

République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université de Mohamed El Bachir El Ibrahimi de Bordj Bou Arréridj
Faculté des Mathématiques et d'Informatique
Département d'informatique



MEMOIRE

Présenté en vue de l'obtention du diplôme

Master en informatique

Spécialité : réseaux & multimédia

THEME

Prédiction du prix de pétrole en utilisant les méthodes
analytiques ARMA et ARIMA

Présenté par :

YAHIAOUI Tassnim Hala

HAMIMID Yakine

Soutenu publiquement le : 14/09/2023

Devant le jury composé de:

Président : Sonia BENABID

Examineur : Abd Raouf BOUZIANE

Encadreur : Ferhat FILLALI

2022/2023

Dédicace

Je dédie ce modeste travail à :

A mes parents « Aucun hommage ne pourrait être à la hauteur de l'amour dont ils ne cessent de me combler ;
Que Dieu leur procure bonne santé et longue vie ».

Ma sœur Leen et mes frères Mohamed, Anis, Amine pour leur aide et encouragements.

Mes tantes, oncles, cousins et cousines pour m'avoir accompagné.

Ma meilleur amie Maroua qui m'a toujours soutenu et encouragé ainsi que sa famille.

Tous mes amies pour leur compagnie.

Ma binôme Tassnim Hala et sa famille pour leur accueil chaleureux.

Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible, je vous dis merci.

Yakine

Dédicace

Je dédie ce modeste travail à :

Mes chers parents, qui m'ont encouragé à travailler dur
à persévérer et à rester inébranlable pour atteindre mes objectifs

Mon soutien et ma force, mon frère

Mes beaux papillons, mes sœurs

Toute ma mignonne famille

Mes amies proches

Ma binôme

Tous ceux que j'aime

Tous ceux que connais.

Tassnim Hala

Remerciement

Nous voudrions remercier Dieu pour toute l'énergie qu'il nous a donné
durant ces cinq années hamdoulillah.

Nous souhaitons exprimer notre reconnaissance envers le professeur
FILLALI FERHAT, notre directeur de recherche, pour ses précieuses
orientations tout au long de cette étude.

Après cela, nous exprimons notre profonde gratitude envers les
membres des jurys qui ont généreusement accepté d'évaluer notre
mémoire de fin d'études.

Nous exprimons notre gratitude envers le département
d'informatique, son responsable ainsi que l'ensemble de ceux qui y
contribuent.

Nous souhaitons exprimer notre reconnaissance envers nos familles et
nos amies qui nous ont soutenus tout au long de ce parcours
académique, leur amour, leurs encouragements et leur compréhension
ont été essentiels pour surmonter les défis et atteindre ces jalons
importants.

Pour conclure, nos remerciements s'adressent à toutes les personnes
qui, de près ou de loin, ont participé à la réalisation de ce modeste
travail.

Yakine & Tassnim Hala

Résumé

Il est incontestable que la capacité à prévoir l'évolution des prix, qu'il s'agisse de prix généraux ou plus spécifiquement des tarifs des produits essentiels tels que le pétrole, joue un rôle crucial dans la prévention des répercussions néfastes qui en découlent. C'est pourquoi les études menées dans ce domaine revêtent une importance considérable et se posent comme un domaine de recherche scientifique distinct. Cette importance découle du rôle essentiel que les estimations des prix du pétrole, ainsi que d'autres indicateurs économiques et sociaux, jouent dans l'élaboration et l'orientation des programmes et des politiques économiques requises pour relever les défis à venir.

Dans ce contexte, nous tenterons d'explorer la réalité de la relation entre l'évolution des prix du pétrole et la conduite de la politique budgétaire. De plus, nous nous efforcerons d'évaluer les changements futurs des prix du pétrole en nous appuyant sur l'analyse de leurs évolutions passées et actuelles. Pour ce faire, nous utilisons les modèles ARIMA et ARMA.

Mots-clés : prix du pétrole, prévision, Modèle ARIMA.

Abstract

It is indisputable that the ability to predict the evolution of prices, whether general prices or more specifically the prices of essential products such as oil, plays a crucial role in preventing the harmful repercussions that ensue. That is why the studies carried out in this field are of considerable importance and arise as a separate field of scientific research. This importance stems from the critical role that oil price estimates, along with other economic and social indicators, play in shaping and guiding the economic programs and policies required to meet the challenges ahead.

In this context, we will attempt to explore the reality of the relationship between the evolution of oil prices and the conduct of fiscal policy. In addition, we will endeavor to assess future changes in oil prices based on the analysis of their past and current developments. To do this, we use the ARIMA and ARMA models.

Keywords: oil price, forecasting, ARIMA Model.

ملخص

ولا جدال في أن القدرة على التنبؤ بتطور الأسعار، سواء الأسعار العامة أو بشكل أكثر تحديدا أسعار المنتجات الأساسية مثل النفط، تلعب دورا حاسما في منع التداعيات الضارة التي تترتب على ذلك. ولهذا السبب فإن الدراسات التي يتم إجراؤها في هذا المجال لها أهمية كبيرة وتنشأ كمجال منفصل للبحث العلمي. وتتبع هذه الأهمية من الدور الحاسم الذي تلعبه تقديرات أسعار النفط، إلى جانب المؤشرات الاقتصادية والاجتماعية الأخرى، في تشكيل وتوجيه البرامج والسياسات الاقتصادية اللازمة لمواجهة التحديات المقبلة. وفي هذا السياق سنحاول استكشاف واقع العلاقة بين تطور أسعار النفط وسير السياسة المالية. بالإضافة إلى ذلك، سنحاول تقييم التغيرات المستقبلية في أسعار النفط بناءً على تحليل تطوراتها الماضية والحالية. للقيام بذلك، نستخدم نماذج ARMA وARIMA .

الكلمات المفتاحية: أسعار النفط، التنبؤ، نماذج الانحدار الذاتي التكاملي المتوسط المتحرك (ARIMA).

Table des matières

Liste des abréviations	ix
Liste des figures	xi
Chapitre 01 : Introduction Générale.....	1
1.2. Objectifs	1
1.3. Méthodologie et résultats	1
1.4. Structure du rapport	2
Chapitre 02 : Revue de la littérature.....	3
2.1. Introduction.....	3
2.2. Vue d'ensemble des méthodes de prédiction des prix du pétrole	3
2.3. Les modèles ARIMA : principe de fonctionnement et utilisation.....	4
2.4. Les modèles basés sur deep learning : principe de fonctionnement et utilisation.....	6
2.4.1. Quelques domaines basés sur le deep learning.....	8
2.6. Conclusion	10
Chapitre 03 : Collecte des données.....	11
3.1. Introduction.....	11
3.2. Sources de données sur les prix du pétrole	11
3.3. Prétraitement des données	18
3.4. Conclusion	19
Chapitre 04 : Modèles ARIMA.....	20
4.1. Introduction.....	20
4.2. Explication du modèle ARIMA et de ses composants (ordre AR, ordre I, ordre MA)	20
4.3. Entraînement du modèle ARIMA sur les données de prix du pétrole	22
4.4. Evaluation des performances des modèles ARIMA.....	24
4.5. Conclusion	25
Chapitre 05 : Méthodes de prédiction basée sur l'intelligence artificielle.....	26
5.1. Introduction.....	26
5.2. Méthodes basées sur le Machine learning.....	26
5.2.1. Trois catégories de machine Learning	26
5.3. Méthodes basées sur le deep learning	29

5.4. Conclusion	34
Chapitre 06 : Résultat d'exécution	35
6.1. Introduction.....	35
6.2. Explication.....	35
6.3. Conclusion	39
Chapitre 07 : Conclusion générale	40
7.1. Contributions.....	40
7.2. Critique du travail :	40
7.3. Travaux futurs et perspectives :	41
Les références	42

Liste des abréviations

AIE : L'Agence internationale de l'énergie.

ARIMA : Moyenne mobile intégrée autorégressive.

ARM : Modaux autorégressifs.

ARMA : Moyenne mobile autorégressive.

DME : dossiers médicaux électroniques.

GPU : Processeur graphique.

IAO : ingénierie assistée par ordinateur.

MAPE: Erreur moyenne absolue en pourcentage.

OPEP : Organisation des Pays Exportateurs de Pétrole.

PIB : produit intérieur brut.

PNB : produit national brut.

WTI: Intermédiaire de l'ouest du Texas.

IA : Intelligence artificiel.

ANN : Réseaux neuronaux artificiels.

SVM : machines à vecteurs de support.

TD : différence temporelle.

DQN : Réseau Q profond.

SARSA : État-Action-Récompense-État-Action

CNN : Réseaux de neurones convolutifs.

MLP : perceptron multicouche.

RNN : Réseaux de Neurones Récurrents.

GRU : Unité Récurrente Fermée.

LSTM : Mémoire à Long Terme à Court Terme.

RBFN : Réseaux de fonction de base radiale.

GAN : Réseaux adversariaux génératifs.

RBM : Machines de Boltzmann restreintes.

Liste des figures

Figure1: Réseaux neurone artificiel.....	7
Figure2: Prix annuel moyen de pétrole brut de l'OPEP entre 2013 et 2022 (en dollars des Etats-Unis par baril).....	13
Figure3: Prix moyen annuel de pétrole brut Brent entre 2013 et 2023 (en dollars des Etats-Unis par baril).....	15
Figure4: Prix annuel moyen de pétrole brut de WTI entre 2013 et 2023 (en dollars des Etats-Unis par baril).....	16
Figure5: Prix en comptant du pétrole brut Dubaï (Fateh) entre 2013 et 2022 (en dollars des Etats-Unis par baril).....	17
Figure6 : La relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage en profondeur.....	30
Figure7: L'environnement de développement intégré « pycharm ».....	35
Figure8: La sélectionne du fichier « Excel ».....	36
Figure9: Affichage du fichier « Excel ».....	36
Figure10: Choix des paramètres.....	37
Figure11: Graphe initial.....	37
Figure12: Graphe initial et le graphe prédit avec le modèle ARIMA.....	38
Figure13: Graphe initial et le graphe prédit avec le modèle ARMA.....	38

Chapitre 01 : Introduction Générale

1.1. Contexte

Le pétrole est considéré comme l'une des ressources énergétiques les plus importantes sur la scène mondiale et régionale, car il est devenu l'un des piliers les plus importants de l'économie des pays consommateurs comme des pays producteurs. Il est également considéré comme une ressource économique stratégique.

Les marchés mondiaux du pétrole sont l'un des marchés les plus volatils, car les prix du pétrole ont connu de grands développements et des fluctuations brusques et soudaines, en raison de leur association avec de nombreux facteurs difficiles à prévoir.

Et la fluctuation des prix du pétrole a un impact sur les marchés financiers, diverses industries et même sur la vie quotidienne, et cela conduit à réfléchir aux moyens de connaître ses prix et ainsi créer et développer différents modèles pour les prédire.

1.2. Objectifs

Les objectifs de la prévision des prix du pétrole sont nombreux, notamment la gestion des risques, la planification budgétaire, l'analyse des bénéfices, la gestion de l'offre et de la demande, le plus important étant de prendre des décisions stratégiques et de fournir des prévisions ou des estimations futures.

Ces prévisions peuvent être utiles à de nombreuses compagnies pétrolières, investisseurs, gouvernements, etc., grâce à la planification de la production, à la gestion des risques liés aux fluctuations des prix, à l'établissement de budgets et de prévisions financières réalistes et elles sont également utilisées à des fins politiques.

1.3. Méthodologie et résultats

Il existe de nombreuses méthodes qui permettent et qui facilitent la prévision et la prédiction des prix du pétrole, parmi les méthodes nous mentionnons : le modèle du marché mondial du pétrole, les modèles basés sur les séries chronologiques, les modèles de régressions et les modèles économétriques. Le modèle le plus important parmi les modèles basés sur les séries chronologiques c'est le modèle ARIMA (AutoRegressive Integrated Moving Average).

Les résultats finaux de cette étude sont des graphes, le premier représente les prix du pétrole des années précédentes et l'autre représente les prix du pétrole qui à prévoir.

1.4. Structure du rapport

Dans ce travail, nous présenterons d'abord la liste d'abréviations, la liste d'images et la liste des tableaux, puis nous présenterons les chapitres dont chaque chapitre a son introduction et conclusion.

Dans le premier chapitre nous parlerons de l'introduction générale, tandis que dans le deuxième chapitre nous passerons en revue de la littérature nous aborderons chacun des éléments suivants : vue l'ensemble des méthodes de prédiction, les modèles ARIMA et les modèles basés sur le deep learning en générale.

Dans le troisième chapitre, dans lequel les données sont collectées, nous découvrirons les sources des données sur les prix du pétrole et le prétraitement des données. Nous expliquerons en détail les composants du modèle ARIMA, comment entraîner le modèle ARIMA sur les données des prix du pétrole et évaluer ses performances dans le quatrième chapitre.

Dans le cinquième chapitre, nous commencerons par une introduction à l'algorithme ARM (Auto Regressive Modals) et comment il fonctionne, nous discuterons ensuite de la façon de construire le modèle ARM pour prédire les prix du pétrole, et enfin former le modèle ARM sur des données statistiques temporelles. Le sixième et dernier chapitre est une conclusion générale. Enfin, n'oubliez pas les références.

Chapitre 02 : Revue de la littérature

2.1. Introduction

La revue de la littérature explore en profondeur les méthodes de prédiction des prix du pétrole. Il met en évidence à la fois les méthodes traditionnelles comme les modèles ARIMA, et les approches novatrices basées sur l'apprentissage profond. De plus, cette étude se nourrit des recherches antérieures pour éclairer l'importance et les conséquences des différentes méthodologies dans le contexte actuel de la volatilité des marchés pétroliers.

2.2. Vue d'ensemble des méthodes de prédiction des prix du pétrole

La fluctuation des prix du pétrole a conduit à l'utilisation de différentes méthodes pour l'estimer et prédire ses prix réels. Voici quelques-unes des méthodes de prévision des prix du pétrole :

➤ **Modèle du marché mondial du pétrole :**

La première des nouvelles approches de projection élaborées se base sur l'utilisation d'un modèle du marché mondial du pétrole brut qui intègre les principaux facteurs de variation des prix du pétrole, tels que définis par la théorie économique. Plus spécifiquement, ce modèle exprime le prix actuel réel du pétrole en fonction de ses prix passés et des niveaux précédents de production pétrolière mondiale, qui servent d'indicateur de l'activité économique réelle, reflétant ainsi les fluctuations du cycle économique mondial et les variations des stocks mondiaux de pétrole brut [10].

➤ **Les modèles basés sur des séries chronologiques :**

Les modèles basés sur des séries chronologiques constituent l'approche la plus élémentaire de la prévision. Elle implique l'examen des tendances passées afin d'anticiper les événements futurs. Ces modèles varient en fonction des méthodes statistiques et de régression employées. L'ARIMA (Autoregressive Integrated Moving Average) est notamment l'un des modèles les plus reconnus pour prédire les consommations énergétiques à partir de séries chronologiques. Cependant, pour que cette analyse soit valide, il est nécessaire de disposer de données sur plusieurs

décennies et portant sur les mêmes variables explicatives. Par conséquent, cette approche peut ne pas être applicable universellement, notamment dans certaines nations en développement [5].

➤ **Les modèles de régression :**

Les modèles de régression évaluent la relation entre la consommation d'énergie et diverses variables en effectuant des régressions sur des ensembles de données. Un exemple illustratif de cette approche est l'application d'un modèle dans l'État du Kerala, en Inde, pour simuler la demande en énergie commerciale. Cette analyse est réalisée par secteur sur des séries de données précises, à savoir les secteurs résidentiels, des services et industriel. Chaque secteur est représenté par une équation distincte contenant des variables spécifiques, notamment le produit intérieur brut (PIB) et ses composantes sectorielles, le revenu par habitant, l'indice de production industrielle, le nombre de consommateurs, l'indice des prix du pétrole, ainsi que la consommation annuelle d'énergie dans l'État, à la fois globalement et par secteur [5].

➤ **Les modèles économétriques :**

Les modèles économétriques reposent sur l'idée que des relations existent entre la consommation d'énergie et des variables macro-économiques telles que le produit national brut (PNB), les prix de l'énergie, les technologies employées, la population, la production industrielle et agricole, le commerce extérieur, et d'autres facteurs similaires. Les variables pertinentes varient en fonction du pays, de la région ou du secteur étudié. Il a été démontré que ce type de modèle est assez efficace pour anticiper la demande globale dans les économies émergentes [5].

2.3. Les modèles ARIMA : principe de fonctionnement et utilisation

ARIMA est l'acronyme de "AutoRegressive Integrated Moving Average" qui peut être traduit en français par "Moyenne Mobile Autorégressive Intégrée". Le modèle combine trois composantes (p, d, q) ; dont « p » représente l'autorégression (AR), « d » représente l'intégration (I) et « q » représente la moyenne mobile (MA). L'autorégression concerne la dépendance des valeurs d'une série temporelle par rapport à ses valeurs passées. Il peut être utilisé pour appliquer

des fonctions prédictives à des cas isolés, c'est-à-dire pour prédire la valeur d'une observation individuelle sans dépendre d'autres séries de données [33] [7].

Il s'agit d'un modèle qui prédit les valeurs futures d'une série temporelle sur certains aspects de la structure statistique de la série observée. Autrement dit, c'est un modèle statistique, utilisé pour faire des prédictions sur les valeurs futures d'une série temporelle. Le modèle analyse la série temporelle observée pour identifier certains motifs, tendances ou caractéristiques statistiques qui peuvent influencer les valeurs futures. En se basant sur ces informations, le modèle est capable de faire des prédictions pour les valeurs qui viendront ensuite dans la série temporelle. Cela permet aux utilisateurs de mieux comprendre l'évolution du phénomène étudié et de prendre des décisions éclairées en anticipant les futurs comportements de la série temporelle [26].

L'utilisation principale d'ARIMA est de suivre et d'anticiper l'évolution d'un phénomène au fil du temps. Étant donné que les séries temporelles présentent souvent des tendances, des variations saisonnières ou des motifs spécifiques, ARIMA permet de modéliser ces structures et d'effectuer des prévisions sur la base des informations passées [33].

ARIMA "est une notion mathématique que l'on retrouve souvent dans le domaine de l'économétrie et dans l'analyse statistique" c'est-à-dire il étudie les relations économiques à l'aide de modèles statistiques. Il est également fréquemment utilisé dans l'analyse statistique pour analyser et prévoir des séries temporelles dans divers domaines tels que l'économie, la finance, la météorologie, les ventes, etc [33].

ARIMA est un modèle statistique et économétrique conçu pour analyser des événements de séries temporelles qui se déroulent sur une période spécifique. Il est utilisé pour comprendre les données passées et prévoir les tendances futures au sein d'une série. ARIMA est particulièrement adapté aux mesures enregistrées à des intervalles réguliers, allant des fractions de seconde aux périodes quotidiennes, hebdomadaires ou mensuelles [34].

L'utilisation d'ARIMA peut être bénéfique dans différents domaines tels que l'économie, les finances, la météorologie, la démographie, etc. En étudiant les données passées d'une série temporelle, ARIMA permet de mieux comprendre son comportement et de détecter des modèles significatifs. Grâce à ses capacités de prévision, il peut également aider à anticiper les évolutions futures et à prendre des décisions éclairées. C'est un outil puissant pour les analystes et les

chercheurs qui cherchent à exploiter les informations contenues dans des données temporelles pour des prévisions et des prises de décisions précises [34].

L'ARIMA, est un modèle qui prédit une série chronologique en se basant sur ses propres valeurs passées. Il est adapté à toute série numérique non saisonnière présentant des schémas plutôt que des événements aléatoires. Par exemple, les données de ventes d'un magasin constituent une série chronologique car elles sont collectées sur une période donnée. Une caractéristique essentielle de l'ARIMA est que les données sont recueillies à des intervalles réguliers et constants. Une version modifiée peut également être utilisée pour modéliser des prévisions sur plusieurs saisons [34].

Le modèle ARIMA gagne en popularité parmi les data scientists car il leur permet de prédire la demande future dans divers domaines tels que les ventes, la production ou même les prix des actions. En d'autres termes, ARIMA cherche à identifier les motifs et les tendances cachées dans une série chronologique plutôt que de simplement suivre les valeurs numériques brutes. Cela permet aux data scientists d'obtenir des prévisions plus précises et utiles concernant la demande future ou les fluctuations des prix. Grâce à cette approche, les entreprises peuvent prendre des décisions éclairées concernant leurs stocks, leurs niveaux de production ou leurs investissements dans le marché. L'ARIMA s'avère ainsi être un outil puissant dans le domaine de l'analyse des séries chronologiques et de la prévision [34].

2.4. Les modèles basés sur deep learning : principe de fonctionnement et utilisation

Le deep learning ou L'apprentissage en profondeur, représente un sous-domaine de l'intelligence artificielle axé sur la conception de vastes modèles de réseaux neuronaux capables de prendre des décisions précises en s'appuyant sur des données. Ce domaine trouve son application privilégiée dans des contextes où les données sont complexes et où de vastes ensembles de données sont disponibles. Actuellement, la plupart des entreprises en ligne et des technologies haut de gamme destinées au grand public font usage de l'apprentissage en profondeur [9].

Le deep learning repose sur des couches de réseaux neuronaux, des algorithmes inspirés du fonctionnement du cerveau humain. Grâce à un entraînement intensif avec de vastes ensembles de données, les neurones du réseau neuronal sont ajustés. Le résultat est un modèle de deep learning qui peut ensuite traiter de nouvelles données de manière autonome. Ces modèles collectent des informations provenant de multiples sources de données et les analysent en temps réel, sans intervention humaine. Dans le cadre du deep learning, les processeurs graphiques (GPU) sont optimisés pour la formation des modèles, car ils permettent de traiter de multiples calculs simultanément [35].

Le deep learning repose sur un réseau de neurones artificiels conçu pour imiter le fonctionnement du cerveau humain. Ce réseau est composé de plusieurs couches interconnectées comme ci-dessous :

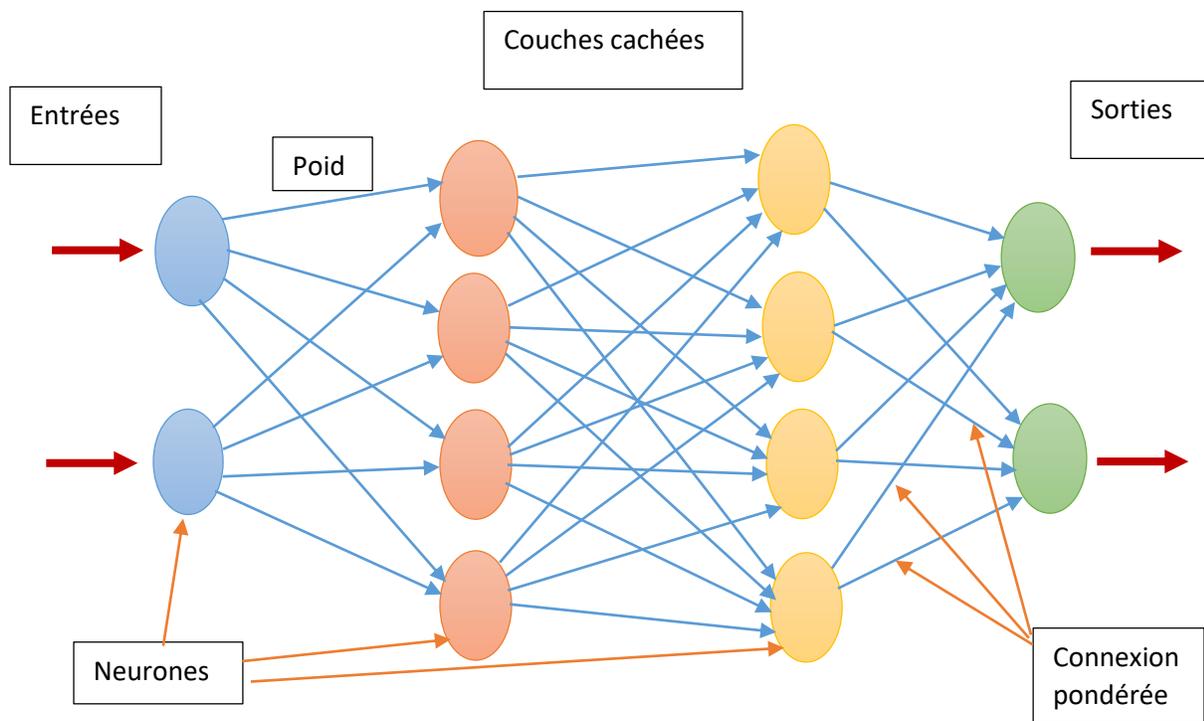


Figure 1 : Réseaux neurone artificiel

Couche d'entrée :

Les neurones de cette couche reçoivent les valeurs d'entrée du réseau et les transmettent aux neurones des couches cachées. Chaque neurone reçoit une valeur, il n'effectue donc pas de sommation [14].

Couches cachées :

Chaque neurone de cette couche reçoit l'information provenant de plusieurs couches précédentes. Il effectue ensuite une sommation pondérée par les poids des connexions et applique sa fonction d'activation pour transformer cette somme. Enfin, il envoie la réponse obtenue aux neurones de la couche suivante [14].

Couche de sortie :

Elle remplit un rôle similaire aux couches cachées, mais la différence est que les sorties des neurones de la couche de sortie ne sont liées à aucun autre neurone du réseau [14].

2.4.1. Quelques domaines basés sur le deep learning

❖ La santé :

La modernisation de l'industrie médicale a entraîné la production massive de données de santé. L'analyse rapide et précise de ces données offre de multiples avantages pour améliorer les résultats des soins. Les algorithmes de deep learning jouent un rôle crucial dans différents domaines de la santé, tels que la recherche médicale, l'analyse d'imagerie, la prévention des maladies, le développement de médicaments guidé et le traitement du langage naturel. Ce dernier aspect s'avère particulièrement utile pour transcrire efficacement des notes cliniques en texte libre dans les dossiers médicaux électroniques (DME). Grâce à ces avancées technologiques, l'application du deep learning dans le secteur de la santé ouvre de nouvelles perspectives pour améliorer la qualité des soins et favoriser des approches médicales plus personnalisées et efficaces [31].

❖ **La fabrication industrielle :**

Dans le secteur de la fabrication industrielle, les fabricants sont confrontés à la nécessité constante de fournir des produits et des services de meilleure qualité, plus rapidement et à moindre coût. Pour atteindre ces objectifs, de nombreuses entreprises adoptent l'ingénierie assistée par ordinateur (IAO), ce qui leur permet de réduire le temps, les coûts et les matériaux nécessaires au développement de prototypes physiques pour tester de nouveaux produits [31].

L'intégration du deep learning dans ce processus offre des avantages significatifs. En effet, le deep learning peut être utilisé pour modéliser des schémas extrêmement complexes dans des données multidimensionnelles, ce qui permet d'améliorer considérablement la précision analytique des données de test [31].

❖ **Conduite automatisée :**

En utilisant le deep learning, les fabricants peuvent donc optimiser leurs processus de développement, gagner en efficacité et en précision, et ainsi rester compétitifs sur le marché en proposant des produits de haute qualité tout en réduisant les coûts et les délais de production [19].

L'apprentissage en profondeur dans l'industrie automobile repose sur la détection automatique d'objets tels que les panneaux d'arrêt et les feux de circulation. De plus, l'apprentissage en profondeur est également utilisé pour détecter les piétons, ce qui contribue à réduire les accidents de la circulation [19].

Les avancées de l'intelligence artificielle ont conduit à des progrès remarquables dans le domaine des véhicules autonomes, allant des voitures sans conducteur aux drones de livraison. C'est-à-dire Les progrès de l'intelligence artificielle (en particulier l'apprentissage en profondeur) ont ouvert la voie à des avancées significatives dans le développement des véhicules autonomes. Ces véhicules autonomes, également appelés voitures sans conducteur, sont des véhicules capables de se déplacer et de prendre des décisions de manière autonome, sans intervention humaine directe [8].

2.6. Conclusion

Ce chapitre propose une vue d'ensemble des approches de prédiction des prix du pétrole. Nous avons commencé par définir les modèles ARIMA, puis exploré les mécanismes des modèles basés sur le deep learning. En outre, nous avons récapitulé des études passées utilisant ces méthodologies pour prédire les prix du pétrole.

Dans le chapitre suivant, nous aborderons la démarche de collecte des données.

Chapitre 03 : Collecte des données

3.1. Introduction

La collecte des données est l'élément central de notre étude, assurant la disponibilité de données pertinentes et de haute qualité. Cette étape explore en détail les multiples sources de données liées aux prix du pétrole, mettant en évidence l'importance primordiale du prétraitement. En effet, le prétraitement des données joue un rôle essentiel pour obtenir des résultats fiables et pertinents dans nos analyses et nos modèles de prédiction.

3.2. Sources de données sur les prix du pétrole

➤ Organisation des Pays Exportateurs de Pétrole (OPEP) :

Plusieurs facteurs différents se conjuguent et conduisent finalement à la nécessité de former un bloc de pays exportateurs de pétrole, une organisation qui coordonne ses politiques et unifie sa voix dans une position de négociation contre les monopoles mondiaux qui ont dominé l'industrie pétrolière, et porte le nom " Organisation des Pays Exportateurs de Pétrole " dite « OPEP » [20].

Cette organisation intergouvernementale a été créée par cinq pays (l'Arabie saoudite, le Koweït, l'Irak, l'Iran et le Venezuela) après la fin de la conférence de Bagdad le 14 septembre 1960 pour défendre les intérêts des pays producteurs et affronter le conglomérat des grandes entreprises internationales. Ainsi, l'OPEP est devenue la plus importante organisation mise en place par les pays en développement pour veiller à leurs intérêts et préserver le pétrole en tant que ressource épuisée et une source essentielle de revenus pour ces pays [20] [29].

L'OPEP comprend 13 pays membres organisés dans le tableau ci-dessous :

Continent	pay	Année d'adhésion
Afrique	l'Algérie	depuis 1969
Afrique	l'Angola	depuis 2007
Afrique	le Congo	depuis 2018

Afrique	le Gabon	depuis 2016 (il a quitté l'organisation en 1995)
Afrique	la Guinée équatoriale	depuis 2017
Afrique	la Libye	depuis 1962
Afrique	le Nigéria	depuis 1971
Amérique du Sud	le Venezuela	depuis 1960
Moyen-Orient	l'Arabie saoudite	depuis 1960
Moyen-Orient	les Émirats arabes unis	1967
Moyen-Orient	l'Iran	1960
Moyen-Orient	l'Irak	1960
Moyen-Orient	le Koweït	1960

Le tableau représente les pays membres de l'OPEP, Année d'adhésion et

L'un des motifs les plus importants qui ont motivé la création de l'OPEP est le désir des pays producteurs et exportateurs de pétrole de parvenir à un juste équilibre des pouvoirs monopolistiques dans la production de pétrole. Le premier objectif de l'OPEP est l'orientation à la hausse du cours du pétrole et la maîtrise de son cycle de production, et il y a d'autres objectifs comme : travailler avec diligence et efficacité pour développer des politiques de prix visant à assurer la stabilité des prix du pétrole sur les marchés mondiaux, afin d'éviter les fluctuations négatives des revenus pétroliers. L'organisation vise également à obtenir un rendement équitable pour les personnes qui ont investi leur argent dans l'industrie pétrolière [20] [29].

Les membres de l'OPEP produisent plusieurs pétroles, et leurs prix moyens, ainsi que le prix de pétrole de l'OPEP, sont affichés dans une norme appelée le panier. Ce panier est considéré comme l'une des références les plus importantes pour les prix du pétrole brut dans le monde.

[25] Les principales références sont :

- ✓ Le Brent.
- ✓ West Texas Intermediate (WTI).

✓ Le Dubai Crude (Fateh).

- **Statistiques des Prix moyen du pétrole brut de l'OPEP :**



Figure 2 : prix annuel moyen du pétrole brut de l'OPEP entre 2013 et 2022

(En dollars des États-Unis par baril) [25]

- **Analyse les statistiques de l'OPEP :**

Dans ces statistiques, le prix de brut de l'OPEP a augmenté à 105.87 \$ et 102.97 \$ en 2013 et 2022, respectivement, et a diminué à 40.76 \$ et 41.47 \$ en 2016 et 2020, respectivement, et atteint 69.78 en 2018.

➤ **L'Agence internationale de l'énergie (AIE) :**

L'Agence internationale de l'énergie est une organisation intergouvernementale indépendante associée à l'organisation de coopération et de développement économiques, qui travaille à produire une énergie propre et sûre. Cette agence a été créée en 1974 après le premier choc pétrolier et avait pour mission de coordonner les actions à entreprendre en cas de rupture d'approvisionnement en pétrole, ainsi que d'assurer la sécurité énergétique [28] [30].

Les prix d'importation du pétrole brut sont influencés par plusieurs facteurs, dont les fluctuations courantes de l'offre et de la demande sur le marché mondial du pétrole. L'Agence internationale de l'énergie (AIE) tient un registre appelé "Crude Oil Import Register" dans lequel elle enregistre les données relatives aux importations de pétrole brut réalisées par différents pays [24].

En plus des fluctuations habituelles de l'offre et de la demande, d'autres facteurs influencent également les prix du pétrole brut, notamment les facteurs géopolitiques. Les événements politiques, les conflits internationaux, les sanctions économiques et les tensions géopolitiques peuvent perturber l'approvisionnement en pétrole, entraînant ainsi des fluctuations significatives des prix [24].

L'AIE collecté ces informations auprès des organismes nationaux, qui fournissent des données sur les importations de pétrole brut en fonction de différents critères, tels que le type de brut l'origine géographique (le pays exportateur) et la qualité du pétrole brut. Cette collecte de données permet à l'AIE et aux acteurs du marché de mieux comprendre les tendances de l'importation de pétrole brut et de prendre des décisions éclairées concernant les échanges et les investissements dans le secteur pétrolier [24].

➤ **Brent :**

Le pétrole Brent est un élément essentiel de l'économie pétrolière mondiale, car il affecte les prix de l'énergie, les matières premières et les économies nationales et mondiales. Il peut être exporté et a une bonne qualité qui le rend approprié pour une utilisation dans un large éventail d'industries. Les marchés du Brent ont évolué au fil du temps pour inclure les marchés au comptant et à terme, et le pétrole Brent est devenu un type important de pétrole brut en Europe, et l'une des références pour les prix mondiaux du pétrole [23].

- **Statistiques des Prix moyen du pétrole brut de Brent :**

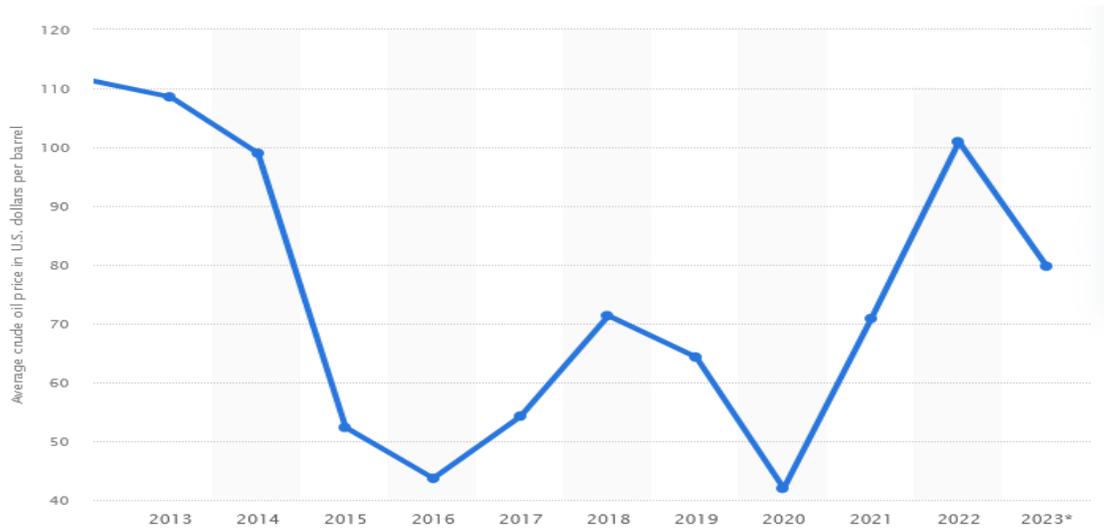


Figure 3: Prix moyen annuel du pétrole brut Brent entre 2013 et 2023

(En dollars des États-Unis par baril) [36]

- **Analyse les statistiques de Brent :**

Dans ces statistiques, on note l'instabilité des prix du Pétrole brut de Brent de 2013, où il a atteint 108.56 \$, jusqu'en 2023, où il a atteint 79.75 \$. Le prix du Pétrole brut de Brent a baissé en 2016 et 2020, atteignant respectivement 43.67 \$ et 41.96 \$.

➤ **West Texas Intermediate (WTI) :**

- **Statistiques des Prix moyen du pétrole brut de WTI :**

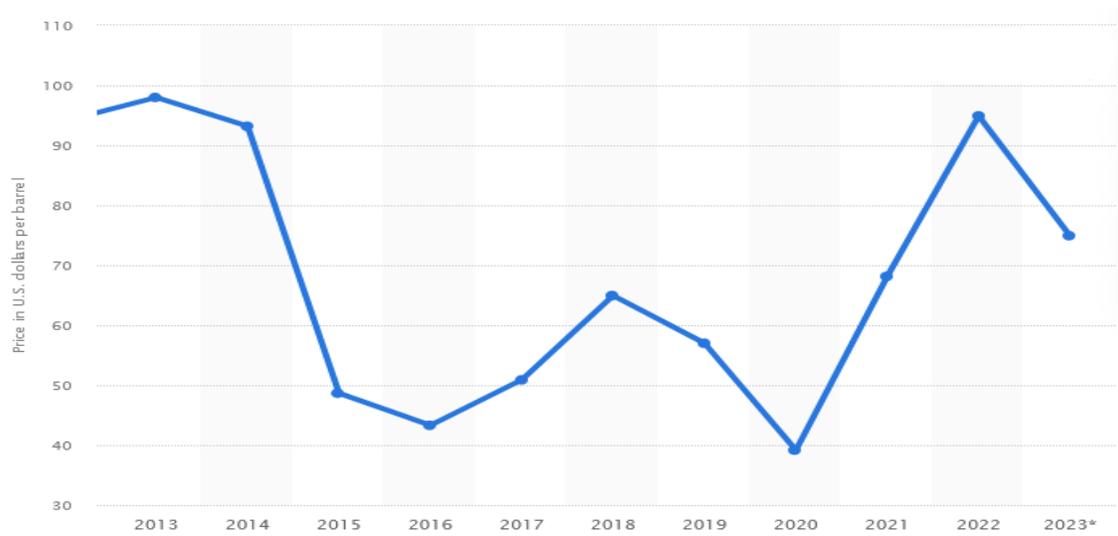


Figure 4: Prix annuel moyen du pétrole brut (WTI) entre 2013 et 2023

(En dollars des États-Unis par baril) [37]

- **Analyse des statistiques de West Texas Intermediate (WTI) :**

Dans ces statistiques, on note la fluctuation des prix du Pétrole brut de Brent de 2013, où il a atteint 97.98 \$, jusqu'en 2023, où il a atteint 74.92 \$. Le prix du Pétrole brut de Brent a chuté en 2016 et 2020, atteignant respectivement 43.29 \$ et 39.16 \$.

➤ **Dubaï (Fateh) :**

➤ **Statistiques des Prix moyen du pétrole brut Dubaï (Fateh) :**

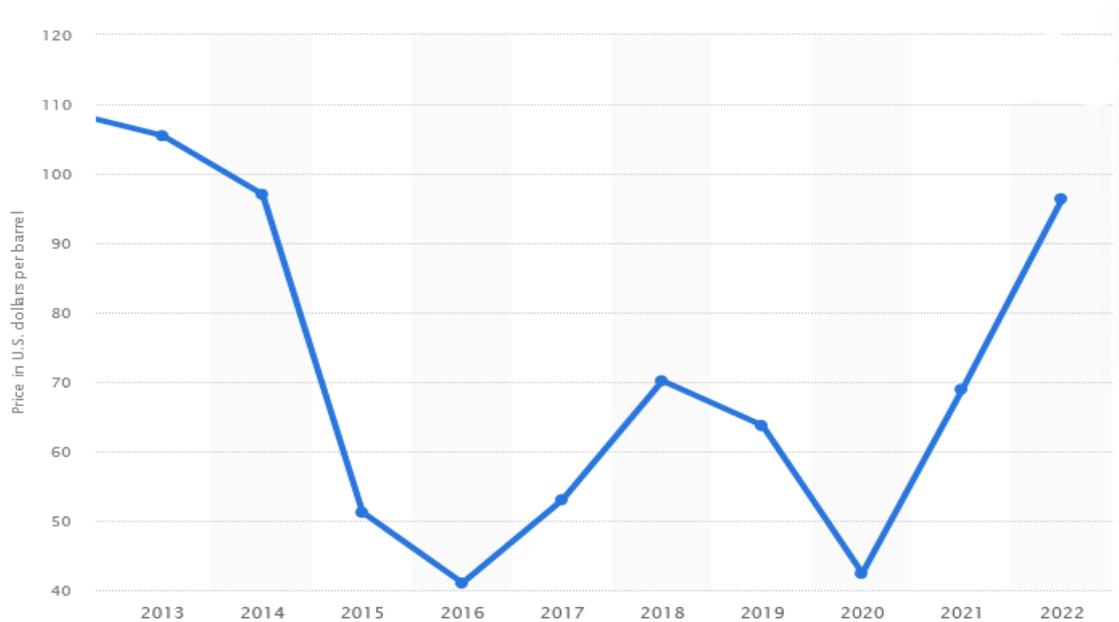


Figure 5: Prix au comptant du pétrole brut Dubaï (Fateh) entre 2013 et 2022

(En dollars des États-Unis par baril) [39]

• **Analyse des statistiques de Dubai (Fateh) :**

On observe une baisse du prix du brut de 105.47 en 2013 à 41.02 en 2016, suivie d'une hausse à 70.15 en 2018, puis il a de nouveau diminué en 2020 à 42.41 il atteint 96.38 en 2022.

➤ **Les pays du Conseil de coopération du Golfe :**

Les pays du Conseil de coopération du Golfe dépendent en grande partie du pétrole dans le domaine économique, car Bahreïn, le Qatar, le Koweït, les Emirats arabes unis, le Sultanat d'Oman et le Royaume d'Arabie saoudite possèdent la moitié des réserves mondiales prouvées ou plus des matières premières les plus importantes, puisque les coûts de production de pétrole dans la région

sont les plus bas au monde, où le cout est de trois et quatre dollars le baril dans chacun des royaume d'Arabie saoudite et à Oman, respectivement, et ce conduit à une augmentation du profit dans l'économie dans le domaine du pétrole en particulier [22].

L'agence internationale de l'énergie prévoit une augmentation de la production des pays exportateurs de pétrole entre 2005 et 2030 3% par an, de sorte que leur part dans la production mondiale augmentera de 1.4%. Il indique une augmentation de la production des pays exportateurs de pétrole du Moyen-Orient de 28.8% à 54.1% entre 2000 et 2030. On estime que les exportations de pétrole en Moyen-Orient et en Afrique du Nord augmenteront pour atteindre 18.1 millions de barils par jour en 2020 à partir de la production mondiale, qui s'élevait à 43.63barils par jour.

3.3. Prétraitement des données

Le prétraitement des données représente une étape fondamentale au sein du processus de fouille de données. Cette étape englobe la collecte, le tri et la conversion des données initiales dans un format exploitable pour l'analyse ultérieure. L'objectif de cette phase est de garantir que les données soient de qualité supérieure, exhaustive et apte à être soumises à l'analyse [40].

➤ Collecte de données :

L'amorce du prétraitement des données se fait par la collecte des données, marquant ainsi sa phase inaugurale. Cette étape consiste à réunir des données non raffinées provenant de diverses sources, telles que des fichiers, des bases de données, des sites web, etc. Ces données peuvent adopter diverses modalités, incluant des textes, des images, ainsi que des fichiers audio et vidéo [40].

➤ Nettoyage des données :

Par la suite dans le processus de prétraitement des données, intervient la phase de purification des données. Cette étape implique l'élimination des valeurs absentes, des doublons, des informations incompatibles et des données erronées, garantissant ainsi la précision et la fiabilité des données [40].

➤ **Transformation des données :**

La métamorphose des données englobe le processus de convertir les données originales en un format exploitable pour l'analyse, englobant potentiellement des actions telles que normaliser les données, réduire leur dimensionnalité, ou encore discrétiser les valeurs. L'objectif global est de façonner un ensemble de données qui se révèle plus accessible et propice à l'analyse [40].

➤ **Intégration des données :**

L'étape d'intégration des données consiste à fusionner les données recueillies auprès de sources diverses en un ensemble unique. Ce processus englobe fréquemment la résolution des divergences entre les données [40].

➤ **Réduction des données :**

La phase de diminution des données consiste à restreindre le volume des informations en éliminant les variables sans pertinence ou celles étroitement liées. Cette démarche vise à simplifier la complexité des données et à augmenter l'efficacité de l'analyse accomplie [40].

Le prétraitement des données revêt une signification capitale au sein du processus de fouille de données, assurant que les données soient de qualité supérieure, exhaustives et prêtes à être soumises à l'analyse. Cette procédure englobe la collecte, le dépouillement, la métamorphose, la fusion et la réduction des données [40].

3.4. Conclusion

En conclusion de ce chapitre, la collecte des données se révèle être le pilier fondamental de notre étude, garantissant l'accès à des données pertinentes et de qualité supérieure. Nous avons minutieusement exploré les diverses sources de données liées aux prix du pétrole, soulignant ainsi l'importance cruciale du prétraitement.

Dans le prochain chapitre nous ferons connaissance avec les modèle ARIMA.

Chapitre 04 : Modèles ARIMA

4.1. Introduction

Ce chapitre nous offrira une compréhension approfondie du modèle ARIMA, de ses composants essentiels, de son application aux données de prix du pétrole, ainsi que de sa performance en tant qu'outil de prévision.

4.2. Explication du modèle ARIMA et de ses composants (ordre AR, ordre I, ordre MA)

Le modèle ARIMA est appelé modèles Box-Jenkins car l'approche ARIMA a été popularisée par Box et Jenkins [6].

Le modèle ARIMA fait des prévisions pour une variable en s'appuyant sur sa relation linéaire avec les valeurs passées. Le modèle ARIMA est appliqué lorsque la série de données est cohérente dans le temps (stationnaire). Cependant, dans les scénarios réels, les données des séries chronologiques ne sont souvent pas cohérentes (non stationnaires). Pour appliquer des modèles stationnaires, il est crucial d'éliminer les fluctuations causées par des facteurs non stationnaires dans la série chronologique. Une solution pour relever ce défi a été introduite par Box et Jenkins à travers le modèle ARIMA, qui convertit efficacement les données non stationnaires en données stationnaires en introduisant un processus de différenciation. Cette approche surmonte avec succès les limites [1].

Dans les modèles ARIMA, la première étape consiste à supprimer la non-stationnarité par différenciation. Ce processus consiste à soustraire l'observation actuelle de la précédente. Par exemple, une différenciation du premier ordre peut être obtenue en remplaçant la valeur d'origine Y_t par:

Le modèle ARIMA(p, d, q) se caractérise par sa forme générale, qui peut s'exprimer comme suit:

$$Y'_t = Y_t - Y_{t-1}$$

$$Y'_t = C + \sum_{i=1}^p \phi_i y'_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

Ici, Y'_t représente la série nouvellement différenciée, ε représente le bruit blanc, et $\{\phi = 1, 2, \dots, p\}$ et $\{\theta = 1, 2, \dots, q\}$ correspondent aux coefficients des composantes autorégressives (AR(p)) et moyenne mobile (MA(q)), respectivement [1].

Le modèle ARIMA est une généralisation d'un modèle de moyenne mobile autorégressive (ARMA), dans un modèle ARMA, la valeur attendue de Y_t à un instant donné est représentée en fonction des observations précédentes, $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-p}$, et les innovations passées, $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$. Le degré AR, noté p , signifie dans quelle mesure y s'appuie sur des observations antérieures, tandis que le degré MA représente la dépendance de y à l'égard des innovations passées [6].

Le modèle ARIMA passe par différentes étapes : premièrement, les données de la série temporelle sont transformées pour obtenir une structure de covariance cohérente ; Ensuite, le modèle est identifié, approché, diagnostiqué et utilisé à des fins de prédiction [18].

Le « I » dans ARIMA signifie que l'ensemble de données est différencié pour atteindre la stationnarité et qu'après modélisation, les résultats sont intégrés pour générer des prédictions et des estimations finales. Le modèle ARIMA est représenté par : ARIMA (p, d, q), où [18] :

- ✓ "p" est l'ordre du modèle autorégressif (AR).
- ✓ "d" représente le nombre d'opérations de différenciation appliquées pour obtenir la stationnarité.
- ✓ "q" désigne l'ordre du modèle de moyenne mobile (MA).

Lors de la différenciation du modèle ARIMA par rapport au temps, il devient un modèle ARMA (p, q) stationnaire. Le fondement d'ARMA réside dans le modèle autorégressif (AR) avec

un ordre de "p", le cadre de moyenne mobile (MA) avec un ordre de "q", et les expressions mathématiques pour MA, AR et ARMA sont utilisées pour construire le modèle ARIMA et sont [18]:

✓ Le modèle AR:

$$\varphi'_t = \theta_1 \varphi_{t-1} + \theta_2 \varphi_{t-2} + \dots + \theta_p \varphi_{t-p} + \varepsilon_t = \sum_{i=1}^p \theta_i \varphi_{t-i} + \varepsilon_t$$

✓ Le modèle MA :

$$\varphi'_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots + \phi_q \varepsilon_{t-q} = \sum_{i=1}^q \phi_i \varepsilon_{t-i}$$

✓ Le modèle ARMA:

$$\varphi'_t = \sum_{i=1}^p \theta_i \varphi_{t-i} + \varepsilon_t + \sum_{i=1}^q \phi_i \varepsilon_{t-i}$$

Où θ_t est le paramètre d'autorégression au temps t, ε_t est le terme d'erreur au temps t et ϕ_t est le paramètre de moyenne mobile au temps t [18].

4.3. Entraînement du modèle ARIMA sur les données de prix du pétrole

La méthode de Box et Jenkins constitue une approche scientifique appliquée à la modélisation de séries temporelles. Pour mettre en œuvre cette méthode, les étapes suivantes peuvent être suivies, avec possibilité de revoir ces étapes si les résultats ne sont pas satisfaisants :

1) La familiarisation avec les données :

La phase de familiarisation avec les données implique l'exploration approfondie du contexte d'application, des théories existantes, des objectifs poursuivis, ainsi que de la qualité des données en termes de précision et d'exactitude. Cela comprend également une compréhension de la périodicité naturelle du phénomène étudié, la constance temporelle, et l'identification d'événements qui auraient pu influencer la série. Une partie intégrante de cette étape consiste en une analyse visuelle des données, ayant pour but de détecter les variations structurelles dans la série, d'identifier les erreurs flagrantes, et d'appréhender les conséquences d'éventuelles interventions extérieures, telles que des modifications législatives ou économiques, des incidents majeurs, ou des grèves [13].

2) L'analyse préliminaire :

La phase préliminaire d'analyse débute par des choix décisionnels : cela peut englober la décision de négliger les données initiales dans la série, la rectification des valeurs aberrantes, le comblement des lacunes de données, la transformation des données, voire le recours à une nouvelle variable. Par la suite, l'effort est dirigé vers la réalisation de la stationnarité de la série, souvent appuyée par des visualisations graphiques [13].

3) Spécification du modèle :

La phase de spécification du modèle, également appelée identification, repose sur l'observation des structures d'autocorrélation et d'autocorrélation partielle, ce qui guide vers un ou plusieurs modèles ARMA. Cependant, en général, il est nécessaire d'effectuer plusieurs cycles de la phase (3) à (5). Dans ce processus, les caractéristiques les plus saillantes sont prises en compte. Les paramètres du modèle sont ensuite estimés, et la procédure recommence avec l'analyse des résidus obtenus [13].

4) Estimation des paramètres :

Les paramètres se réfèrent aux coefficients des termes AR (Auto-Regressive) et MA (Moving Average) dans les polynômes associés [13].

5) L'étude de l'adéquation du modèle :

La phase d'évaluation de la pertinence du modèle, aussi appelée validation, implique de déterminer si l'optimisation non linéaire a réussi et si le modèle est adéquat. En cas d'invalidité du modèle, il est nécessaire de réexaminer l'analyse à partir de l'une des étapes précédentes. Idéalement, cela devrait être fait en utilisant les informations acquises [13].

6) La prévision :

Définit le modèle pour suivre directement les prévisions. De plus, les variances des erreurs de prévision pour l'horizons 1, 2, ... sont également obtenues. En supposant que la distribution soit normale, il devient possible de définir la répartition des valeurs à venir ainsi que les plages de prévision [13].

7) L'interprétation des résultats :

L'interprétation des résultats peut parfois être complexe, mais elle n'est pas toujours indispensable [13].

4.4. Evaluation des performances des modèles ARIMA

Une fois qu'un modèle a généré ses prédictions, il est judicieux d'évaluer sa qualité pour garantir la fiabilité de ces prédictions. Cette section présente les métriques de performance les plus fréquemment employées dans la littérature académique [16].

La performance de la prévision est évaluée en examinant les différences entre les valeurs prédites par le modèle et les valeurs réelles sur toute la durée de la période de test. Les principales métriques utilisées à cet effet incluent l'erreur moyenne absolue, l'erreur quadratique moyenne, la racine carrée de l'erreur quadratique moyenne et la moyenne des écarts absolus [11].

Soit $\{y_{1:t+h}^{(i)}\}_{i \in 1:N}$ un ensemble de N séries temporelles est disponible, et la mission consiste à générer des prévisions pour ces séries $y'_{t+1:t+h}$ Pour les prochaines h périodes temporelles après t , pour chaque l'une des N séries temporelles comprises dans l'ensemble de données analysé [16].

L'identification du modèle de prévision optimal est une étape cruciale car elle garantit l'exactitude des prédictions. Il est essentiel d'évaluer la qualité des prédictions, ce qui revient essentiellement à déterminer leur exactitude. Une prévision précise implique de comparer les

valeurs prévues avec les valeurs réelles. Diverses méthodes et approches ont été explorées par les chercheurs pour évaluer les performances des prévisions. Dans cette étude, nous utilisons le pourcentage d'erreur absolu moyen (MAPE) pour évaluer l'efficacité de notre modèle de prévision [12].

MAPE signifie « Mean Absolute Percentage Error », et il s'agit d'une mesure largement utilisée pour comparer (pour évaluer) et mesurer les performances de prévision, ainsi :

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{\text{Actuel} - \text{Prévision}}{\text{Actuel}} \right|$$

Par conséquent, le choix de la méthode de mesure de la performance la plus appropriée repose sur les préférences et l'expertise de l'analyste. L'efficacité de chaque outil de mesure peut être améliorée grâce à une application appropriée et peut entraîner des inexactitudes en cas de calcul incorrect. Par conséquent, la sélection de l'instrument de mesure peut être adaptée aux exigences spécifiques de l'analyse [12].

4.5. Conclusion

Dans ce chapitre, nous avons introduit en détail le modèle ARIMA. Notre exploration a débuté par une explication approfondie de ses composants essentiels : l'ordre AR, l'ordre I et l'ordre MA. Ensuite, nous avons mis en lumière le processus d'entraînement du modèle ARIMA. Enfin, nous avons procédé à une évaluation rigoureuse des performances du modèle ARIMA.

La phase suivante qui est les modèles basés sur le deep Learning qui sera détaillé dans le chapitre suivant.

Chapitre 05 : Méthodes de prédiction basée sur l'intelligence artificielle

5.1. Introduction

Ce chapitre, nous découvrirons les principes fondamentaux, les applications concrètes, dans et les avantages respectifs des méthodes basées sur le machine Learning et celles basées sur le deep Learning. Nous serons ainsi mieux équipés pour comprendre comment l'IA est utilisée pour la prédiction dans divers domaines, de la finance à la santé, en passant par la technologie et bien d'autres.

5.2. Méthodes basées sur le Machine learning

La machine Learning représente une méthode de programmation informatique qui tire parti de probabilités statistiques pour conférer aux ordinateurs la capacité d'apprendre de manière autonome, sans nécessiter une programmation explicite. Son objectif fondamental est de doter les ordinateurs de la capacité "d'apprendre à apprendre", et par conséquent, de s'adapter et de réagir de manière similaire aux êtres humains. Cette adaptation implique l'amélioration continue de leurs compétences d'apprentissage et de leurs connaissances au fil du temps. L'ultime aspiration de la machine Learning est de parvenir à ce que les ordinateurs puissent agir et réagir sans nécessiter une programmation explicite pour chaque action ou réaction spécifique. Cette approche repose sur des programmes évolutifs qui ajustent leurs performances à chaque exposition à des données d'entrée variées [42].

5.2.1. Trois catégories de machine Learning

Le machine Learning n'est pas une technologie nouvelle, mais a des racines qui remontent loin dans le temps. En effet, le tout premier réseau neuronal artificiel, connu sous le nom de "Perceptron", fut inventé en 1958 par le psychologue américain Frank Rosenblatt. À l'origine, le Perceptron était envisagé comme une machine, plutôt qu'un simple algorithme. En 1960, il a été intégré dans le développement de la "Mark 1 Perceptron", une machine de reconnaissance d'images pionnière. La Mark 1 Perceptron fut le premier ordinateur à utiliser des réseaux neuronaux artificiels (ANN) pour simuler la réflexion humaine et apprendre par tâtonnements [42].

Avec l'avènement des bibliothèques et des Framework open source, ainsi que l'incroyable augmentation de la puissance de calcul des ordinateurs, multipliée par des milliards entre 1956 et 2021, la machine Learning s'est largement répandue. Aujourd'hui, le machine Learning est omniprésent, que ce soit dans les transactions boursières, la protection contre les logiciels malveillants, ou encore la personnalisation du marketing. Quelle que soit sa complexité, la machine Learning peut être regroupée en trois principales catégories :

➤ **Machine Learning supervisé :**

Le machine Learning supervisé est une approche fondamentale mais rigoureuse de la technologie. Dans ce cadre, les opérateurs fournissent à l'ordinateur des exemples d'entrées et leurs résultats attendus, incitant ainsi l'ordinateur à rechercher des solutions qui permettent d'obtenir ces résultats en fonction des entrées données. L'objectif ultime est que l'ordinateur acquière une compréhension générale des règles qui associent les entrées aux résultats [42].

La machine Learning supervisé trouve une application importante dans la capacité à effectuer des prédictions sur des données qui ne sont pas encore disponibles ou sur des données futures, ce que l'on appelle souvent la "modélisation prédictive". L'algorithme s'efforce de construire une fonction capable de prédire de manière précise les résultats en se basant sur les variables d'entrée. Par exemple, il peut être utilisé pour prédire la valeur d'un bien immobilier (résultat) en utilisant des données d'entrée telles que le nombre de pièces, l'année de construction, la superficie du terrain, l'emplacement, et ainsi de suite [42].

La machine Learning supervisé se divise généralement en deux catégories principales :

1. Classification : Lorsque la variable de sortie consiste en des catégories discrètes.

2. Régression : Lorsque la variable de sortie est une valeur numérique spécifique.

Parmi les principaux algorithmes utilisés en machine Learning supervisé, on retrouve les forêts aléatoires, les arbres de décision, l'algorithme des k plus proches voisins (k-Nearest Neighbors), la régression linéaire, l'algorithme de Naïve Bayes, les machines à vecteurs de support (SVM), la régression logistique et les méthodes de boosting de gradient [42].

➤ **Machine Learning non supervisé :**

Dans le domaine de la machine Learning non supervisé, l'algorithme est chargé de découvrir la structure inhérente aux données sans qu'aucune étiquette ne lui soit fournie. Cette approche peut être poursuivie pour elle-même, dans le but de révéler des structures sous-jacentes dans les données, ou elle peut être utilisée comme un moyen d'atteindre un objectif spécifique. On qualifie souvent cette démarche d'"apprentissage des caractéristiques" (feature Learning).

Un exemple concret d'application de la machine Learning non supervisé est l'algorithme de reconnaissance faciale prédictive de Facebook, qui identifie automatiquement les personnes présentes sur les photos publiées par les utilisateurs, sans avoir été explicitement enseigné pour reconnaître chaque individu.

La machine Learning non supervisé se divise généralement en deux principales catégories :

1. Cluster ING : L'objectif consiste à détecter des regroupements naturels au sein des données, où des éléments similaires sont regroupés ensemble.

2. Association : L'objectif est d'identifier des règles ou des relations permettant de définir de larges ensembles de données, mettant en lumière des associations entre différents éléments.

Parmi les principaux algorithmes couramment utilisés en machine Learning non supervisé, on peut citer le K-Means, le cluster ING hiérarchique, ainsi que les techniques de réduction de la dimensionnalité [42].

➤ **Machine Learning par renforcement :**

Dans le domaine de la machine Learning par renforcement, un logiciel interagit avec un environnement en constante évolution dans le but de réaliser une tâche spécifique, comme la conduite d'un véhicule ou la compétition dans un jeu. Le programme en phase d'apprentissage reçoit un retour d'information sous forme de "récompenses" et de "sanctions" pendant qu'il évolue dans l'espace des problèmes, apprenant ainsi à identifier les comportements les plus efficaces dans le contexte donné [42].

En 2013, un algorithme de machine Learning par renforcement, le E-learning, a déjà suscité l'attention en devenant célèbre pour sa capacité à apprendre à gagner dans six jeux vidéo Atari, sans intervention d'un programmeur.

La machine Learning par renforcement se divise généralement en deux principales catégories :

1. Méthode de Monte Carlo : Le programme reçoit des récompenses à la fin de chaque "état terminal".

2. Méthode de différence temporelle (TD) : Les récompenses sont évaluées et attribuées à chaque étape du processus.

Parmi les principaux algorithmes couramment utilisés en machine Learning par renforcement, on peut citer le E-learning, le Deep Q Network (DQN), ainsi que le SARSA (State-Action- Reward-State-Action) [42].

5.3. Méthodes basées sur le deep learning

➤ L'apprentissage profond (Deep learning) :

L'apprentissage profond est une nouvelle branche de la recherche en apprentissage automatique, il fait lui-même partie de l'intelligence artificielle. Il a été créé dans le but de rapprocher l'apprentissage automatique de son objectif fondamental : l'émulation de l'intelligence artificielle. Il se concentre sur les algorithmes inspirés par la structure et le fonctionnement du cerveau, leur permettant d'apprendre plusieurs niveaux de représentation afin de modéliser des relations complexes entre les données [2].

Le Deep Learning se focalise sur l'utilisation de multiples couches interconnectées pour convertir les entrées en caractéristiques qui peuvent ensuite être utilisées pour anticiper les sorties correspondantes. En utilisant un ensemble de données assez vaste comprenant des paires d'entrées et de sorties, un algorithme d'apprentissage peut être employé pour automatiquement apprendre comment faire correspondre les entrées aux sorties en ajustant un ensemble de paramètres au sein de chaque couche du réseau [17].

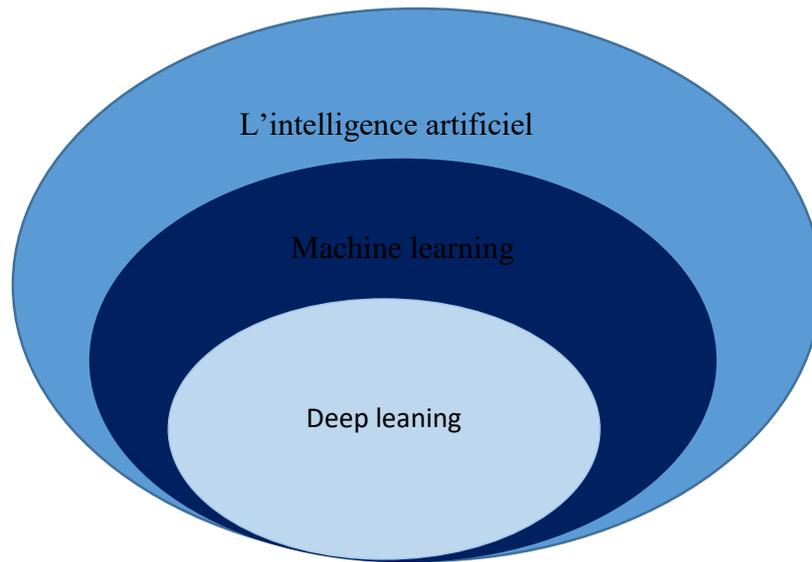


Figure 6: La relation entre l'intelligence artificielle, l'apprentissage automatique et L'apprentissage en profondeur [2]

Il existe plusieurs méthodes de prédiction basées sur l'apprentissage profond, nous aborderons ci-dessous quelques approches et techniques qui permettent aux machines d'apprendre des modèles complexes à partir des données et de les utiliser pour faire des prédictions significatives.

✓ **Réseaux de neurones convolutifs (CNN) :**

Les réseaux de neurones convolutifs (CNN) sont une architecture spécialement conçue pour traiter des données qui ont une structure spatiale régulière, comme les jetons d'une séquence ou les pixels d'une image. Dans une couche CNN, cette régularité est exploitée en appliquant un ensemble de filtres convolutifs locaux à travers différentes positions des données. Cette approche présente deux avantages majeurs : elle permet de réduire le risque de surapprentissage en ayant un nombre restreint de poids à ajuster par rapport à une couche d'entrée classique, tout en conservant la dimensionnalité pour les couches suivantes, et elle assure l'invariance par rapport aux translations. Typiquement, un module CNN est composé de plusieurs couches CNN successives, ce qui permet

aux nœuds des couches plus avancées de disposer de champs de réception plus larges pour capturer des caractéristiques plus complexes [17].

Les réseaux de neurones convolutifs sont de puissants modèles de programmation qui permettent la reconnaissance automatique d'images en attribuant une étiquette à chaque image fournie en entrée qui correspond à sa classe. Il s'agit d'un type spécifique de réseau neuronal basé sur le processus de convolution. Les réseaux convolutifs proviennent d'une architecture connue sous le nom de perceptron multicouche (MLP), mais ils utilisent des poids conjoint associés à la fenêtre convolutive, leur permettant d'extraire implicitement des caractéristiques locales. Les CNN sont particulièrement adaptés aux tâches de reconnaissance d'images [2].

✓ **Réseaux de Neurones Récurrents (RNN) :**

Les Réseaux de Neurones Récurrents (RNN) ont été développés dans le but d'extraire des caractéristiques globales à partir de données séquentielles. Lorsqu'ils traitent une séquence d'entrée, un module RNN fait usage d'un vecteur d'état interne pour synthétiser les informations provenant des éléments successifs de la séquence. Ce module intègre un sous-ensemble paramétré qui prend en compte le vecteur d'état interne précédent ainsi que l'élément d'entrée courant de la séquence pour générer le vecteur d'état interne actuel. Finalement, le vecteur d'état final résume l'intégralité de la séquence d'entrée.

Puisque la même opération est répétée sur chaque élément de la séquence, les modules RNN sont souvent sujets à des problèmes de disparition ou d'explosion du gradient lors de l'entraînement avec l'algorithme de rétropropagation. Pour résoudre ces problèmes, des modules de réseau neuronal récurrents fermés ont été développés, tels que la Mémoire à Long Terme (LSTM) et l'Unité Récurrente Fermée (GRU). Ces modules sont spécifiquement conçus pour atténuer ces problèmes et permettre un meilleur apprentissage sur des données séquentielles [17].

✓ Réseaux de Mémoire à Long Terme à Court Terme (LSTM) :

Les Réseaux de Mémoire à Long Terme à Court Terme, couramment désignés sous l'acronyme LSTM, représentent une catégorie spéciale de Réseaux Neuronaux Récurrents (RNN). Ils ont été proposés par Hochreiter et Schmidhuber en 1997. Contrairement aux RNN classiques,

qui, en théorie, ont la capacité d'apprendre des règles de mise à jour de séquences de manière arbitraire, ces modèles ont tendance à oublier rapidement les informations du passé dans la pratique. Ce phénomène est connu sous le nom de "problème de la disparition du gradient", ce qui a conduit au développement des LSTM.

La cellule LSTM représente une adaptation de la couche récurrente traditionnelle, permettant ainsi aux signaux provenant d'étapes antérieures des couches profondes de persister et d'influencer la cellule à l'étape actuelle [2].

✓ Unités récurrentes fermées (GRU) :

Les cellules GRU (Gated Recurrent Unit) améliorent la capacité des réseaux récurrents à conserver davantage d'informations historiques pour des prédictions plus précises. Elles incorporent un mécanisme de mise à jour (update gate) qui détermine quelle quantité d'informations du passé doit être conservée, ainsi qu'une porte de réinitialisation (reset gate) qui contrôle la quantité d'informations à oublier [2].

Le GRU représente une solution très efficace pour atténuer le problème de la disparition du gradient dans les réseaux de neurones récurrents. La disparition du gradient est un défi courant en apprentissage automatique, se produisant lorsque le gradient devient excessivement petit, ce qui entrave la mise à jour des poids du réseau. De plus, les GRU présentent de meilleures performances que les LSTM sur des ensembles de données de petite taille [2].

✓ Réseaux de fonction de base radiale (RBFN):

Ces méthodes sont des types spécifiques de réseaux neuronaux à propagation avant. Ils tirent parti de fonctions d'activation radiales comme élément clé. Leur architecture comprend typiquement une couche d'entrée, une couche cachée, et une couche de sortie. Habituellement, les RBFN sont déployés pour des tâches telles que la classification, la prédiction de séries temporelles et la régression linéaire [41].

✓ Réseaux adversariaux génératifs (GAN):

Les GAN créent de nouvelles instances de données qui s'apparentent aux données d'apprentissage profond. Ils possèdent deux principaux composants : un générateur et un discriminateur. Si le générateur apprend à produire des informations erronées, le discriminateur, quant à lui, apprend à exploiter ces fausses informations [41].

✓ Machines de Boltzmann restreintes (RBM) :

Cet algorithme a été élaboré par le professeur Geoffrey Hinton. En d'autres termes, les Restricted Boltzmann Machines (RBM) sont des réseaux neuronaux stochastiques qui se composent de deux couches : les unités visibles et les unités cachées. Ces réseaux artificiels ont la capacité d'apprendre à partir d'une distribution de probabilité sur un ensemble d'entrées [41].

Nous allons maintenant découvrir la relation entre l'apprentissage profond, l'apprentissage automatique et l'intelligence artificielle :

Le domaine de l'intelligence artificielle est une branche de la science qui cherche des méthodes pour résoudre des problèmes de grande complexité, que ce soit d'un point de vue logique ou algorithmique. Dans le cadre de l'intelligence artificielle, on trouve le champ de l'apprentissage automatique, qui vise à développer des techniques pour que les machines puissent apprendre à partir des données. L'apprentissage profond, également connu sous le nom de deep learning, deep structured learning, ou apprentissage hiérarchique, regroupe un ensemble de méthodes dans le

domaine de l'apprentissage automatique. Ces méthodes tentent de modéliser des données de manière abstraite en utilisant des architectures composées de transformations non linéaires [2].

5.4. Conclusion

En conclusion de ce chapitre, nous avons exploré en détail les principes fondamentaux, les applications concrètes, ainsi que les avantages inhérents aux méthodes reposant sur le machine Learning et celles s'appuyant sur le deep Learning. Grâce à ces connaissances, vous êtes désormais mieux préparés à appréhender le rôle crucial de l'intelligence artificielle dans la prédiction, que ce soit dans des domaines aussi variés que la finance, la santé, la technologie, et bien d'autres.

Chapitre 06 : Résultat d'exécution

6.1. Introduction

Dans ce chapitre, nous allons explorer deux aspects cruciaux de cette démarche :

L'environnement de développement PyCharm et les modèles ARIMA (AutoRegressive Integrated Moving Average) et ARMA (AutoRegressive Moving Average). Notre objectif principal sera de démontrer de manière éclairante la prédominance d'ARIMA par rapport à ARMA dans des contextes spécifiques essentiels.

6.2. Explication

Dans ce projet nous utilisons le pycharm pour exécuté les codes de ARIMA et ARMA :

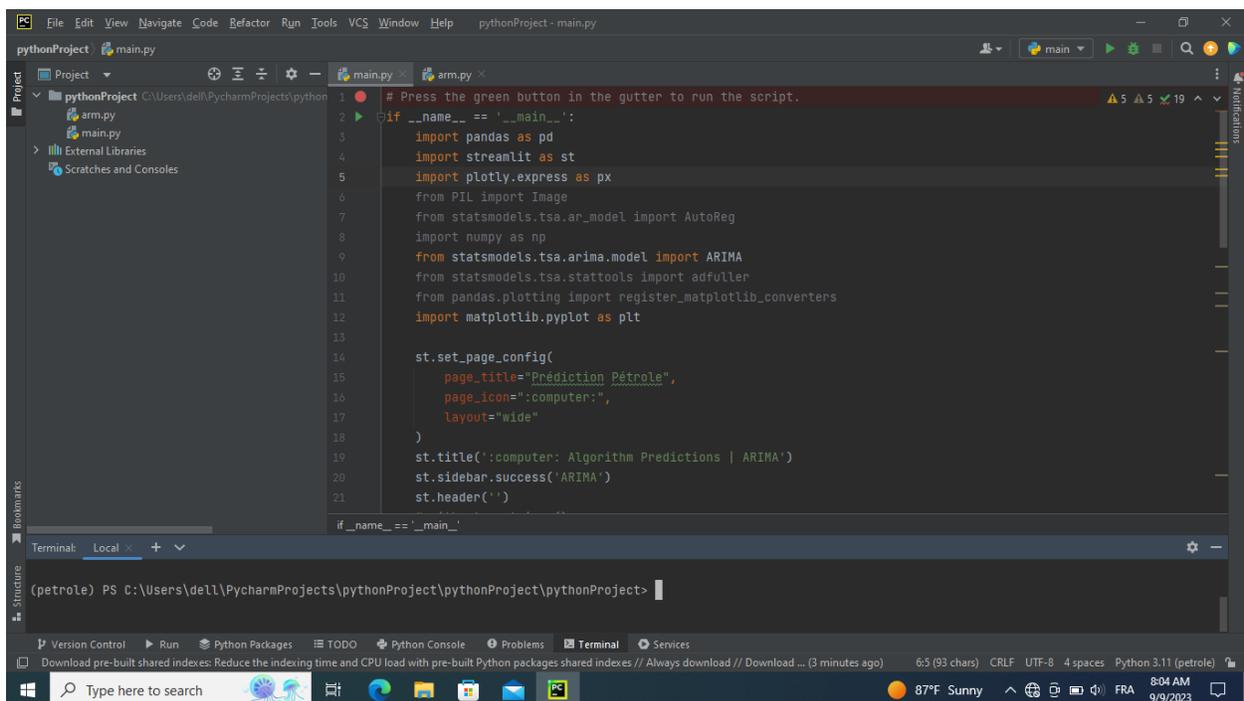


Figure7 : L'environnement de développement intégré « pycharm ».

Les résultats de l'exécution du modèles ARIMA et ARM sont afficher dans les figures ci-dessous:
Cliquez sur le « Browse files » pour sélectionner le fichier « Excel ».

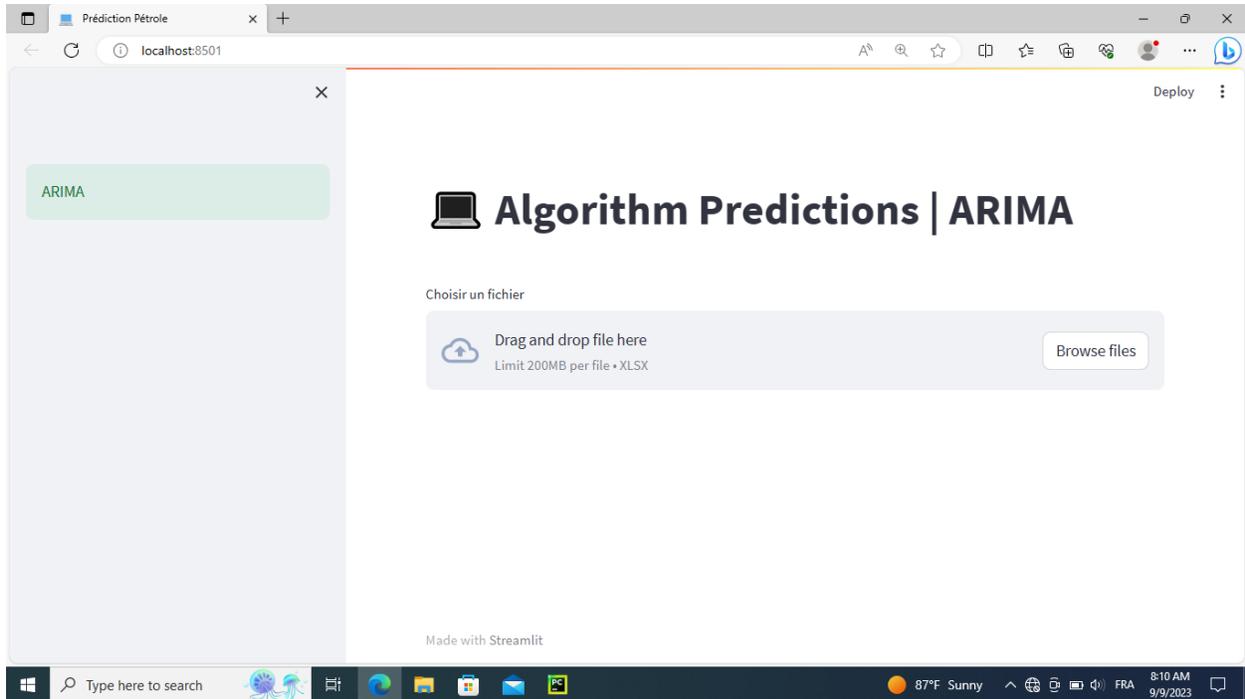


Figure 8 : La sélectionne du fichier « Excel »

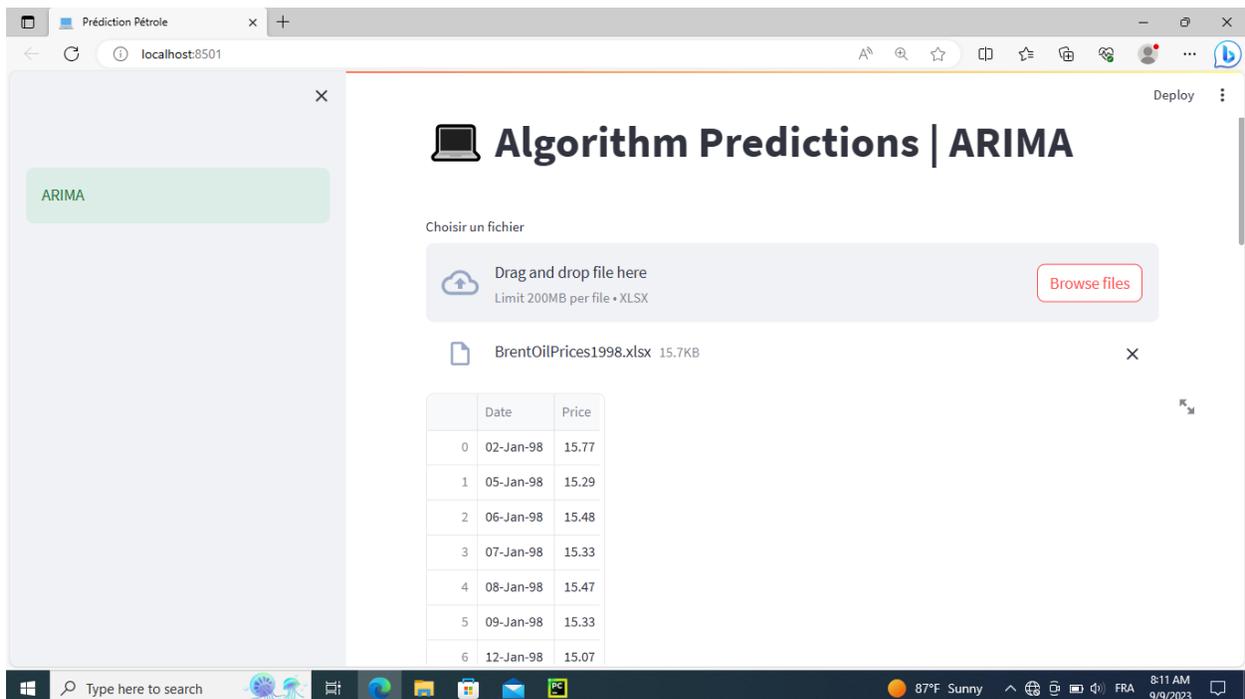


Figure 9 : Affichage du fichier « Excel »

Choisi les paramètres « Date » et « Price »

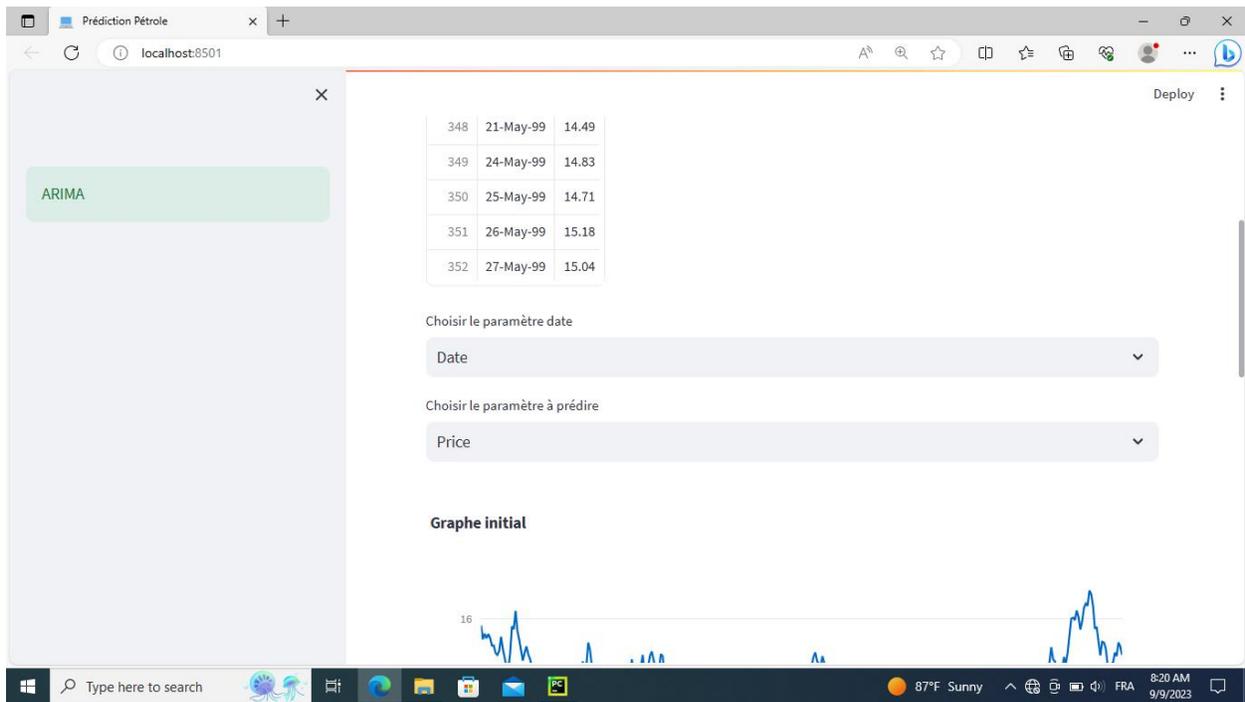


Figure 10 : Choix du paramètres

Voici le graphe initiale :

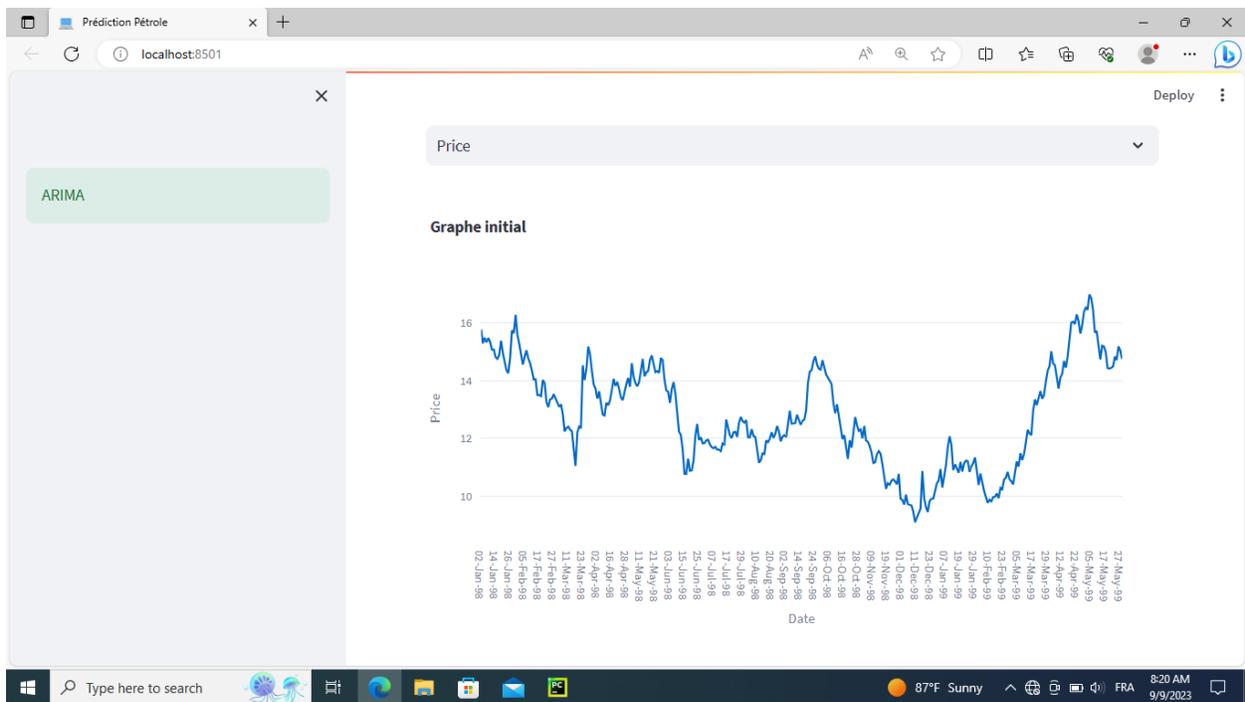


Figure 11 : Grappe initial

Voici la prédiction avec le modèle ARIMA :

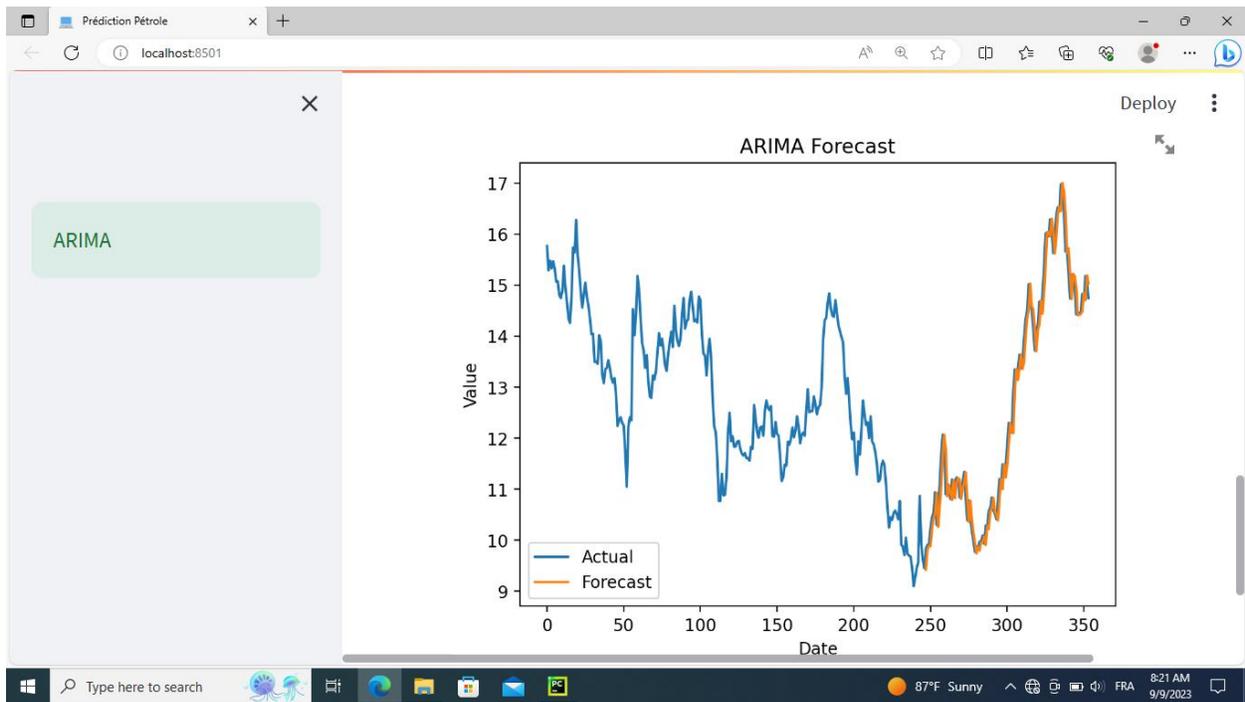


Figure 12 : Le graphe initial et le graphe prédit avec le modèle ARIMA

Voici la prédiction avec le modèle ARMA :

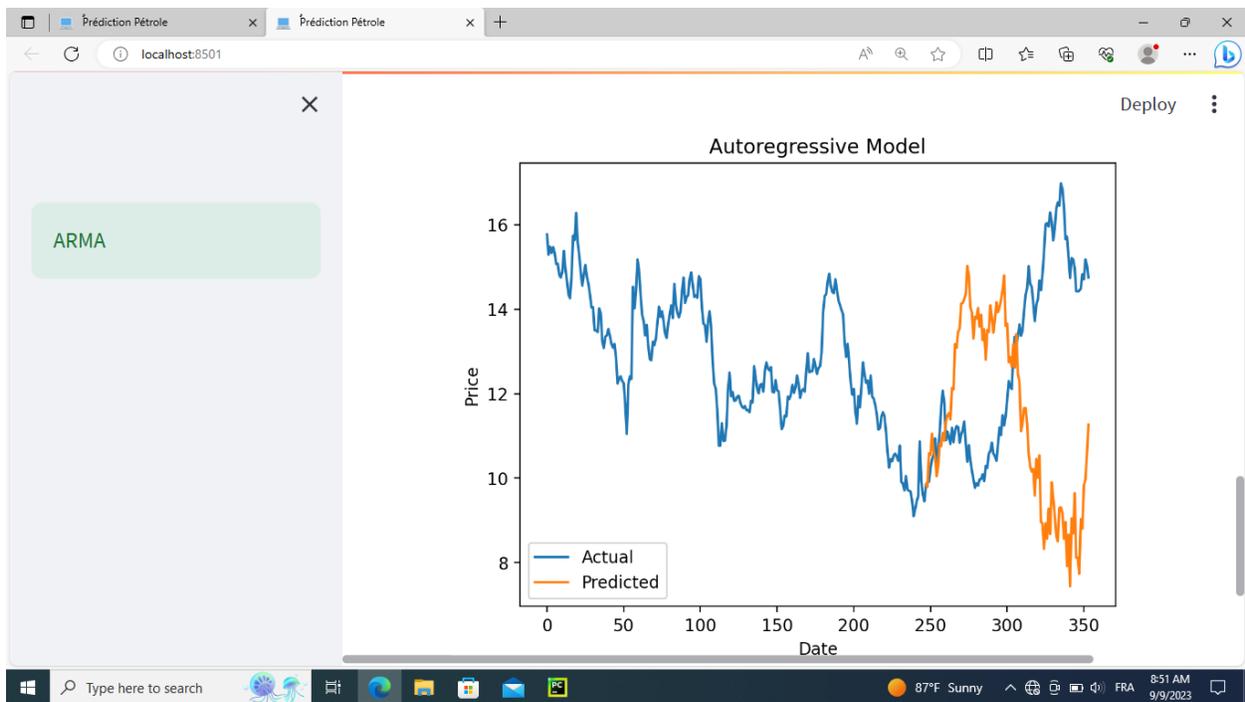


Figure 13 : Le graphe initial et le graphe prédit avec le modèle ARMA

6.3. Conclusion

A la fin de ce chapitre, nous avons examiné en détail l'usage fondamental de PyCharm en tant qu'IDE pour mettre en œuvre les modèles ARIMA et ARMA dans le cadre de l'analyse des séries temporelles. De plus, nous avons mis en avant la supériorité d'ARIMA par rapport à ARMA dans de nombreuses circonstances.

Chapitre 07 : Conclusion générale

7.1. Contributions

Notre projet, centré sur la comparaison des algorithmes de prédiction dans le domaine commercial avec une étude de cas sur les prix du pétrole, constitue une avancée majeure pour l'analyse prédictive et la prise de décision économique. Son objectif principal est d'évaluer et de confronter l'efficacité de diverses approches algorithmiques dans la prévision des fluctuations des prix du pétrole, une problématique cruciale pour les acteurs du secteur énergétique et financier.

Autrement dit, notre projet de comparaison des algorithmes de prédiction dans le contexte des prix du pétrole représente une avancée significative dans le domaine commercial. Il offre des outils et des connaissances précieuses pour améliorer la prise de décision et la gestion des risques dans le secteur énergétique, tout en stimulant de nouvelles pistes de recherche et d'innovation dans le domaine de l'analyse prédictive.

7.2. Critique du travail :

L'étude comparative des algorithmes de prédiction dans le domaine commercial, en se fondant sur l'exemple des prix du pétrole, comporte des limitations à prendre en considération. Tout d'abord, elle peut être entravée par l'imprévisibilité inhérente au marché pétrolier, où des événements géopolitiques majeurs tels que les conflits internationaux peuvent exercer une influence inattendue sur les prix. De plus, cette analyse repose fréquemment sur des données historiques qui pourraient ne pas pleinement refléter les dynamiques émergentes du marché. Enfin, certains modèles de prédiction peuvent s'avérer complexes à interpréter pour les acteurs commerciaux, ce qui peut altérer leur pertinence dans le processus décisionnel.

Néanmoins, en dépit de ces limites, l'étude présente des aspects positifs indéniables, elle contribue à la recherche et au développement de modèles de prédiction plus sophistiqués, capables d'apporter une certaine anticipation des fluctuations des prix du pétrole. Ces modèles fournissent aux entreprises une compréhension améliorée des tendances passées et actuelles du marché, facilitant ainsi des décisions plus éclairées en matière de gestion des risques, de planification stratégique et d'investissement. Par ailleurs, l'étude de cas sur les prix du pétrole peut servir de point de départ pour des travaux de recherche ultérieurs visant à renforcer la robustesse des

modèles de prédiction et à les appliquer à d'autres secteurs commerciaux. En définitive, malgré les défis persistants, cette étude constitue une étape cruciale vers une utilisation plus avancée des algorithmes de prédiction dans le domaine commercial.

7.3. Travaux futurs et perspectives :

Les conclusions tirées de notre étude comparative sur les algorithmes de prédiction dans le domaine commercial, en mettant en avant notre étude de cas sur les prix du pétrole, ouvrent la voie à diverses pistes d'extension et de perspectives prometteuses.

En résumé, notre travail actuel constitue une base solide pour des recherches et des développements ultérieurs dans le domaine de la comparaison des algorithmes de prédiction dans le contexte commercial. Les opportunités d'extension et les perspectives évoquées précédemment offrent des horizons stimulants pour la poursuite de l'amélioration de la précision des prévisions, pour une meilleure compréhension des marchés commerciaux, et pour un soutien continu à la prise de décision éclairée dans un environnement en constante évolution.

Les références

- [1] Alrweili, H., & Fawzy, H. (2022). Forecasting crude oil prices using an ARIMA-ANN hybrid model. *J Stat Appl Probab*, 11(3), 845-855.
- [2] Application des réseaux de neurones profonds pour la prédiction de la puissance produite par un micro-réseau, dédié au chargement des véhicules électriques, AYYOUB LEFZA, Octobre 2020.
- [3] Brownlee, J. (2017). *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery.
- [4] Cryer, Jonathan D. Kung-Sik Chan. Time Series Analysis With Applications in R(Vol. 286).p 66-68
- [5] Dabat, M. H., Litvine, D., & Mouras, S. Comprendre, évaluer et prédire la demande en énergie: enjeux pour l'accès à l'énergie dans les pays en développement.
- [6] Devi, B. U., Sundar, D., & Alli, P. (2013). An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50. *International Journal of Data Mining & Knowledge Management Process*, 3(1), 65.
- [7] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 1847979018808673.
- [8] Josée J.Intelligence artificielle pour les débutants.
- [9] Kelleher, J. D. (2019). *Deep learning*. MIT press. Page1.
- [10] L'ART ET LA SCIENCE DE LA PRÉVISION DU PRIX RÉEL DU PÉTROLE, REVUE DE LA BANQUE DU CANADA • PRINTEMPS 2014 - Christiane Baumeister.
- [11] Laurent, E. (2006). Optimiser la gestion de trésorerie par la modélisation économétrique des flux financiers: théorie et application à la trésorerie du régime général de la sécurité sociale en France. Editions Publibook.
- [12] Makala, D., & Li, Z. (2021, February). Prediction of gold price with ARIMA and SVM. In *Journal of Physics: Conference Series* (Vol. 1767, No. 1, p. 012022). IOP Publishing.
- [13] Mélard, G. (2006). Initiation à l'analyse des séries temporelles et à la prévision. *Revue Modulad*, 82(35).

- [14] Prédiction des paramètres pétrophysiques à l'aide des méthodes de l'intelligence artificielle, cas du réservoir TAGI champs de Chebet El-Nakhla bassin d'Oued M'Ya, sud-est de l'Algérie. SEKHER Khaled, BOUTAGHANE Ayoub. 2021.page 22.
- [15] SÉRIES CHRONOLOGIQUES, Adam Salvail, 19 janvier 2021.
- [16] SUR L'UTILISATION DE L'APPRENTISSAGE PROFOND POUR LA PRÉVISION PROBABILISTE EN
- [17] Torrisi, M., Pollastri, G., & Le, Q. (2020). Deep learning methods in protein structure prediction. *Computational and Structural Biotechnology Journal*, 18, 1301-1310.
- [18] Tularam, G. A., & Saeed, T. (2016). Oil-price forecasting based on various univariate time-series models. *American Journal of Operations Research*, 6(3), 226-235.
- [19] السيد اسماء د. محمد محمود كريمة / د. التعليم تكنولوجيا ومستقبل الاصطناعي الذكاء تطبيقات 84صفحة .محمد
- [20] Volume 6, Numéro 1, Pages 109-111 . 2022-06-30 المجلة العلوم الإدارية و المالية
- [21] الطاقة في مصر و تحديات المستقبل د. سيد عاشور احمد صفحة 16
- [22] الاقتصاد السياسي للنفط في الشرق الأوسط و شمال افريقيا و افاق سوق الطاقة الدولية. منذر ماخوس. ترجمة مونيك كامل. مراجعة ابتسام الخضرا. صفحة 31.
- [23] زهيرة & دراسة تنبؤية لأسعار النفط باستخدام نماذج ARCH-ARMA دراسة حالة برنت خلال الفترة 2019-1987 عمور(2019)
- [24] <https://data.oecd.org/fr/energy/prix-du-petrole-brut-a-l-importation.htm> ,
Juil,20.2023, 22 :15 h
- [25] <https://fr.statista.com/statistiques/564926/prix-annuel-du-petrole-de-l-opec-1960/>
, Aout,15.2023, 18:10h
- [26] <https://ledatascientist.com/arima/> , Juil,13.2023, 13:45h
- [27] <https://www.actuia.com/tutoriel/analyse-de-series-temporelles/> , Juil.28.2023,
6 :50h
- [28] <https://www.connaissancedesenergies.org/fiche-pedagogique/agence-internationale-de-lenergie-aie> , Juil,15.2023, 3 :30h
- [29] <https://www.connaissancedesenergies.org/fiche-pedagogique/opec-organisation-des-pays-exportateurs-de-petrole> , Aout,2.2023, 19:00h

- [30] <https://www.glossaire-international.com/pages/tous-les-termes/agence-internationale-de-l-energie-aie.html> , Juil,13.2023, 23 :20h
- [31] <https://www.hpe.com/fr/fr/what-is/deep-learning.html> ,Juil,30.2023,18 :18h
- [32] <https://www.investopedia.com/terms/a/autoregressive.asp> , Aout,5.2023,21:55h
- [33] <https://www.jedha.co/formation-ia/algorithmme-arma> , Juil,31.2023, 2 :30h
- [34] <https://www.mastersindatascience.org/learning/statistics-data-science/what-is-arma-modeling/> , Aout,7.2023, 13:55h
- [35] <https://www.oracle.com/dz/artificial-intelligence/machine-learning/what-is-deep-learning/> ,Juil,11.2023, 23:00h
- [36] <https://www.statista.com/statistics/262860/uk-brent-crude-oil-price-changes-since-1976/> , Aout,5.2023, 21:25h
- [37] <https://www.statista.com/statistics/266659/west-texas-intermediate-oil-prices/>,
Aout,5.2023, 22:20h
- [38] <https://www.statista.com/statistics/266660/price-development-of-dubai-fateh-oil/> ,
Aout,5.2023, 1:20h
- [39] <https://www.studocu.com/row/document/universite-djillali-liabes-de-sidi-bel-abbes/data-mining/le-pretraitement-des-donnees/48143738> , Juil,14.2023, 3 :00h
- [40] <https://www.talend.com/fr/resources/methode-modeles-data-mining/> ,
Aout,18.2023, 5:35h
- [41] <https://www.jedha.co/formation-ia/algorithmes-deep-learning>
- [42] <https://www.talend.com/fr/resources/what-is-machine-learning>