Résultats modèles

* LSTM (Gold) : (avec suppression des données manquantes, normalisation + mise à l’échelle des données)

Résultat obtenu en réalisant 10 essais (de 200 époques chacun) afin de rechercher les meilleurs hyperparamètres et les meilleurs performances

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Cela signifie que, en moyenne, l'erreur quadratique moyenne (MSE) sur l'ensemble de validation est d'environ 0.0403, et la perte moyenne est d'environ 0.4039 (obtenu au 10e essai)

Le meilleur val\_loss obtenu jusqu'à présent est de 0.2312. Cela signifie que lors d'un essai antérieur, le modèle a atteint une meilleure performance avec une perte de validation plus faible

(Pour rappel, les données de validation sont des données que le modèle n’a jamais rencontrées.)

loss (val\_loss pour l’ensemble de validation) est la mesure de l’erreur entre les valeurs prédites par le modèle et les vrais valeurs.

Récap : (results\_summary()) classés par ordre décroissant de performance

(Si on regarde les autres essais, on voit que val\_loss et loss sont souvent assez proches ce qui veut dire que le modèle généralise assez bien les données (pas de surapprentissage))

Problème : total elapsed time est trop élevée (17 min) => peut être du au nb élevé d’époques, si l’ordinateur est peu performant etc..

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Exemple : essai 4

Hyperparamètres :

Input\_unit : nb neurones dans la 1ere couche

N\_layer : nb couches

Dropout\_rate : tx de désactivation aléatoire des neurones (ici 5% des neurones seront désactivés de manière aléatoire après chaque apprentissage pdt l’entrainement)

Lstm\_2\_units et lstm\_3\_units ne sont pas utilisés dans cet essai car il n’y a que 2 couches

Lstm\_0\_units : nb de neurones dans la 1ere couche lstm après la couche d’entrée

Une image contenant texte, capture d’écran, Police, conception

Description générée automatiquement

Prédiction prix de l’or sur les 500 derniers jours du dataset

Une image contenant capture d’écran, texte, Tracé, diagramme

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

* **Type de modèle**: Sequential, ce qui signifie que les couches sont empilées les unes sur les autres de manière séquentielle.
* **Couches**:
  + **LSTM**: Trois couches LSTM sont utilisées. Chaque couche LSTM est suivie d'une couche de dropout pour régulariser le modèle.
  + **Dropout**: Une couche de dropout est utilisée pour éviter le surapprentissage.
  + **Dense**: Une couche Dense avec une seule sortie. Cela semble être la couche de sortie qui prédit une seule valeur de sortie.
* **Paramètres**:
  + **Total params**: Le nombre total de paramètres du modèle, y compris les poids et les biais.
  + **Trainable params**: Le nombre de paramètres qui sont entraînables lors de l'apprentissage.
  + **Non-trainable params**: Le nombre de paramètres qui ne sont pas entraînables. Dans ce cas, il semble que tous les paramètres soient entraînables.
  + **Optimizer params**: Le nombre de paramètres de l'optimiseur utilisé pour entraîner le modèle.
* ARIMA (prix du pétrole de 2000 à 2024)

Résultats :

Une image contenant texte, capture d’écran, nombre

Description générée automatiquement

* Les coefficients des termes AR et MA dans votre modèle sont proches de zéro et ont des p-valeurs élevées, ce qui indique qu'ils ne sont pas significativement différents de zéro. Par exemple, le coefficient AR.L1 est de 0.1059 avec une p-valeur de 0.964, ce qui suggère qu'il n'y a pas de relation significative entre la série temporelle et sa propre valeur retardée de 1 période. De même, les coefficients pour les termes MA sont également non significatifs.
  + Les coefficients non significatifs et les diagnostics des résidus indiquent que le modèle ARIMA(2, 1, 2) ne capture peut-être pas correctement la dynamique de la série temporelle. Cela peut être dû à des spécifications inappropriées du modèle ou à des caractéristiques non capturées de la série temporelle, telles que des tendances non linéaires ou des saisons.
* Modele SARIMA (avec prise en compte de la saisonnalité)