers jours à partir de nos données d'entraînement.



**Prédiction
avec les
données
Test**

Conclusion

Si la précision du modèle est **augmentée**, cela signifie que vous serez en mesure de découvrir le comportement du futur marché, vous aidant à savoir exactement **quand** et dans **quelles** actions investir.

Les perspectives d'améliorations :

- ▶ **Alimenter le réseau neuronal LSTM**
- ▶ **Tester d'autres modèles**
- ▶ **Tester d'autres métriques**
- ▶ **Enrichir la base de données**



Confiance et Sécurité

- ▶ **Sécurité psychologique** : capacité à s'exprimer et à prendre des risques ou à se mettre en situation de vulnérabilité en face des autres.
- ▶ **Fiabilité** : confiance partagée dans les objectifs à atteindre.
- ▶ **Clarté des structures et des buts** : l'équipe partage des rôles et des objectifs clairs.
- ▶ **Sens** : le travail en équipe est important pour chacun.
- ▶ **Impact** : l'équipe pense que son travail compte et a un impact.

Confiance et Qualité

Pour Dan Teran, il faut penser au long terme :

- ▶ Payer mieux ces employés
- ▶ Apporter un service de meilleur qualité
- ▶ Critères d'embauche : optimisme et empathie





Merci !

Des questions ?