

Faculté des Sciences

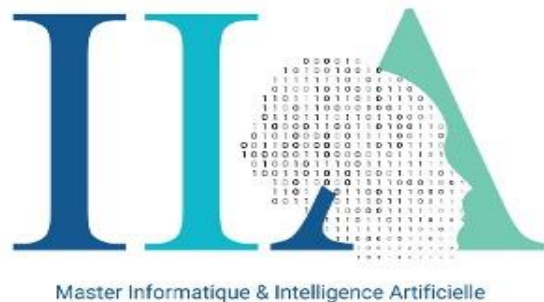
Kenitra

M2IA



Mini Projet de Module de processus stochastique

Exemples d'Application Marché financier et Santé



Réalisé par :

Lahlimi Oussama

Zeroual Ismail

Makdad Ahmed

Encadrée par :

M. Kaicer Mohammed

2022/2023

Table des matières

Introduction :	1
La différence entre “Prévision” et “Prédiction”:	2
1-Prévision	2
2-Prédiction	2
Exemples bien clairs pour la prédiction et la Prévision :	3
1. Exemple de Prévision	3
2. Exemple de Prédiction	4
Le Stochastiques est utilisé pour “Prédire” ou “ Prévoir” :	4
La série chronologique et son domaine d’appartenance aux algorithmes de Machine Learning :	4
1. L’équation mathématique générale d'une série chronologique est la suivante :	4
2. Détails importants à considérer lors de l'analyse d'une série chronologique comprennent :	5
3. La stationnarité :	6
4. Domaines d'application :	6
Modèle ARMA (p,q) :	7
1. Modèle Autorégressive (AR) :	8
2. Modèle Moyenne Mobile (MA):	8
3. Modèle ARMA :	9
4. Méthode Box-Jenkins :	10
Application de modèle ARMA sur deux bases de données Finance et Santé:	10
1. Identification du modèle :	10
2. Estimation du modèle :	11
3. Vérification du modèle :	11
4. Exemple d’application pour une base de données de Santé :	11
5. Exemple d’application pour une base de données de Finance ::	13
Les réseaux de neurones avec ARMA :	16

Introduction:

Dans un monde en constante évolution, où les décisions éclairées et la planification stratégique sont devenues des impératifs, l'analyse des données temporelles revêt une importance capitale. Les séries chronologiques, qui enregistrent des données dans l'ordre temporel, sont omniprésentes dans notre vie quotidienne, qu'il s'agisse de données financières reflétant les marchés monétaires, de données médicales enregistrant l'état de santé de la population, ou encore de données climatiques retraçant les variations météorologiques.

Au cœur de cette démarche d'analyse, les modèles ARMA (AutoRegressive Moving Average) se révèlent être des outils d'une importance indéniable. Ces modèles, issus de la théorie des séries chronologiques, conjuguent rigueur statistique et puissance prédictive pour nous permettre de décrypter et d'anticiper des tendances, de détecter des modèles récurrents et de comprendre des phénomènes complexes.

Ce rapport se consacre à l'exploration et à l'application des modèles ARMA dans deux domaines cruciaux : le marché financier et la santé. Au-delà de leur application spécifique, notre objectif est d'illustrer comment ces modèles offrent des solutions à des problèmes concrets et variés.

Nous commencerons par jeter un regard sur la différence subtile, mais fondamentale, entre la "prédiction" et la "prévision", des concepts qui sous-tendent la philosophie même de l'analyse des séries temporelles. Nous explorerons ensuite en profondeur les séries chronologiques en tant que sources riches d'informations sur les tendances, les saisons et les fluctuations.

Nous plongerons ensuite dans les équations du modèle ARMA, éléments-clés de cette analyse. En comprenant ces équations, nous serons mieux armés pour aborder les applications concrètes.

Enfin, nous nous concentrerons sur les applications spécifiques des modèles ARMA dans le marché financier et le domaine de la santé. Dans chacun de ces domaines, nous découvrirons comment les modèles ARMA peuvent contribuer à des prises de décision plus éclairées, à une gestion plus efficace des ressources et à une anticipation des évolutions futures.

À travers cette exploration, nous démontrerons que l'analyse des séries temporelles et l'application des modèles ARMA sont au cœur de la transformation des données brutes en informations exploitables, contribuant ainsi à des décisions plus informées et à un avenir meilleur.

I. La différence entre "Prévision" et "Prédiction" :

Prévision et prédiction sont des paronymes. Ce sont des mots dont la prononciation et la graphie se ressemblent mais dont le sens est sensiblement différent. Ainsi que Les termes "prédiction" et "prévision" sont souvent utilisés de manière interchangeable dans le langage courant, mais en statistiques, en analyse des données et en planification, ils ont des significations légèrement différentes.

1. **Prévision :**

Définition : Faire une prévision, c'est prévoir suite à un raisonnement (Logique Mathématique) qu'une chose va se produire. On fait des prévisions basées sur un calcul (méthodes analytiques ou statistiques). Les prévisions ont une forte probabilité d'arriver bien que le futur soit toujours incertain.

La "prévision" se réfère généralement à l'estimation de valeurs futures en se basant sur des données historiques et des tendances passées. Elle implique l'utilisation d'outils statistiques pour anticiper des résultats potentiels.

Les synonymes de « prévision » sont « probabilité, prospective, anticipation » ainsi que « divination » lorsque la prévision a une origine magique.

2. **Prédiction :**

Définition : Faire une prédiction, c'est annoncer qu'un événement ayant peu de chance de se passer arrivera, c'est dire qu'une chose se déroulera malgré son improbabilité. On peut faire des prédictions à partir d'une intuition, d'un raisonnement personnel, d'une prémonition, à partir de conjectures...

La "prédiction" consiste à faire des déclarations spécifiques sur les résultats futurs, souvent en utilisant des modèles et des données pour évaluer des événements à venir.

Synonymes de « prédiction » : divination, prophétie, augure (et plus rarement prévision).

Prévision	prédiction
Forecasting	Predictive
Series Temporalis	Estimation Decision Ranking
ARMA , ARIMA	Regressions Arber

Le table suivante montre La différence entre "Prévision" et "Prédiction"

En résumé, bien que les termes "prévision" et "prédiction" puissent être utilisées de manière similaire dans de nombreux contextes, la "prévision" tend à avoir une connotation plus objective, basée sur des données et des méthodes analytiques, tandis que la "prédiction" peut inclure une dimension plus subjective ou spéculative, parfois basée sur des facteurs non quantitatifs. Cependant, il n'y a pas de règles strictes concernant l'utilisation de ces termes, et ils peuvent être interchangeables dans de nombreuses situations courantes.

II. Exemples bien clairs pour la prédiction et la Prévision :

1. Exemple de Prévision :

Faire des prévisions météorologiques précises.

Faire des prévisions économiques à partir de données statistiques stables.

2. Exemple de Prédiction :

Une banque, accorde ou refuse un crédit à un individu ou à une entreprise

Prédire le prix d'une maison en fonction de ses caractéristiques.

III. Le Stochastiques est utilisé pour "Prédire" ou " Prévoir" :

Le Stochastique est principalement utilisée pour "**prévoir**". Les mouvements potentiels des prix dans les marchés financiers, tels que les actions, les devises ou les matières premières.

Le Stochastique est un indicateur technique couramment utilisé en analyse technique pour évaluer la position actuelle d'un prix par rapport à ses plus hauts et plus bas récents

IV. La série chronologique et son domaine d'appartenance aux algorithmes de Machine Learning :

Série chronologique :

Ensemble des observations d'une variable statistique économique faites à intervalles réguliers dans le temps (année, trimestre, mois, jour...)

Série Temporelle :

Est l'observation d'une quantité X_t mesurée à des temps différents $t_1...t_p$.

1. L'équation mathématique générale d'une série chronologique est la suivante :

$$X_t = T_t + S_t + C_t + \varepsilon_t$$

- X_t : représente la valeur de la série chronologique à un moment donné t.
- T_t : est la composante de tendance, qui représente la variation à long terme ou la direction générale de la série chronologique au fil du temps.
- S_t : est la composante saisonnière, qui correspond à un phénomène périodique de période identifiée, tels que les saisons de l'année.
- C_t : est la composante cyclique, qui représente les cycles économiques ou autres tendances à plus long terme qui se répètent à des intervalles moins réguliers.

- ε_t : est la composante d'erreur ou résiduelle, qui décrit un changement qui n'est pas prévisible ni par la saisonnalité ni par la tendance. Dans l'idéal le résidu est aléatoire.

2. Détails importants à considérer lors de l'analyse d'une série chronologique comprennent :

- **Stationnarité** : Une série chronologique stationnaire est celle dont les propriétés statistiques, telles que la moyenne et la variance, restent constantes avec le temps.
- **Décomposition** : La décomposition d'une série chronologique implique de séparer les composantes de tendance, saisonnière et résiduelle pour mieux comprendre chaque aspect de la série.

- **Tendance** : La tendance décrite la dynamique global de série
Exemple : Tendance à la hausse sur l'évolution des températures.
- **Saisonnalité** : Correspond à un phénomène périodique identifié.
Exemple : Les Saisons.
- **Bruite** : Le bruit (composant stochastique) décrit un changement qui n'est pas prévisible ni par la saisonnalité ni par la tendance dans l'idéal le résidu est aléatoire.
Exemple : Les catastrophes naturelles.
- **Les décompositions de bases :**

Additive:
$$X_t = T_t + S_t + \varepsilon_t$$

Multiplicative:
$$X_t = T_t * S_t * \varepsilon_t$$

- **Identification des Modèles** : L'identification du modèle implique de choisir le modèle mathématique approprié pour représenter la série chronologique.
- **Prévision et Prédiction** : L'objectif principal de l'analyse de séries chronologiques est souvent de faire des prévisions ou des prédictions pour les valeurs futures de la série, ce qui peut être utile pour la planification et la prise de décision.

3. La stationnarité :

Définition :

- Une série temporelle est stationnaire lorsque ses moments ne dépendent pas au temps.
- Ses propriétés statistiques ne varient pas dans le temps :
 - **Moyenne** : Son espérance est constante au cours du temps, il n'y a donc pas de tendance.
 - **Variance** : Mesure de dispersion des valeurs.
 - **Covariance** : Mesure de la variabilité conjointe de deux variables. Elle ne dépend pas de temps.
- Pourquoi a-t-on besoin d'une Série stationnaire ?
 - Il est plus facile de faire des **prédictions** sur série où les propriétés futures sont connues.
- On vérifie la stationnarité avec un test statique comme le test de Dickey Fuller augmenté (ADF).
- Les hypothèses de stationnarité doivent être respectées pour appliquer les modèles AR et ARMA.

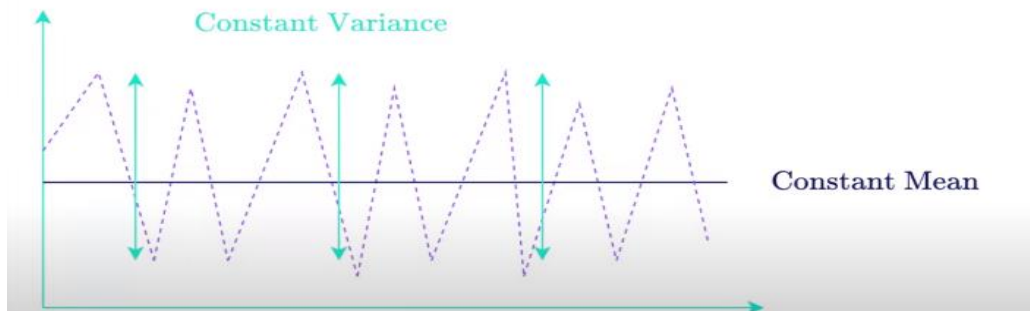


Figure représente d'une série stationnaire

4. Domaines d'application :

Les séries chronologiques sont largement utilisées dans des domaines tels que l'économie, la finance (pour la prévision des prix des actions, par exemple), la météorologie (pour prédire la météo), la santé (pour surveiller les données de patients), l'ingénierie (pour la maintenance préventive), etc.

En ce qui concerne l'apprentissage automatique (Machine Learning), les séries chronologiques sont souvent traitées à l'aide de techniques spécifiques d'apprentissage automatique appelées "modèles de séries chronologiques" ou "modèles de prévision temporelle". Ces algorithmes sont conçus pour extraire des motifs et des relations temporels à partir des données de séries chronologiques afin de faire des prédictions ou d'effectuer des analyses, notamment :

1. Modèles ARMA (AutoRegressive Moving Average) : Ils sont utilisés pour modéliser et prévoir des séries chronologiques stationnaires.
2. Réseaux de neurones récurrents (RNN) : Les RNN, notamment les LSTM (Long Short-Term Memory) et les GRU (Gated Recurrent Unit), sont adaptés à la modélisation de séquences et sont couramment utilisés pour les prévisions temporelles.
3. Modèles de régression : Ils peuvent être utilisés pour modéliser les séries chronologiques en fonction de variables explicatives.
4. Méthodes de lissage exponentiel : Elles sont utiles pour la prévision de données temporelles, en particulier lorsque des modèles plus complexes ne sont pas nécessaires.
5. Sélection de caractéristiques temporelles : L'extraction de caractéristiques pertinentes à partir de séries chronologiques peut être un préalable important à de nombreuses tâches de machine learning.

V. Modèle ARMA (p,q) :

Modèle AR, MA, ARMA :

Le modèle ARMA combine deux composantes principales : l'autorégressif (AR) et la moyenne mobile (MA). Généralement, un modèle ARMA est représenté sous la forme ARMA (p, q), où p est l'ordre de l'autorégression et q est l'ordre de la moyenne mobile.

Ordres du modèle ARMA

- **p (Ordre AR) :** L'ordre AR spécifie le nombre de termes autorégressifs à inclure dans le modèle. Il indique combien de valeurs passées sont utilisées pour prédire la valeur actuelle.

- **q (Ordre MA)** : L'ordre MA spécifie le nombre de termes de la moyenne mobile à inclure dans le modèle. Il indique combien de résidus passés sont utilisés pour prédire l'observation actuelle.

1. Modèle Autorégressif (AR) :

Définition :

Le Modèle AR consiste à modéliser notre série temporelle **à l'aide des termes passés**. Il s'agit d'une régression de la variable par rapport à elle-même.

AR (Autorégression) : L'AR fait référence à la régression linéaire des valeurs actuelles sur les valeurs précédentes de la série temporelle. Il capture les effets de dépendance linéaire à partir des valeurs passées.

Equation Autoregressive (AR):

$$X_t = C + \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t$$

$$X_t = C + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

Où :

- X_t : est la valeur de la série chronologique à un moment donné t.
- C : est une constante.
- φ_i : sont les coefficients autorégressifs.
- ε_t : est un terme d'erreur à chaque instant t.

2. Modèle Moyenne Mobile (MA):

Définition :

Le Modèle MA consiste à modéliser notre série temporelle **à l'aide des erreurs**. Il s'agit d'une régression sur les erreurs.

- **MA (Moyenne mobile)** : Le MA utilise la moyenne mobile des erreurs précédentes pour modéliser la relation entre les résidus et les observations

actuelles de la série temporelle. Il capture les effets de dépendance entre les résidus.

Equation Moyenne Mobile (MA):

$$X_t = \mu + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

$$X_t = \mu + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t$$

Où :

- X_t : est la valeur de la série chronologique à un moment donné t.
- μ : est la moyenne de la série chronologique.
- θ_i : sont les coefficients de la moyenne mobile.
- ϵ_t : est un terme d'erreur à chaque instant t.

3. Modèle ARMA :

Equation associée au modèle ARMA (p,q.):

$$X_t = C + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t$$

- Pour mettre en place ce modèle, il faut s'assurer de la stationnarité de notre série temporelle.
- Un passage au log peut suffire (Modèle Multiplicatif+ log = Modèle Additif).
- Si ce n'est pas le cas, on va alors différencier notre série afin de la stationnarité.
- ARMA (0,0) : bruit Blanc.

[Exemple ARMA : code en python](#)

4. Méthode Box-Jenkins :

Acf et Pacf :

Fonction d'autocorrélation : c'est la corrélation (Pearson) entre les valeurs d'une série.

Par exemple entre X_t et X_{t-2} cette valeur comprend à la fois effets directs et indirects

Partial Autocorrélation fonction : C'est la corrélation (Pearson) entre les valeurs d'une série isolée de l'impact des autres valeurs de la série.

Pour obtenir cette valeur, il suffit de créer un modèle linéaire et d'estimer les coefficients.

$$X_t = a_1X_{t-1} + a_2X_{t-2} + a_3X_{t-3} + V_t$$

VI. Application de modèle ARMA sur les deux bases de données de Finance et Santé:

On appliquer cette méthode pour les deux bases de données Finance et santé :

Méthodologie de Box-Jenkins le modèle ARMA

La méthodologie de Box-Jenkins est une approche couramment utilisée pour modéliser et prévoir les séries temporelles. Elle comprend les étapes suivantes :

1. Identification du modèle :

- **Analyse des données :** Examiner les données de la série temporelle pour détecter les tendances, les saisonnalités et les comportements anormaux.

- **Différenciation** : Si la série temporelle présente une tendance ou une saisonnalité, appliquer une différenciation pour rendre les données stationnaires.

- **Identification des ordres** : Utiliser les graphiques ACF (fonction d'autocorrélation) et PACF (fonction d'autocorrélation partielle) pour déterminer les ordres p et q du modèle ARMA.

2. Estimation du modèle :

- Estimation des paramètres : Utiliser les méthodes d'estimation (telles que la méthode des moindres carrés) pour estimer les paramètres du modèle ARMA.

3. Vérification du modèle :

- Diagnostic du modèle : Vérifier si les résidus du modèle ARMA sont bruit blanc (c'est-à-dire s'ils ne présentent pas de structure ou de corrélation significative).

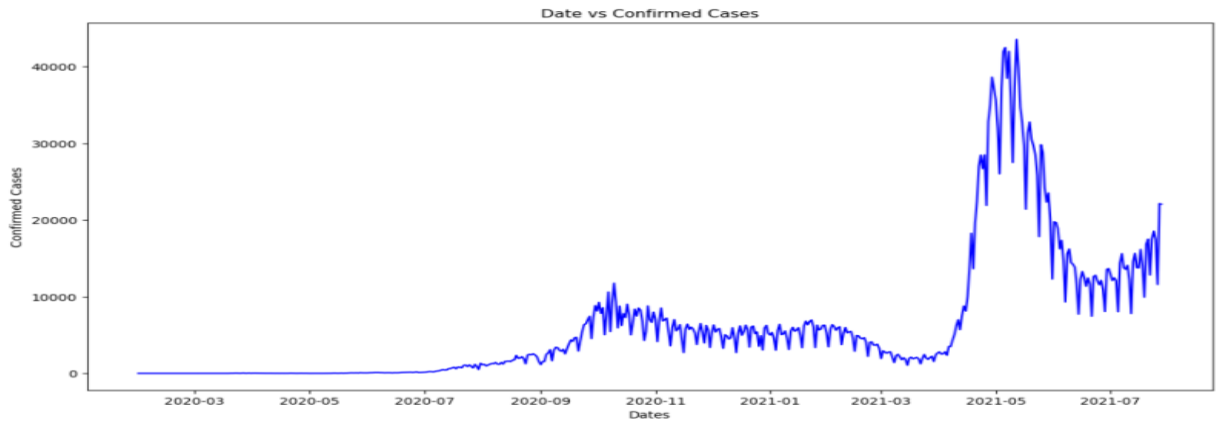
- Réajustement : Si le modèle ne satisfait pas les critères de bruit blanc, ajuster les ordres du modèle ARMA et répéter les étapes précédentes.

- Validation : Valider les performances du modèle en effectuant des prédictions sur des données de validation ou en utilisant des mesures d'évaluation telles que l'erreur quadratique moyenne (RMSE) ou le critère d'information d'Akaike (AIC).

La méthodologie de Box-Jenkins est itérative, ce qui signifie que les étapes d'identification, d'estimation et de vérification peuvent être répétées plusieurs fois pour améliorer le modèle. L'objectif est de trouver le meilleur modèle ARMA qui capture les motifs et les caractéristiques importantes de la série temporelle, et qui peut être utilisé pour effectuer des prédictions précises.

4. Exemple d'application ARMA pour une base de données de Santé :

- On travailler sur une série chronologique de évolution des cas confirmé de la pandémie Covid-19



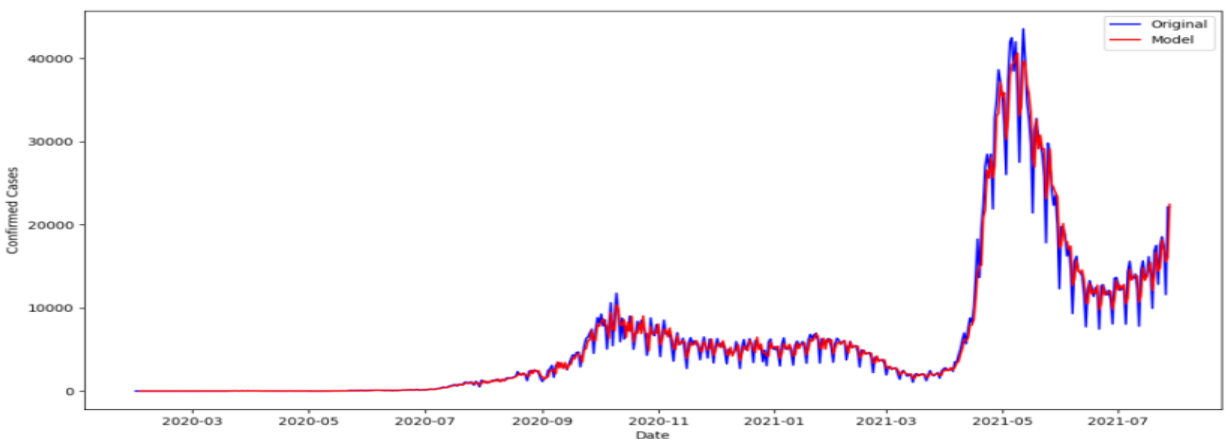
Cette figure représente notre série chronologique des cas confirmée de COVID-19

- d'après l'identification de notre modèle et l'analyse des données et identification des ordres par la méthode **arma_order_select_ic** (qui permet de sélectionner automatiquement les ordres optimaux (p, q) pour un modèle ARMA) :

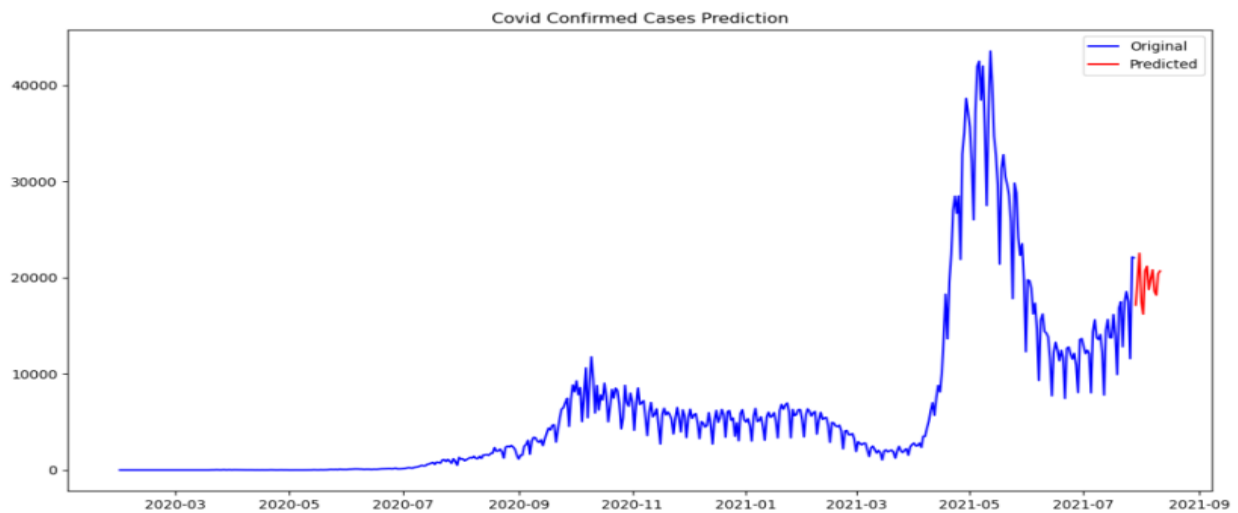
```
{'bic':
      0      1      2
0  9458.750250  9436.042797  9342.509633
1  9458.919126  9373.099893  9344.063136
2  9365.011725  9358.157852  9216.737041
3  9360.889329  9363.767099  9369.247344
4  9347.176470  9342.256723  9155.484205,
'bic_min_order': (4, 2)}
```

Cette figure représente les d'ordres optimale p=4 et q=2

Et finalement on trouve le résultat suivant :



Cette figure représente une comparaison entre la série chronologique Originale et la prédiction de notre modèle ARMA(4,2)



Cette figure représente une prévision pour les deux prochaines semaines

5. Exemple d'application ARMA pour une base de données de Finance:

- **Description de la base de données "BTC-USD" :**

yfinance : pour télécharger les données historiques du prix de la crypto-monnaie Bitcoin (BTC) en dollars américains (USD) depuis Yahoo Finance.

Prix de clôture (Close) : Le prix de clôture est l'un des champs les plus couramment utilisés pour prédire les prix futurs.

Évolution du Close le dernier prix auquel une transaction a été effectuée à la fin de la période de négociation

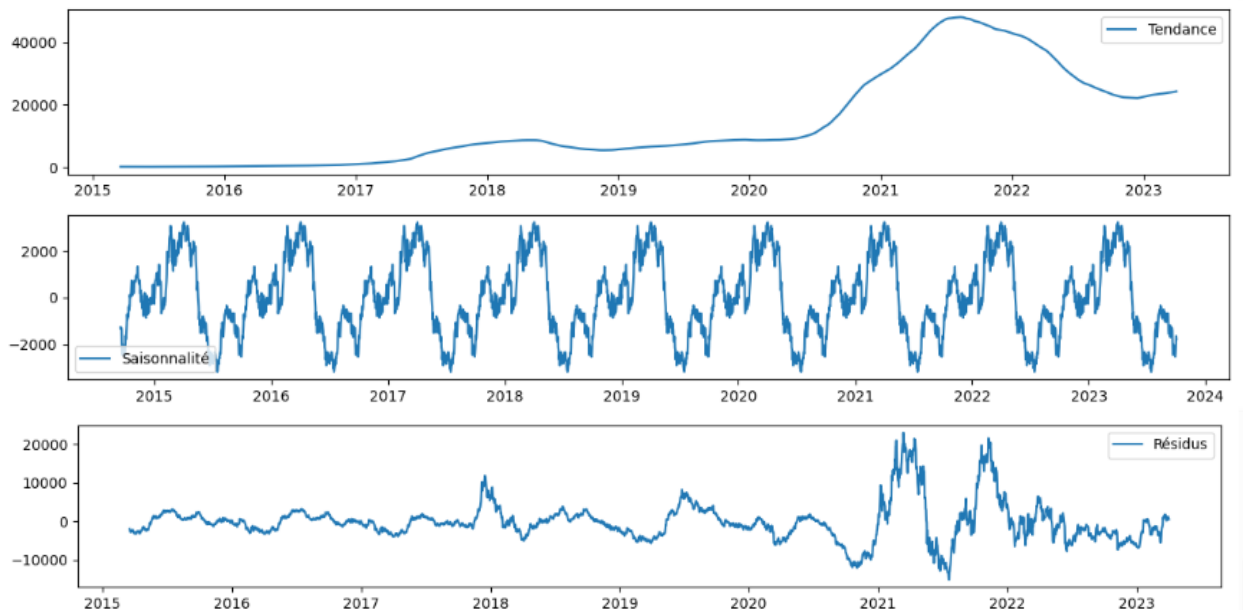


Cette figure représente Evolution du Close

Métrique	Valeur
Valeur de test	-1.5955661886632957
P-valeur	0.4858203648608006
Conclusion	La série est non stationnaire
Métrique	Valeur

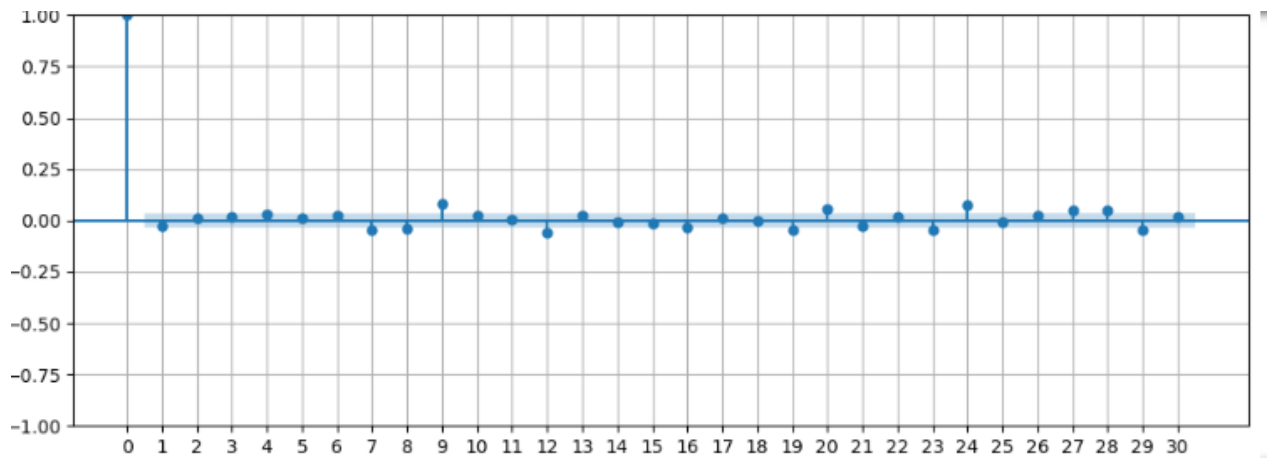
Cette figure représente teste de stationnarité de notre série chronologique Par ADF

- **Valeur de test** : Il s'agit de la statistique de test du test de Dickey-Fuller. Plus la valeur de test est négative, plus il est probable que la série soit stationnaire.
- **P-valeur**: Il s'agit de la p-valeur associée à la statistique de test. La p-valeur mesure la probabilité que les résultats du test soient obtenus par hasard. Plus la p-valeur est faible (inférieure à un seuil tel que 0.05), plus il est probable que la série soit stationnaire.
- **la décomposition saisonnière** :



Cette figure représente la décomposition saisonnière (Tendance, Saisonnalité, Résiduel) de notre série pour une période d'une année

- après notre série reçoit une série stationnaire on choisir l'ordre p et q par la méthode de ACF pour le modèle AR(p) et PACF pour le modèle MA(q).



Cette figure représente ACF - différence des prix de clôture BTC on choisir pour $p=2$ comme ordre de modèle AR



Ce figure représente un modèle ARMA(2,1) pour effectuer des prédictions sur les prix de clôture du BTC

VII. Les réseaux de neurones avec ARMA :

Les réseaux de neurones récurrents (RNN) et les modèles ARMA (AutoRegressive Moving Average) sont deux approches différentes pour modéliser les séries chronologiques. Les RNN sont des algorithmes d'apprentissage automatique qui peuvent être utilisés pour analyser les séries chronologiques et faire des prédictions sur des résultats futurs

Nous avons utilisé le modèle ARMA dans la question précédente

Nous avons obtenu ce résultat :



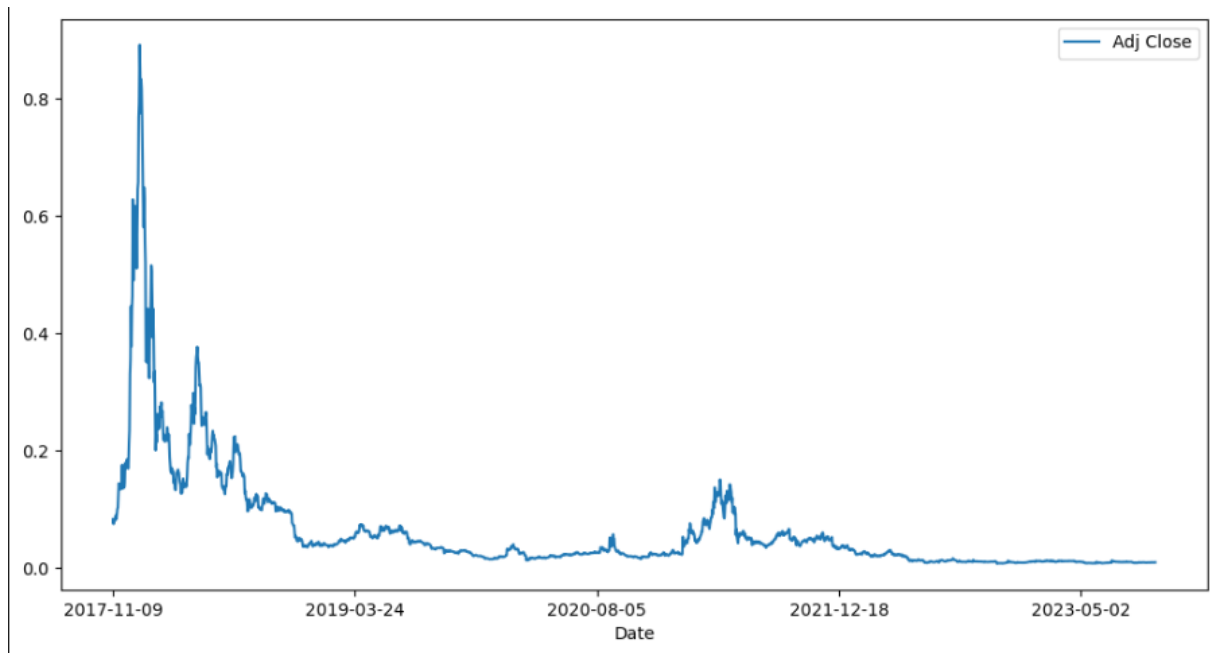
Ce figure représente un modèle ARMA pour effectuer des prédictions sur les prix de clôture du BTC

Nous essayons maintenant de répondre à la question 6 en nous basant sur les réseaux de neurones

Les réseaux de neurones récurrents (RNN) sont des algorithmes d'apprentissage automatique qui peuvent être utilisés pour analyser les séries chronologiques et faire des prédictions sur des résultats futurs, comme les prévisions boursières ou les prévisions de ventes

Nous utilisons LSTM

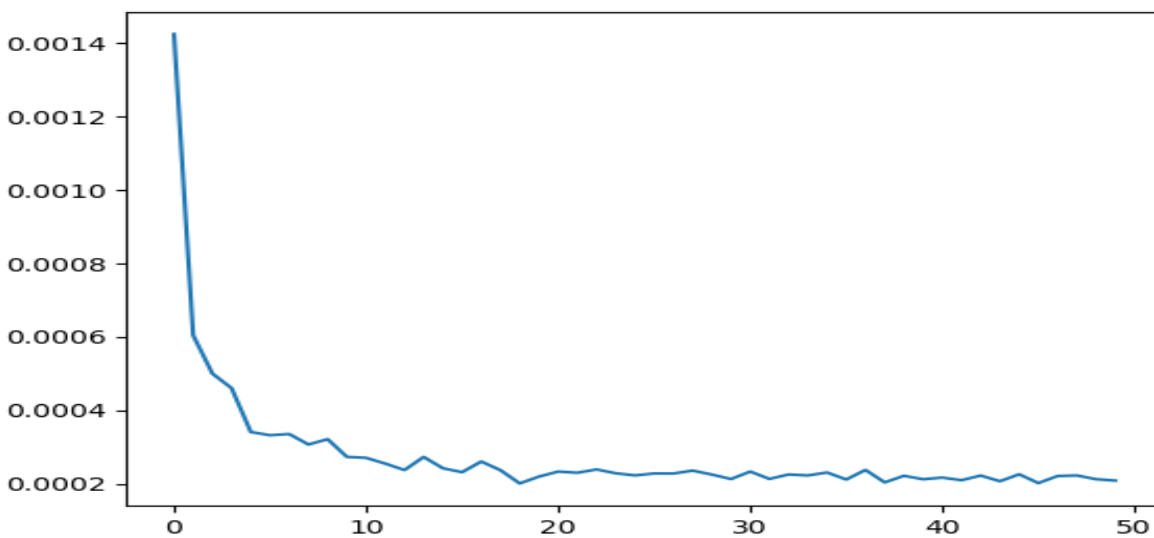
Les réseaux LSTM (Long Short-Term Memory) sont une variante des RNN qui ont été conçus pour résoudre le problème du gradient évanescent, qui peut se produire lorsque l'on entraîne un RNN sur des séquences très longues.



Représentation graphique base de données BTS-USD :

Tracer la perte par époque d'un modèle LSTM entraîné.

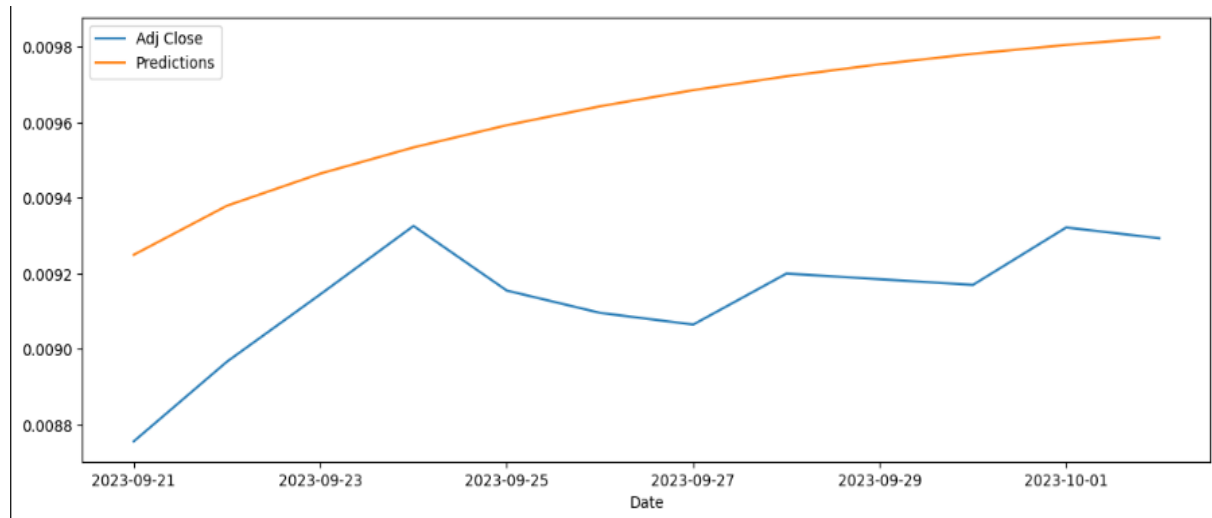
L'axe des abscisses représente le nombre d'époques, tandis que l'axe des ordonnées représente la valeur de la perte pour chaque epoch.



Prédiction du model RNN :

L'axe des abscisses représente le temps, tandis que l'axe des ordonnées représente les performances du groupe BTS. Les points bleus représentent les performances réelles du

groupe, tandis que la ligne orange représente les prévisions générées par le modèle RNN.



Dans cet exemple, prévoyez le prix de clôture de BTS en utilisant les réseaux de neurones

Nous avons essayé d'hybrider les réseaux de neurones avec ARMA, mais nous n'y sommes pas parvenus. Nous avons uniquement fait des prédictions basées sur chaque méthode séparément. Nous essaierons de l'appliquer à une autre occasion

REFERENCES

- 1) [Les processus AR, MA et ARMA - Analysez et modélisez des séries temporelles - OpenClassrooms](#)
- 2) [SERIE \(centre-univ-mila.dz\)](#)
- 3) [\(1577\) Prédiction de Séries Temporelles Avec Python et ARIMA : Guide Pratique Pour Débutants - YouTube](#)
- 4) [Prévision ou prédiction \(podcastfrancaisfacile.com\)](#)
- 5) [\(1530\) Bitcoin Price Prediction using ARIMA Model - YouTube](#)
- 6) [SERIE CHRONOLOGIQUE - YouTube](#)
- 7) [\(1577\) Modèles ARIMA pour la prévision du cours des actions X Comment choisir les termes p, d, q pour construire le modèle ARIMA \(1/2\) - YouTube](#)
- 8) [Autoregressive Moving Average \(ARMA\): Artificial data - statsmodels 0.15.0 \(+59\)](#)
- 9) [AppliedTimeSeriesAnalysisWithPython/HOTSAP_ARMA.ipynb at main · marcopeix/AppliedTimeSeriesAnalysisWithPython \(github.com\)](#)
- 10) [Processus stationnaires – modèles ARMA \(dauphine.fr\)](#)
- 11) [*Les processus AR et MA \(universite-paris-saclay.fr\)](#)

