Mémoire de Master 2022-2023

**Comment peut-on utiliser le modèle LSTM, pour prédire les prix journaliers du Bitcoin en intégrant des données externes ?**

Paul, BERGEON

Marché financier et analyse quantitative

1701950

Mémoire encadré par : Miia Chabot

**RÉSUMÉ**

Ce mémoire explore l'application des réseaux de neurones LSTM pour prédire les prix journaliers du Bitcoin, en intégrant des données externes comme le sentiment de marché "Crypto Fear and Greed Index", le prix de clôture du Bitcoin, la valeur de l'Ethereum, le volume de transactions, les prix les plus hauts et les plus bas journaliers, le taux de variation et la volatilité du BTC. Le modèle LSTM a démontré une performance significative, soulignant son potentiel en prédiction financière. Cependant, des défis demeurent tels que la dépendance à la qualité et l'accessibilité des données et l'importance d'une interprétation correcte des résultats par les investisseurs. En conclusion, l'avenir du LSTM dans la prédiction des prix du Bitcoin et plus largement les crypto-monnaies sont prometteur, malgré les défis à relever pour une mise en œuvre efficace.

**Mots clés :** Bitcoin, LSTM, prédiction du prix, données externes, journalier.

**Attestation**

Je soussigné Paul BERGEON certifie être l’auteur exclusif/exclusive de ce Mémoire de Master.

Je certifie que ce mémoire n’a pas déjà ou ne fait pas l’objet d’une soumission dans le cadre d’un autre diplôme.

J’assure avoir pris connaissance et compris la politique de l’ESSCA en matière de plagiat. Je certifie que toutes les sources documentaires sur lesquelles s’appuie ce travail sont portées à la connaissance du lecteur de manière explicite et conformément aux règles définies dans le guide du mémoire de Master de l’ESSCA.

Je certifie que les données sur lesquelles s’appuie ce mémoire ont été collectées et analysées en respectant les règles de l’intégrité académique. Je me tiens à la disposition de l’ESSCA pour produire les données originales et les fichiers informatiques sur lesquels j’ai effectué l’analyse.

L’ESSCA n’entend donner ni approbation, ni improbation aux opinions émises dans ce Mémoire de Master.  
Ces opinions doivent être considérées comme propres à leur.

**Sommaire**

Table des matières

[I. Introduction 6](#_Toc135079434)

[1. Analyse fondamentale et technique 8](#_Toc135079435)

[2. L'évolution des modèles de machine learning et de deep learning dans la prédiction des prix des cryptos monnaies 10](#_Toc135079436)

[3. Intégration des données externes pour améliorer la prédiction des prix des cryptos monnaies 11](#_Toc135079437)

[4. Défis et opportunités pour l'amélioration des modèles de prédiction des prix des cryptos monnaies 13](#_Toc135079438)

[5. Définition du concept principal 15](#_Toc135079439)

[6. Identification d'un manque dans la littérature 16](#_Toc135079440)

[A. Les limites des modèles de machine learning actuels 17](#_Toc135079441)

[B. Les avantages potentiels des modèles de deep learning 17](#_Toc135079442)

[C. Modèles hybrides pour prédire les prix des cryptomonnaies 18](#_Toc135079443)

[D. Défis et perspectives de recherche 19](#_Toc135079444)

[E. Formulation de la problématique 20](#_Toc135079445)

[F. Présentation de la stratégie de recherche approfondie, annonce des contributions et du plan 21](#_Toc135079446)

[II. Revue de littérature 23](#_Toc135079447)

[III. Méthodologie 32](#_Toc135079448)

[1. Introduction 32](#_Toc135079449)

[A. Contexte de la recherche 32](#_Toc135079450)

[B. Objectifs de la recherche 32](#_Toc135079451)

[C. Rappel des principaux concepts : modèle LSTM, prédiction des prix du Bitcoin, intégration de données externes 33](#_Toc135079452)

[2. Orientation méthodologique 33](#_Toc135079453)

[A. Type de méthode utilisée 33](#_Toc135079454)

[3. Terrain de recherche 35](#_Toc135079455)

[A. Présentation du cadre de recherche 35](#_Toc135079456)

[4. Justification du choix de la période d'étude 36](#_Toc135079457)

[A. Raisons pour sélectionner une période spécifique 36](#_Toc135079458)

[B. Explication de l'importance de cette période pour la recherche 37](#_Toc135079459)

[5. Collecte des données 38](#_Toc135079460)

[A. Échantillonnage des données 38](#_Toc135079461)

[B. Modes de collecte des données 39](#_Toc135079462)

[C. Justification de l'adéquation des données collectées pour l'entraînement du modèle LSTM 41](#_Toc135079463)

[6. Traitement et analyse des données 43](#_Toc135079464)

[A. Prétraitement des données 43](#_Toc135079465)

[B. Entraînement du modèle LSTM 44](#_Toc135079466)

[IV. Analyse et Résultats 47](#_Toc135079467)

[1. Mise en œuvre du modèle LSTM 47](#_Toc135079468)

[A. Configuration et entraînement du modèle 47](#_Toc135079469)

[B. Importance des paramètres : 48](#_Toc135079470)

[C. Validation et test du modèle 51](#_Toc135079471)

[D. Analyse des résultats 52](#_Toc135079472)

[2. Synthèse des résultats 57](#_Toc135079473)

[3. Importance des données externes 61](#_Toc135079474)

[V. Discussion 64](#_Toc135079475)

[1.Interprétation des résultats 64](#_Toc135079476)

[2. Limitations de l'étude 65](#_Toc135079477)

[3. Implications pour les traders et les investisseurs 66](#_Toc135079478)

[4. Suggestions pour les recherches futures 67](#_Toc135079479)

[VI. Conclusion 69](#_Toc135079480)

[1. Résumé des résultats clés 69](#_Toc135079481)

[2. Réponse à la question de recherche 70](#_Toc135079482)

[VII. Références 75](#_Toc135079483)

[Annexe 1 78](#_Toc135079484)

[Annexe 2 79](#_Toc135079485)

[Annexe 3 80](#_Toc135079486)

[Annexe 4 81](#_Toc135079487)

[Annexe 4 82](#_Toc135079488)

[Annexe 5 83](#_Toc135079489)

[Annexe 6 84](#_Toc135079490)

[Annexe 7 85](#_Toc135079491)

[Annexe 8 86](#_Toc135079492)

[Annexe 10 87](#_Toc135079493)

# Introduction

*« Citation éventuelle »*Source

La prédiction des prix des cryptos monnaies est devenue un enjeu majeur dans le domaine de la finance et de l'économie numérique, en raison de l'essor fulgurant et de l'adoption croissante des cryptos monnaies à travers le monde. En tant qu'actifs numériques décentralisés, les cryptos monnaies présentent des caractéristiques distinctes par rapport aux actifs financiers traditionnels, telles que la volatilité élevée des prix, l'absence de régulation centralisée et la dépendance à l'égard de la technologie blockchain. Ces particularités ont conduit les chercheurs à explorer de nouvelles méthodologies et approches pour mieux comprendre et anticiper les fluctuations des prix des cryptos monnaies.

Dans ce contexte, la prédiction des prix des cryptos monnaies s'est avérée être une question de recherche complexe et multidimensionnelle, nécessitant une compréhension approfondie des facteurs sous-jacents et des mécanismes qui régissent les marchés des cryptos monnaies. Les chercheurs ont abordé cette question en adoptant diverses perspectives, en combinant des approches traditionnelles telles que l'analyse fondamentale et l'analyse technique avec des méthodes innovantes issues du domaine de l'intelligence artificielle, telles que le machine learning et le deep learning.

L'analyse fondamentale vise à évaluer la valeur intrinsèque d’une crypto-monnaie en examinant les facteurs économiques, financiers et politiques qui peuvent influencer sa demande et son offre, tels que la croissance économique, les politiques monétaires et fiscales, et les développements technologiques. Toutefois, l'application de l'analyse fondamentale aux cryptos monnaies est limitée en raison de la nature décentralisée et anonyme de ces actifs, ainsi que de l'absence de données financières et économiques comparables à celles des actifs traditionnels.

L'analyse technique, quant à elle, se concentre sur l'étude des mouvements passés des prix et des volumes de transactions pour identifier des modèles et des tendances qui pourraient se reproduire à l'avenir. Cette approche est largement utilisée dans les marchés des cryptos monnaies en raison de leur forte volatilité et de la disponibilité de données historiques sur les prix et les volumes de transactions. Néanmoins, l'efficacité de l'analyse technique dans la prédiction des prix des cryptos monnaies reste un sujet de débat, en raison de la nature spéculative et irrationnelle des comportements des investisseurs sur ces marchés.

C'est dans ce contexte que les modèles de machine learning et de deep learning ont suscité un intérêt croissant parmi les chercheurs, en tant qu'outils puissants et flexibles pour traiter des volumes importants de données et apprendre des modèles complexes à partir de ces données. En particulier, les modèles LSTM, en tant que réseaux de neurones récurrents, ont démontré leur potentiel dans la prédiction des séries temporelles, comme les prix des cryptos monnaies. Grâce à leur architecture unique, les LSTM sont capables de traiter des séquences de données de longue durée et de prendre en compte les dépendances temporelles à long terme, ce qui les rend particulièrement adaptés aux problèmes de prédiction des prix des cryptos monnaies.

Les modèles LSTM ont été appliqués avec succès dans divers domaines tels que la reconnaissance vocale, la traduction automatique et la génération de texte, et ont montré des performances prometteuses dans la prédiction des prix des cryptos monnaies. Cependant, la majorité des travaux de recherche sur les LSTM et la prédiction des prix des cryptos monnaies se sont concentrés sur le Bitcoin avec des données basiques comme le prix, haut, bas, capitalisation, sans tenir compte des facteurs externes qui pourraient influencer les mouvements des prix. De plus, les recherches sont axées sur le Bitcoin et peu sur les autres crypto-monnaies ce qui représentent un manquement.

Il est important de noter que les prix des cryptos monnaies sont également influencés par des facteurs externes, tels que les événements géopolitiques, les développements technologiques et les réactions des réseaux sociaux. L'intégration de ces données externes dans les modèles LSTM pourrait permettre de capturer des informations supplémentaires pertinentes pour la prédiction des prix des cryptos monnaies et d'améliorer la précision des prédictions.

Il faut noter que les travaux sur le sujet prennent comme période des données comprises avant 2018 pour la plus part. Cette temporalité pénalise fortement les résultats car la volatilité était plus importante avant, mais aussi il manque de données pour entrainer le modèle.

Ainsi, il existe un besoin croissant de recherches qui explorent l'utilisation des modèles LSTM pour la prédiction des prix des cryptos monnaies en intégrant des données externes et en développant des méthodologies robustes pour évaluer les performances des modèles. Le présent travail s'inscrit dans cette direction, en proposant une approche innovante pour intégrer des données externes, telles que le crypto fear and greed, des indicateurs de prix comme le volume de transaction ou encore sa volatilité dans un modèle LSTM pour la prédiction des prix des cryptos monnaies.

## Analyse fondamentale et technique

L'analyse fondamentale et l'analyse technique constituent les deux approches traditionnelles pour la prédiction des prix des actifs financiers, y compris les cryptos monnaies. Ces approches visent à évaluer la valeur d'un actif en se basant sur des facteurs internes et externes, ainsi que sur des modèles et des tendances historiques des prix. Dans le contexte des cryptos monnaies, ces approches présentent des avantages et des limites qui méritent d'être discutés en détail.

#### Analyse fondamentale

L'analyse fondamentale cherche à déterminer le prix d'un actif en étudiant tous les facteurs économiques, financiers et politiques qui influencent sa demande et son offre. Pour les actifs traditionnels, cela inclut des facteurs tels que la croissance économique, les politiques monétaires et fiscales, les développements technologiques, les régulations gouvernementales et bien d’autres. L'analyse fondamentale repose sur l'hypothèse que les actifs sont correctement évalués sur le marché à long terme, et que les écarts entre la valeur intrinsèque et le prix du marché offrent des opportunités d'investissement.

Toutefois, l'application de l'analyse fondamentale aux cryptos monnaies présente plusieurs défis. Premièrement, la nature décentralisée et anonyme des cryptos monnaies rend difficile l'évaluation de leur valeur intrinsèque, car il n'existe pas de données financières et économiques comparables à celles des actifs traditionnels. Deuxièmement, les cryptos monnaies sont sujettes à des facteurs externes, tels que les événements géopolitiques et les développements technologiques, qui peuvent influencer leur valeur de manière imprévisible. Enfin, les marchés des cryptos monnaies sont souvent caractérisés par des comportements spéculatifs et irrationnels, ce qui peut entraîner des fluctuations de prix importantes indépendamment des facteurs fondamentaux.

#### Analyse technique

L'analyse technique, en revanche, se concentre sur l'étude des mouvements passés des prix et données relatives à l’actif telle que (le halving, gas fee, nombre de tx, le crypto fear and greed ) pour identifier des paternes et des tendances qui pourraient se reproduire à l'avenir. Cette approche repose sur l'hypothèse que les prix reflètent toutes les informations disponibles et que les mouvements passés des prix peuvent fournir des indices sur les mouvements futurs.

L'analyse technique est largement utilisée dans les marchés des cryptos monnaies, en raison de leur forte volatilité et de la disponibilité de données historiques. Cependant, l'efficacité de l'analyse technique dans la prédiction des prix des cryptos monnaies est sujette à controverse. D'une part, certains chercheurs soutiennent que l'analyse technique peut aider à identifier des opportunités d'investissement et à réduire les risques. D'autres soulignent que l'efficacité de l'analyse technique repose sur l'hypothèse que les comportements passés des investisseurs se répéteront à l'avenir, ce qui pourrait ne pas être valable dans le cas des cryptos monnaies, en raison de leur nature spéculative et de leur sensibilité aux facteurs externes. De plus, il existe un risque de biais de confirmation, où les analystes techniques pourraient interpréter les données historiques pour soutenir leurs préjugés et leurs attentes plutôt que de les évaluer objectivement.

Compte tenu des limites de l'analyse fondamentale et de l'analyse technique dans la prédiction des prix des cryptos monnaies, il est nécessaire d'explorer de nouvelles approches qui pourraient améliorer la précision des prédictions et fournir des informations supplémentaires sur les facteurs qui influencent les mouvements des prix. Les modèles de machine learning et de deep learning, en particulier les modèles LSTM, offrent une voie prometteuse pour combler ces lacunes en tirant parti de leur capacité à traiter des volumes importants de données et à apprendre des modèles complexes à partir de ces données.

Dans la partie suivante, nous discuterons de l'évolution des modèles de machine learning et de deep learning dans la prédiction des prix des cryptos monnaies et de la manière dont ils peuvent compléter et renforcer les approches traditionnelles de l'analyse fondamentale et de l'analyse technique.

## L'évolution des modèles de machine learning et de deep learning dans la prédiction des prix des cryptos monnaies

Au cours de la dernière décennie, les modèles de machine learning et de deep learning ont gagné en popularité dans la prédiction des prix des cryptos monnaies, en raison de leur capacité à traiter de grandes quantités de données et à identifier des modèles complexes qui échappent aux approches traditionnelles. Dans cette partie, nous examinerons l'évolution de ces modèles et discuterons de leur potentiel pour améliorer la précision des prédictions des prix des cryptos monnaies.

#### Modèles de machine learning

Les modèles de machine learning, tels que les régressions linéaires, les arbres de décision et les forêts aléatoires, ont été largement utilisés pour la prédiction des prix des cryptos monnaies. Ces modèles apprennent des relations entre les variables d'entrée, telles que les prix historiques, les volumes de transactions et les indicateurs techniques, et les variables de sortie, telles que les prix futurs ou les rendements. Les modèles de machine learning ont montré des performances prometteuses dans la prédiction des prix des cryptos monnaies, en particulier lorsqu'ils sont combinés avec des techniques d'optimisation des paramètres et de sélection des variables.

Cependant, les modèles de machine learning présentent plusieurs limites dans la prédiction des prix des cryptos monnaies. Tout d'abord, ils sont souvent sensibles aux problèmes d'overfitting, où les modèles apprennent des modèles spécifiques aux données d'entraînement, mais ne généralisent pas bien aux données de test ou aux conditions de marché inédites. Deuxièmement, les modèles de machine learning ont généralement du mal à capturer des modèles complexes et non linéaires, ce qui limite leur capacité à traiter des données de séries temporelles, telles que les prix des cryptos monnaies.

#### Modèles de deep learning

Les modèles de deep learning, en revanche, ont le potentiel de surmonter certaines des limites des modèles de machine learning grâce à leur capacité à apprendre des représentations de données de haut niveau et à traiter des données complexes, telles que des images, du texte et des séries temporelles. Les réseaux de neurones convolutifs (CNN), les réseaux de neurones récurrents (RNN) et les modèles LSTM sont quelques-uns des modèles de deep learning les plus populaires utilisés pour la prédiction des prix des cryptos monnaies.

Les modèles LSTM, en particulier, sont particulièrement prometteurs pour la prédiction des prix des cryptos monnaies, en raison de leur capacité à prendre en compte les dépendances temporelles à long terme et à éviter les problèmes de disparition des gradients. Les recherches récentes ont montré que les modèles LSTM peuvent surpasser les modèles de machine learning traditionnels et les approches d'analyse fondamentale et technique dans la prédiction des prix des cryptos monnaies.

## Intégration des données externes pour améliorer la prédiction des prix des cryptos monnaies

L'une des principales limitations des approches traditionnelles de prédiction des prix des cryptos monnaies, telles que l'analyse fondamentale et l'analyse technique, ainsi que des modèles de machine learning et de deep learning, est leur dépendance aux données historiques des prix. Cependant, les cryptos monnaies sont également influencées par de nombreux facteurs externes, tels que les réactions des réseaux sociaux, le sentiment du marché, les événements géopolitiques et les développements technologiques. Dans cette partie, nous examinerons comment l'intégration de données externes peut améliorer la prédiction des prix des cryptos monnaies.

#### Données des réseaux sociaux

Les réseaux sociaux jouent un rôle crucial dans la diffusion des informations et des opinions sur les cryptos monnaies, et peuvent influencer la perception des investisseurs et les décisions d'achat et de vente. Les tweets, les publications sur les forums et les articles de blog sont autant de sources potentielles d'informations qui peuvent être utilisées pour enrichir les modèles de prédiction des prix des cryptos monnaies. Par exemple, l'analyse du nombre de tweets mentionnant #bitcoin, ou du sentiment général exprimé dans ces tweets, peut fournir des indications sur l'engouement du marché pour une crypto monnaie particulière et sur les attentes des investisseurs concernant ses performances futures.

#### Indicateurs de sentiment du marché

Le sentiment du marché, c'est-à-dire l'attitude collective des investisseurs à l'égard d'un actif financier, peut également influencer les mouvements des prix des cryptos monnaies. Les indicateurs de sentiment, tels que le crypto fear and greed index, mesurent les émotions et les perceptions des investisseurs en fonction de divers facteurs, tels que les fluctuations des prix, les volumes de transactions et les nouvelles du marché. L'intégration de ces indicateurs dans les modèles de prédiction des prix des cryptos monnaies peut aider à mieux comprendre les dynamiques du marché et à anticiper les mouvements futurs des prix.

#### Données économiques et géopolitiques

Les événements économiques et géopolitiques, tels que les politiques monétaires et fiscales, les régulations gouvernementales et les conflits internationaux, peuvent également avoir un impact sur les prix des cryptos monnaies. L'intégration de ces données dans les modèles de prédiction des prix peut fournir des informations supplémentaires sur les facteurs externes qui influencent les mouvements des prix et aider à améliorer la précision des prédictions.

L'intégration des données externes dans les modèles de prédiction des prix des cryptos monnaies, tels que les modèles LSTM, peut améliorer la compréhension des facteurs qui influencent les mouvements des prix et renforcer la précision des prédictions. Dans les sections suivantes de ce mémoire, nous présenterons une méthodologie pour intégrer des données externes dans notre modèle LSTM et évaluerons l'impact de cette intégration sur les performances du modèle en termes de prédiction des prix. Nous discuterons également des défis associés à la collecte, au traitement et à l'analyse de ces données externes, ainsi que des stratégies pour surmonter ces défis.

Enfin, nous explorerons des pistes pour des recherches futures dans ce domaine, notamment en ce qui concerne l'identification de nouvelles sources de données externes, l'amélioration des techniques d'analyse du sentiment et l'évaluation de l'importance relative de différents facteurs externes dans la prédiction des prix des cryptos monnaies. En combinant ces efforts avec l'application de modèles de deep learning avancés, tels que les LSTM, nous espérons contribuer à l'avancement des connaissances et des pratiques dans le domaine de la prédiction des prix des cryptos monnaies et fournir des outils utiles pour les investisseurs, les chercheurs et les décideurs.

## Défis et opportunités pour l'amélioration des modèles de prédiction des prix des cryptos monnaies

La prédiction des prix des cryptos monnaies est une tâche complexe et délicate, compte tenu de la nature volatile et imprévisible de ces actifs numériques. Toutefois, le choix du modèles de deep learning, LSTM, avec l'intégration de données externes, offrent de nouvelles opportunités pour améliorer la précision des prédictions. Dans cette partie, nous discuterons des défis et des opportunités associés à l'amélioration des modèles de prédiction des prix des cryptos monnaies.

#### Amélioration de la qualité des données

L'un des principaux défis dans la prédiction des prix des cryptos monnaies est la qualité et la fiabilité des données utilisées pour entraîner et évaluer les modèles. Les données incomplètes, inexactes, biaisées ou tout simplement trop compliqué ou chère à récupérer peuvent entraîner des erreurs de prédiction, des impossibilités et limiter la performance des modèles. Pour relever ce défi, les chercheurs doivent s’investir dans la collecte, le nettoyage et l'analyse des données, ainsi que dans le développement de méthodes pour évaluer et améliorer la qualité des données.

#### Exploration de nouvelles sources de données

Comme mentionné précédemment, l'intégration de données externes, telles que les réactions des réseaux sociaux, les indicateurs de sentiment du marché et les données économiques et géopolitiques, peut améliorer la prédiction des prix des cryptos monnaies. Cependant, il reste encore de nombreuses sources de données potentielles à explorer, notamment les données sur les transactions et les portefeuilles des investisseurs, les développements technologiques dans l'espace des cryptos monnaies, la volatilité, son adoption, son utilisation dans des cas de la vie de tous les jours ou les événements majeurs du marché sans autant de facteur qui peuvent influencer son prix. L'exploration et l'intégration de ces sources de données représentent une opportunité importante pour améliorer les modèles de prédiction des prix des cryptos monnaies.

#### Avancées dans les techniques de modélisation

Les modèles de deep learning, tels que les LSTM, ont montré des performances prometteuses dans la prédiction des prix des cryptos monnaies. Néanmoins, il existe encore un potentiel d'amélioration en explorant de nouvelles architectures de réseaux de neurones, en optimisant les paramètres et les hyperparamètres des modèles, et en développant des méthodes pour éviter les problèmes d'overfitting et d'underfitting. Les avancées dans les techniques de modélisation, telles que les approches d'apprentissage ensembliste et les techniques de régularisation, offrent des opportunités pour améliorer la performance des modèles de prédiction des prix des cryptos monnaies.

#### Évaluation des modèles et comparaison des performances

Un défi clé dans la prédiction des prix des cryptos monnaies est l'évaluation et la comparaison des performances des différents modèles. Les chercheurs doivent développer des méthodes rigoureuses pour évaluer les performances des modèles en termes de précision, de robustesse et de généralisation. Ceci peut inclure l'utilisation de métriques appropriées, telles que l'erreur quadratique moyenne (MSE), l'erreur absolue moyenne (MAE), ou le coefficient de détermination (R²) ainsi que la mise en œuvre de méthodes de validation croisée pour évaluer la performance des modèles sur différents ensembles de données et dans différentes conditions de marché. La comparaison des performances des modèles de deep learning, tels que les LSTM, avec d'autres approches de prédiction, telles que l'analyse fondamentale, l'analyse technique et les modèles de machine learning traditionnels, permettra d'identifier les forces et les faiblesses de chaque approche et d'orienter les recherches futures dans ce domaine.

#### Adaptabilité et réactivité aux changements de marché

Le marché des cryptos monnaies est en constante évolution, avec de nouvelles cryptos monnaies émergentes, des développements technologiques et des changements réglementaires qui peuvent influencer les prix. Les modèles de prédiction des prix des cryptos monnaies doivent être capables de s'adapter rapidement à ces changements pour maintenir leur précision et leur pertinence. Les chercheurs et les praticiens doivent donc explorer des stratégies pour rendre les modèles plus adaptables, par exemple en mettant à jour régulièrement les données d'entraînement, en ajustant les paramètres des modèles en fonction des conditions de marché, ou en intégrant des mécanismes d'apprentissage en ligne pour permettre aux modèles d'apprendre de nouvelles informations au fur et à mesure qu'elles deviennent disponibles.

Les défis et les opportunités associés à l'amélioration des modèles de prédiction des prix des cryptos monnaies offrent un potentiel important pour les recherches futures dans ce domaine. En abordant ces défis et en exploitant ces opportunités, les chercheurs peuvent contribuer à l'avancement des connaissances et des pratiques en matière de prédiction des prix des cryptos monnaies, et fournir des outils utiles pour les investisseurs, les chercheurs et les décideurs.

## Définition du concept principal

Dans notre travail de prédiction des prix des cryptos monnaies, nous mobilisons principalement deux concepts clés, à savoir le deep learning et le modèle LSTM. Le deep learning est une branche de l'apprentissage automatique qui permet de traiter des données complexes en utilisant des réseaux de neurones profonds. Il s'agit d'une méthode d'apprentissage supervisé qui permet de trouver des modèles dans les données en utilisant des couches successives de neurones qui effectuent des calculs et des transformations sur les données d'entrée.

Le deep learning a connu un essor considérable ces dernières années grâce à l'augmentation de la puissance de calcul et des volumes de données disponibles. Cette technique est aujourd'hui largement utilisée dans des domaines tels que la reconnaissance d'image, la traduction automatique ou encore la reconnaissance de la parole. Dans le domaine de la finance, le deep learning est également utilisé pour la prédiction de prix d'actifs financiers tels que les actions ou les devises.

Le modèle LSTM (Long Short-Term Memory) est un type de réseau de neurones récurrents qui est capable de traiter des séquences de données de longue durée en évitant le problème de la disparition des gradients. Les réseaux de neurones récurrents sont des modèles qui permettent de traiter des séquences de données en utilisant une boucle de rétroaction qui permet de conserver une mémoire des informations précédentes. Cependant, les réseaux de neurones récurrents classiques ont tendance à souffrir de la disparition des gradients, ce qui limite leur capacité à prendre en compte les informations à long terme.

Le modèle LSTM a été proposé pour pallier ce problème en utilisant une architecture particulière qui permet de conserver une mémoire à long terme tout en évitant la disparition des gradients. Ce modèle est capable de traiter des séquences de données de longue durée en prenant en compte les relations entre les différents éléments de la séquence. Il est ainsi particulièrement adapté à la prédiction de prix de cryptos monnaies qui sont caractérisés par une forte volatilité et une complexité importante.

En combinant ces deux concepts, nous pouvons ainsi construire un modèle de deep learning basé sur le modèle LSTM pour prédire les prix des cryptos monnaies. En utilisant les données historiques de prix et le crypto fear and greed, notre modèle pourra apprendre les relations complexes entre ces variables et les prix des cryptos monnaies. Le modèle sera ainsi capable de fournir des prévisions précises et fiables pour aider les investisseurs à prendre des décisions éclairées.

## Identification d'un manque dans la littérature

La prédiction des prix des cryptomonnaies est un sujet de recherche en pleine croissance, en raison de l'importance croissante de ces actifs numériques dans l'économie mondiale. Les modèles de deep learning ont été largement utilisés pour prédire les prix des crypto-monnaies, mais ils présentent des limites significatives en termes de précision et de capacité à intégrer des facteurs externes tels que les réactions des réseaux sociaux ou le sentiment général du marché. Cette partie met en évidence le manque dans la littérature concernant l'utilisation de modèles de deep learning, comme le LSTM, pour prédire les prix des cryptomonnaies en intégrant des données externes.

### Les limites des modèles de machine learning actuels

#### Précision des prédictions

Les modèles de machine learning traditionnels, tels que la régression linéaire, les arbres de décision et les forêts aléatoires, ont été largement utilisés pour prédire les prix des cryptomonnaies. Cependant, ces modèles souffrent souvent de problèmes de surajustement ou de sous-ajustement, ce qui limite leur précision et leur capacité à généraliser les tendances sur de nouvelles données. De plus, les cryptomonnaies étant extrêmement volatiles et sujettes à des variations rapides de prix, les modèles de machine learning classiques peuvent ne pas être suffisamment réactifs pour suivre ces fluctuations.

#### Capacité à intégrer des facteurs externes

Un autre défi majeur pour les modèles de machine learning traditionnels est leur capacité limitée à intégrer des facteurs externes qui influencent les prix des cryptomonnaies. Par exemple, les réactions des réseaux sociaux et le sentiment général du marché ont été identifiés comme des facteurs importants pour prédire les prix des cryptomonnaies. Cependant, les modèles de machine learning classiques ont généralement du mal à prendre en compte ces données non structurées et à les intégrer de manière significative dans leurs prédictions.

### Les avantages potentiels des modèles de deep learning

#### Meilleure précision des prédictions

Les modèles de deep learning, comme les réseaux de neurones à long terme à court terme (LSTM), peuvent potentiellement offrir une meilleure précision des prédictions par rapport aux modèles de machine learning traditionnels. Les LSTM sont particulièrement adaptés pour traiter des séquences de données temporelles, comme les séries chronologiques des prix des cryptomonnaies. Grâce à leur capacité à mémoriser des informations sur de longues périodes et à prendre en compte les dépendances temporelles, les LSTM pourraient mieux modéliser les fluctuations rapides et volatiles des prix des cryptomonnaies.

#### Capacité à intégrer des facteurs externes

Les modèles de deep learning peuvent également être plus efficaces pour intégrer des facteurs externes dans leurs prédictions. Par exemple, les LSTM peuvent être combinés avec d'autres types de réseaux de neurones, tels que les réseaux de neurones convolutifs (CNN), pour traiter des données non structurées telles que les textes provenant des réseaux sociaux ou des articles de presse. En intégrant ces données externes, les modèles de deep learning pourraient potentiellement mieux comprendre les facteurs qui influencent les prix des cryptomonnaies et améliorer la précision des prédictions.

### Modèles hybrides pour prédire les prix des cryptomonnaies

#### Combiner LSTM et CNN pour traiter des données multimodales

Un moyen prometteur d'intégrer des données externes et d'améliorer la précision des prédictions consiste à utiliser des modèles hybrides qui combinent plusieurs types de réseaux de neurones. Par exemple, un modèle hybride pourrait combiner un LSTM pour traiter des séries temporelles de prix et un CNN pour traiter des données non structurées provenant des réseaux sociaux et des articles de presse. Le CNN pourrait être utilisé pour extraire des informations pertinentes à partir de textes, telles que le sentiment ou les tendances du marché, qui pourraient ensuite être intégrées au LSTM pour améliorer la prédiction des prix des cryptomonnaies.

#### Transformer les modèles pour une meilleure prise en compte du contexte

Les modèles de type Transformer, tels que BERT ont révolutionné le traitement du langage naturel et se sont avérés particulièrement performants pour la compréhension du contexte et la prise en compte des dépendances à long terme dans les textes. Ces modèles pourraient également être utilisés dans le contexte de la prédiction des prix des cryptomonnaies pour mieux prendre en compte les données non structurées et les intégrer aux prédictions. En combinant les capacités des modèles de type Transformer avec celles des LSTM, il pourrait être possible de créer des modèles de deep learning encore plus efficaces pour prédire les prix des cryptomonnaies.

### Défis et perspectives de recherche

#### Données de qualité et prétraitement

L'un des principaux défis pour l'utilisation des modèles de deep learning dans la prédiction des prix des cryptomonnaies est la qualité et la disponibilité des données. Les données externes, telles que les réactions des réseaux sociaux et le sentiment général du marché, sont souvent non structurées et difficiles à obtenir. Un prétraitement soigneux et une sélection des caractéristiques pertinentes sont essentiels pour améliorer la qualité des données et leur intégration dans les modèles de deep learning.

#### Complexité des modèles et besoins en ressources

Les modèles de deep learning, en particulier les modèles hybrides et les modèles de type Transformer, peuvent être complexes et exiger des ressources de calcul importantes pour l'entraînement et l'inférence. La recherche doit donc se concentrer sur le développement de méthodes d'optimisation et de réduction de la complexité des modèles pour faciliter leur utilisation dans la prédiction des prix des cryptomonnaies.

#### Validation et évaluation des performances

Il est crucial de valider et d'évaluer les performances des modèles de deep learning pour la prédiction des prix des cryptomonnaies en utilisant des méthodologies rigoureuses et des ensembles de données représentatifs. La comparaison des performances des modèles de deep learning avec celles des modèles de machine learning traditionnels est essentielle pour déterminer l'efficacité de ces approches. Les chercheurs doivent utiliser des indicateurs de performance appropriés, tels que l'erreur quadratique moyenne (RMSE), le coefficient de détermination (R²) et l'erreur absolue moyenne (MAE), pour évaluer et comparer les prédictions des différents modèles. Des méthodes de validation croisée et des ensembles de données de test indépendants devraient être utilisés pour évaluer la capacité des modèles à généraliser sur de nouvelles données et éviter le surajustement.

#### Lacune dans la littérature

Il existe un manque significatif dans la littérature concernant l'utilisation de modèles de deep learning, tels que le LSTM, pour prédire les prix des cryptomonnaies en intégrant des données externes. Les modèles de deep learning présentent un potentiel considérable pour améliorer la précision des prédictions et tenir compte des facteurs externes tels que les réactions des réseaux sociaux et le sentiment général du marché. L'exploration de modèles hybrides combinant LSTM, CNN, et Transformer pourrait mener à des avancées importantes dans la prédiction des prix des cryptomonnaies.

Pour combler cette lacune dans la littérature, les chercheurs doivent relever plusieurs défis, notamment la qualité et la disponibilité des données, la complexité des modèles et les besoins en ressources de calcul, ainsi que la validation et l'évaluation des performances. En relevant ces défis et en développant des modèles de deep learning efficaces pour prédire les prix des cryptomonnaies, il sera possible de contribuer à une meilleure gestion des risques pour les investisseurs et les acteurs du marché, tout en élargissant notre compréhension du comportement des cryptomonnaies sur le marché.

### Formulation de la problématique

Notre mémoire vise à répondre à la question suivante :

**Comment peut-on utiliser le modèle LSTM, pour prédire les prix journaliers du Bitcoin en intégrant des données externes ?**

Cette problématique souligne l'importance d'examiner les approches basées sur le deep learning, et plus particulièrement les réseaux LSTM, pour améliorer la précision des prédictions des prix des cryptomonnaies. L'intégration de données externes sont un aspect crucial pour mieux comprendre l'évolution des prix et fournir des prédictions plus précises et robustes. En abordant cette problématique, nous cherchons à combler le manque existant dans la littérature et à contribuer à la conversation académique autour de la prédiction des prix des cryptomonnaies en explorant de nouvelles méthodes et approches pour améliorer les performances des modèles prédictifs.

### Présentation de la stratégie de recherche approfondie, annonce des contributions et du plan

Notre stratégie de recherche se déroulera en plusieurs étapes détaillées, visant à garantir une analyse rigoureuse et une évaluation complète des modèles de deep learning, en particulier des modèles LSTM, pour prédire les prix des cryptomonnaies en intégrant des données externes.

Avant de commencer notre recherche, nous effectuerons une revue approfondie de la littérature pour examiner les approches existantes dans la prédiction des prix du Bitcoin et les modèles de deep learning. Cette étape nous permettra d'identifier les lacunes et les opportunités d'amélioration dans la littérature actuelle.

Nous collecterons des données provenant de diverses sources, telles que les plateformes d'échange de cryptomonnaies, les réseaux sociaux et les sites Web d'analyse du marché. Nous rassemblerons des informations sur les prix, les volumes d'échange, les variations, le plus haut, le plus bas, la volatilité, le prix de l’Ethereum et sentiment de marché avec le crypto fear and greed index.

Les données collectées seront nettoyées, normalisées et transformées pour garantir leur qualité et leur pertinence pour l'entraînement des modèles LSTM. Les étapes de prétraitement incluront le traitement des valeurs manquantes, la normalisation des données et la création de séquences temporelles pour l'entraînement du modèle.

Nous concevrons un modèle LSTM adapté pour prédire les prix du Bitcoin en intégrant les données externes collectées. Nous entraînerons le modèle en utilisant des techniques d'optimisation et de régularisation pour éviter le surajustement et améliorer la généralisation du modèle.

Après l'entraînement, nous utiliserons des métriques telles que l'erreur quadratique moyenne (RMSE), le coefficient de détermination (R²) et l'erreur absolue moyenne (MAE) pour comparer les performances des différents modèles.

Nous analyserons les résultats de notre modèle LSTM en termes de précision des prédictions, en identifiant les forces et les faiblesses de notre approche. Nous discuterons également des implications de nos résultats pour les investisseurs et les acteurs du marché des cryptomonnaies, ainsi que des directions possibles pour améliorer et étendre notre modèle.

Suite à l'analyse des résultats et à la discussion des performances de notre modèle LSTM, nous identifierons les domaines dans lesquels le modèle pourrait être amélioré. Nous pourrions ajuster les hyperparamètres du modèle, modifier l'architecture du réseau ou explorer de nouvelles sources de données pour améliorer la précision des prédictions. Cette étape de validation et de révision sera essentielle pour garantir que notre modèle LSTM est robuste et adapté à la prédiction des prix des cryptomonnaies.

# Revue de littérature

Les cryptomonnaies, telles que Bitcoin et Ethereum, ont attiré une attention considérable ces dernières années en raison de leur nature décentralisée et de leur potentiel de rendement élevé. Cependant, elles sont également connues pour leur imprévisibilité, ce qui rend difficile pour les investisseurs de prendre des décisions d'investissement éclairées. Cette imprévisibilité découle du fait que le marché des cryptomonnaies en est encore à ses débuts et qu'il existe de nombreux facteurs qui peuvent affecter le prix des cryptomonnaies, revue tels que les changements réglementaires, les taux d'adoption et les avancées technologiques.

Pour remédier à cette situation, il y a un intérêt croissant pour le développement de modèles de prédiction pour aider les investisseurs à naviguer sur le marché volatile des cryptomonnaies. Ces modèles permettront à tous investisseurs d’avoir une solution lui permettant la compréhension des prix. Ces derniers peuvent utiliser toutes les données possibles pour alimenter sa manière d’apprendre. Ils peuvent apprendre avec toutes les techniques et méthodes déjà connue allant d’analyse technique en passant par l’analyse fondamental à l'apprentissage automatique pour prédire le prix futur des cryptomonnaies. L'objectif de ces modèles est de fournir aux investisseurs des informations plus précises et plus opportunes, ce qui peut les aider à prendre de meilleures décisions d'investissement. L’intérêt pour les financiers sont de pouvoir diversifier leur investissement dans les crypto-monnaies, (Basher & Sadorsky) indiquent que le Bitcoin peut s’utiliser comme substitue à l’or pour diversifier. De plus ils indiquent que le pétrole peut être utiliser dans l’apprentissage pour baisser la volatilité.

Cependant, le développement de modèles de prédiction pour les cryptomonnaies n'est pas sans défis. L'un des plus grands défis est le manque de données historiques pour de nombreuses cryptomonnaies, ce qui rend difficile l'entraînement et la validation de ces modèles, (Kim & Bock & Lee) soutient que l’intégration d’autres facteurs que le prix est indispensable pour améliorer la prédiction. De plus, le marché des cryptomonnaies est encore largement non réglementé, ce qui signifie qu'il y a beaucoup d'incertitude quant à la façon dont le marché évoluera à l'avenir.

Malgré ces défis, il y a encore beaucoup d'intérêt pour le développement de modèles de prédiction pour les cryptomonnaies avec le deep learning. À mesure que le marché continue d'évoluer et de mûrir, nous pouvons nous attendre à voir des modèles plus sophistiqués développés qui sont mieux capables de prédire le prix futur des cryptomonnaies. (Nazareth & Reddy) répertorient les analyses majeures dans le maching et deep learning sur les différentes classes d’actifs financiers, que la recherche autour de ce sujet a touché toutes les classes d’actifs avec différents modèles et techniques d’apprentissages qui laisse à penser que le chemin vers des modèles efficaces reste long.

Les informations sur les cryptomonnaies sont à la fois à la portée de tous mais nécessitent une connaissance profonde de l’écosystème, ainsi pour les investisseurs se fient aux prix historiques et aux informations limitées disponibles, qui sont souvent agrégées principalement sur Twitter. Contrairement à d'autres classes d'actifs, comme les entreprises, qui disposent de résultats de vente, de directives d'entreprise et d'un compte de résultat qui permet l'évaluation, les cryptomonnaies manquent de ces sources de données. La plupart des articles mettent en évidence que le manque de données est un des facteurs qui diminuent la bonne prédiction des prix de l’article Ethereum forecasting utiliszing ML and DL et (Gerritsen & Lugtigheid & Walther) concluent que le manque de données pénalise l’apprentissage des modèles mais que dans la futur les données plus conséquentes ainsi que le recul sur ces informations donneront des modèles plus performants.

En raison de la nature décentralisée des cryptomonnaies, il n'y a pas de structure de gouvernance ou de reporting centralisé pour les investisseurs. Les prix des cryptomonnaies sont principalement influencés par la demande et l'offre sur les différentes plateformes d'échange. Les investisseurs doivent donc prendre en compte un large éventail de facteurs lorsqu'ils décident d'investir dans une cryptomonnaie, tels que la sécurité du réseau, l'adoption du marché, la concurrence et les réglementations gouvernementales. (Gerritsen & Lugtigheid & Walther) mettent en avance que les impacts externes comme le sentiment de marché, ou des données macro-économiques pourrait améliorer les modèles.

En outre, contrairement aux entreprises traditionnelles qui sont évaluées en fonction de leur performance financière, les cryptomonnaies sont souvent évaluées en fonction de leur potentiel d'adoption future et de leur capacité à résoudre des problèmes existants dans l'économie et la société. Par conséquent, il est souvent difficile de faire des prévisions précises sur l'évolution future des prix des cryptomonnaies en partant de ces postulats car ce sont des paramètres qui sont dur à quantifier et qu’y restent de l’estimation.

Au fil du temps, plusieurs modèles ont été développés pour prédire le prix des actifs. Ces modèles peuvent être classés en trois catégories principales : les modèles de séries temporelles avec GARCH, ARIMA, les modèles de régression avec des régressions linéaires, des régressions logistiques, ainsi que les modèles nouveaux que sont l'apprentissage par le machine et deep learning. (Lahmiri & Bekiros) montrent que le modèle qui fonctionne le mieux pour le Bitcoin est le modèle d’apprentissage qui dépense les deux autres approches.

Les modèles de séries temporelles sont très efficaces pour prédire les prix des actions, mais sont très limités pour les cryptomonnaies en raison de diverses raisons. Les cryptomonnaies sont une classe d'actifs relativement nouvelle, et le marché est moins mature par rapport aux marchés financiers traditionnels, ce qui rend difficile la capture des motifs et des tendances du marché en utilisant des modèles de séries temporelles traditionnels. Les modèles ARIMA et GARCH ne sont pas efficace pour les crypto-monnaies alors qu’ils sont des analyses incontournables dans les actifs classiques comme les actions.

Les modèles de régression tels que la régression linéaire et la régression logistique sont largement utilisés en finance et en économie. Cependant, ces modèles ne sont pas très efficaces pour prédire les prix des cryptomonnaies. L'une des principales raisons est que les prix des cryptomonnaies ne sont pas linéairement liés à leurs facteurs fondamentaux tels que l'offre, la demande et l'adoption. On note que (Galeshchuk) indiquent que les modèles d’apprentissages surpassent les modèles de régressions ce qui va dans le sens que ces modèles peuvent être efficace. L’étude de (Peng & Albuqerque & Sà & Padula & Montenegro) comparent GARCH et SVR et montre que le modèle d’apprentissage est plus performant sur les actifs cryptographiques.

En revanche, les modèles d'apprentissage de deep learning ont montré des résultats prometteurs dans la prédiction des prix des cryptomonnaies. Les modèles d'apprentissage sont efficaces pour analyser les données historiques, identifier les motifs et les tendances, et faire des prévisions en fonction des motifs identifiés. Les modèles de deep learning, en particulier, sont bien adaptés pour gérer les relations complexes et non linéaires dans le marché des cryptomonnaies. Différents modèles existent dans le deep learning et son confondue (Nasirtafreshi) montre que le modèle LSTM surperforme les autres modèles par ces caractéristiques.

Dans l'ensemble, bien que chaque catégorie de modèles ait ses avantages et ses inconvénients, les modèles d'apprentissage, en particulier les modèles de deep learning, ont montré les résultats les plus prometteurs dans la prédiction des prix des cryptomonnaies. À mesure que le marché des cryptomonnaies continue d'évoluer, nous pouvons nous attendre à ce que des modèles plus avancés et sophistiqués soient développés pour mieux capturer les motifs et les tendances du marché. (Jaquart & Dann & Weinhardt) avancent que la régulation sera un avantage décisif pour améliorer les modèles.

Les critiques académiques sur la prédiction des prix des cryptomonnaies mettent en évidence plusieurs limites et défis à relever pour améliorer la précision des modèles de prédiction. En effet, bien que l'apprentissage puisse donner de bons résultats dans certaines périodes, il est important de prendre en compte l'évolution constante des marchés financiers et des cryptomonnaies, ainsi que la nécessité de tenir compte de données externes supplémentaires. Dans la revue de (Rathore, Mishra & Mehra & Pal & Hashim & Shapi’i & Ciano & Shutaywi), les données externes sont indiquées comme des données à intégrer, il met en évidence que la saisonnalité est un facteur à prendre en compte dans son analyse et son modèle.

Le manque de données externes telles que le sentiment, la saisonnalité et les données de publication sur Twitter est souvent souligné comme un facteur d’amélioration dans les modèles de prédiction des prix des cryptomonnaies. En effet, ces données externes peuvent avoir un impact significatif sur la demande et l'offre de cryptomonnaies et influencer les prix de manière importante. L'intégration de ces données externes peut donc améliorer la précision des modèles de prédiction. Dans la revue de (Heston & Sinha), les informations nouvelles sur un actif influencent le prix. Il indique également que le temps de réaction et l’exploration de pattern est également nécessaire pour prendre en compte ces nouvelles.

Par ailleurs, la volatilité des prix des cryptomonnaies est souvent un obstacle pour les modèles de prédiction. L’article de (Mallqi & Fernandes) indique que la volatilité est le challenge dans les modèles d’apprentissages, il met en avant que ces modèles permettent d’améliorer la prédiction. Il conclue en affirmant que son modèle augmente sa prédiction de 10% pour seulement 1-2% d’erreur avec des modèles SVM et ANN, le modèles SVM est le plus performant des deux. La volatilité peut rendre certains modèles inadéquats pour prédire les prix, car les fluctuations rapides et imprévisibles des prix peuvent rendre difficile l'identification des tendances.

Cependant, l'ajout de données externes et la prise en compte de la volatilité des prix des cryptomonnaies peuvent être améliorés avec plus de données. En effet, les modèles d'apprentissage peuvent bénéficier d'un volume plus important de données pour améliorer leur précision et leur capacité à capturer les tendances et les fluctuations du marché. Il est également important de souligner que les résultats des études sur les modèles de prédiction des prix des cryptomonnaies peuvent varier en fonction des résultats souhaités et de la période de temps choisie. Il n'y a pas de modèle unique qui puisse être considéré comme le plus précis, et chaque étude doit être adaptée à ses besoins spécifiques. C’est le point qu’à soulever (Peng & Albuqerque & Sà & Padula & Montenegro), en indiquant que la période de temps est centrale et peut poser de réel problème et peut donner en sortie des données absurdes.

Il est important de noter que de nombreuses études ne prennent pas en compte les frais de transaction, qui peuvent avoir un impact significatif sur la précision des modèles de prédiction. Dans la revue de (Jaquart & Dann & Weinhardt), indique que la prise en compte des frais de transaction est indispensable, de plus ces frais sont variables et fluctue énormément en fonction de la demande, il faut donc les prendre en compte pour entrainer le modèle comme pour les prendre en compte dans l’approche des plus-values. L'inclusion de ces frais peut améliorer la précision des modèles et leur capacité à prédire les prix de manière plus fiable, avec le principe souvent qui est que plus les frais de transaction sont élevés, plus il y a de transaction demandée sur la même période. Ainsi il va que de nombreux investisseurs souhaitent intervenir sur le marché et changer son prix. Il peut être de bon indicateur de volatilité ou d’anticipation de mouvement de prix.

Cependant, la plupart des modèles d'apprentissage se basent principalement sur des données historiques telles que les prix passés, les volumes de transactions, les indicateurs techniques, etc. Ils peuvent manquer d'informations importantes qui peuvent avoir un impact sur les prix des actifs. C’est là que l'ajout de données de médias sociaux peut être bénéfique pour améliorer la précision des modèles de prédiction. Dans la revue (Jaquart & Dann & Weinhardt), le sentiment de marché est un bon indicateur pour entrainer son modèle. Les données de médias sociaux peuvent fournir des informations précieuses sur le sentiment du marché, les opinions des investisseurs et les tendances émergentes. En incorporant ces données, les modèles d'apprentissage machine peuvent capturer des tendances qui peuvent ne pas être apparentes dans les données historiques et améliorer la précision de leurs prédictions.

Par exemple, l'analyse des hashtags sur Twitter peut aider à identifier les tendances émergentes et les sujets de tendances qui peuvent influencer les prix. Dans l’article de (M & Nguyen & Hamdi & Cengiz), les modèles avec les données des réseaux sociaux est indispensable pour une meilleure prédiction mais posse les limites de leur donnée trop faible ou difficile d’accès. L'analyse des commentaires sur Reddit peut aider à comprendre les opinions des investisseurs sur une cryptomonnaie particulière et son potentiel à long terme. Dans la revue de (Marne & Correia & Churi & Gomes), fait une ouverture sur la nécessité d’intégrer les données de Twitter. Cette revue est à mettre en lien avec l’analyse de (Das & Behera & Kumar & Rath) qui met en évidence que dans le marché action, les données de twitter donnent des opportunités d’améliorer l’apprentissage des modèles. En incorporant ces données de médias sociaux dans les modèles de prédiction, les analystes peuvent obtenir une image plus complète de la situation du marché et améliorer leur capacité à prédire les tendances futures des prix des cryptos monnaies. Cette manière de communiquer et d’échanger sur des plateformes telle que Twitter est une nouveauté par rapport à l’actif classique ou la majeure partie des informations est échangé dans la presse spécialisé. Dans la revue de (Keyan & Jianan & Dayong) les recherches sont portées sur différents modèles comme LSTM, GRU, RNN et les réseaux sociaux. Il indique que couplé le modèle LSTM avec les réseaux sociaux augmentent considérablement la prédiction.

L'ajout de données de médias sociaux aux modèles d'apprentissage peut améliorer considérablement leur précision et leur capacité à prédire les tendances futures des prix des actifs. Les analystes doivent être conscients des limites des modèles de prédiction et utiliser des techniques supplémentaires pour obtenir des résultats fiables et précis. En effet, les critiques académiques sur la prédiction des prix des cryptomonnaies soulèvent plusieurs points importants. Tout d'abord, il est vrai que de nombreux modèles de prédiction des prix des cryptomonnaies se basent sur des données historiques qui peuvent être obsolètes ou incomplètes, et que l'utilisation d'apprentissage peut ne pas être suffisante pour capturer tous les facteurs externes qui influencent les prix.

En outre, la volatilité des prix des cryptomonnaies peut rendre certains modèles de prédiction inefficaces, car les prix peuvent fluctuer rapidement et de manière imprévisible. Les modèles de prédiction doivent donc être en mesure de prendre en compte cette volatilité. La revue de (Rezaei & Faaljou & Mansourfar) proposent des modèles LSTM hybride pour baisser la volatilité dans les actions, le modèle CEEM est le plus performant. L’une des limites à son modèle dans la revue de (Marne & Correia & Churi & Gomes) est la volatilité élevée de cet actif qui compliqué l’apprentissage et la prédiction.

Les modèles de prédiction doivent également être adaptés aux besoins et aux objectifs de chaque étude, de nombreuse étude travaillent sur des périodes allant de quelques minutes à la journée et n’ont pas les mêmes résultats sur la pertinence des modèles. L’étude de (Jaquart & Dann & Weinhardt), ont travaillé sur des périodes de temps de 1 à 60 minutes, et les résultats sont probants. D’autres littératures la période de temps se concentre sur des périodes journalières. On note que dans l’article de (Ibrahim & Kashef & Corrigan) travaille sur une période de temps de 5 minutes.

Les investisseurs doivent être conscients des limites des modèles de prédiction et tenir compte des facteurs externes pour prendre des décisions d'investissement éclairées. Il est important de souligner que la prédiction des prix des cryptomonnaies est un processus en constante évolution en raison de la nature changeante des marchés financiers et des cryptomonnaies. Les modèles de prédiction doivent donc être continuellement évalués et ajustés pour être efficaces et fiables. On note que dans la revue de (Liu & Li & Zhu & Yao), les chercheurs alimentent avec 40 facteurs différents et prouvent que plus l’on met de paramètre plus le modèle est performant.

La prédiction des prix des cryptomonnaies est un domaine en évolution constante et qui nécessite des recherches et des analyses approfondies pour améliorer la précision des modèles de prédiction. Les investisseurs doivent être conscients des limites des modèles de prédiction et prendre en compte les facteurs externes pour prendre des décisions d'investissement éclairées. Les avancées futures dans les modèles de prédiction peuvent aider les investisseurs à mieux comprendre les tendances et les fluctuations du marché des cryptomonnaies et prendre des décisions éclairées en matière d'investissement. Dans la revue de (Nagula & Alexakis), les auteurs mettent en avant les vraies potentialités dans le futur avec une régulation qui permettrait d’encadrer les activités de cette industrie ainsi que les forces en présence.

Alors que le marché des cryptomonnaies continue d'évoluer et de mûrir, il est essentiel que les investisseurs restent informés et conscients des dernières tendances et évolutions de l'industrie. Investir dans les cryptomonnaies peut être une proposition à haut risque et à haut rendement, mais avec les bonnes informations et outils, les investisseurs peuvent prendre des décisions avancées et profiter des opportunités présentées par cette classe d'actifs excitante et nouvelle.

De plus, les investisseurs doivent être conscients des limites des modèles de prédiction des prix des cryptomonnaies. Bien que ces modèles puissent fournir des informations utiles, ils ne sont pas infaillibles et ne doivent pas être utilisés comme la seule source d'informations pour prendre des décisions d'investissement. Dans la revue de Oyedel & Ajayi & Oyedel & Bello & Jimoh), indique de compléter l’analyse par une analyse graphique ou d’autres modèles d’analyse permettant d’affiner les résultats de sortie

Les revues académiques sur la prédiction des prix des crypto-monnaies mettent en évidence le défi majeur du manque de données conséquentes et fiables pour entraîner efficacement les modèles. Les ensembles de données disponibles sont souvent limités en taille et en qualité, ce qui limite la capacité des chercheurs à développer des modèles prédictifs précis.

La nature relativement récente des crypto-monnaies est une raison majeure de cette limitation. Les données historiques couvrent une période relativement courte, ce qui se traduit par un nombre limité d'observations pour entraîner les modèles. De plus, la collecte de données sur les crypto-monnaies peut être compliquée en raison de la fragmentation des échanges et de la disponibilité limitée de données publiques.

Pour surmonter ces défis, les chercheurs explorent différentes approches. Certains utilisent des techniques de sous-échantillonnage ou de sur-échantillonnage pour équilibrer les ensembles de données, tandis que d'autres explorent des techniques de transfert d'apprentissage pour utiliser des modèles pré-entraînés sur d'autres ensembles de données plus vastes.

De plus, les chercheurs tentent de combiner plusieurs sources de données pour créer des ensembles de données plus riches et diversifiés. Cela peut inclure l'intégration de données provenant de différents échanges, l'agrégation de données provenant de plusieurs crypto-monnaies ou l'utilisation de données de marché plus générales.

En conclusion, les recherches académiques autour de ce sujet sont relativement nombreuses et permettent d’avoir une bonne base de travail. La majeure partie des travaux se situe avant 2018, avec des données externes très limités, ils basent généralement uniquement sur le prix. Ils mettent en évidence la composante de la volatilité qui dans certains cas rend compliquer la prédiction des prix par les modèles. La recherche autour de ces modèles m’a passionné au point d’en faire mon sujet qui s’articulera autour de la prédiction du prix du Bitcoin avec un modèle de deep learning, le LSTM, avec des données externes sur une période journalière après 2018.

# Méthodologie

## Introduction

### Contexte de la recherche

Le marché des cryptomonnaies, en particulier celui du Bitcoin, suscite un intérêt croissant en raison son potentiel de rendement élevé. Dans ce contexte, la prédiction des prix du Bitcoin est devenue un domaine de recherche important. Les avancées récentes dans le domaine du deep learning ont permis d'obtenir des résultats prometteurs dans la prédiction de séries temporelles, ce qui ouvre de nouvelles perspectives pour la prédiction des prix du Bitcoin.

Cependant, la prédiction précise des prix du Bitcoin reste un défi en raison de la complexité de ce marché, qui est influencé par de nombreux facteurs économiques, financiers, technologique et sociaux. Pour améliorer les prédictions, il est donc nécessaire d'intégrer des données externes pertinentes dans le modèle LSTM.

### Objectifs de la recherche

L'objectif de cette recherche est d'explorer l'utilisation d’un modèle LSTM (Long Short-Term Memory), pour prédire les prix de clôture journaliers du Bitcoin en intégrant des données externes. Plus précisément, nous visons à :

1. Évaluer l'efficacité du modèle LSTM dans la prédiction des prix du Bitcoin.
2. Intégrer des données externes telles que des indicateurs sur le prix, la volatilité, la variation ou le sentiment de marché pour améliorer les performances de prédiction du modèle LSTM.
3. Analyser les variables externes qui ont le plus d'influence sur les prédictions du modèle LSTM et évaluer leur pertinence pour la prédiction des prix du Bitcoin.

### Rappel des principaux concepts : modèle LSTM, prédiction des prix du Bitcoin, intégration de données externes

Dans cette partie introductive, nous rappellerons brièvement les principaux concepts liés à notre recherche. Le deep learning est une branche de l'apprentissage automatique (machine learning) qui utilise des réseaux de neurones artificiels profonds pour extraire des caractéristiques complexes à partir des données et effectuer des tâches de prédiction ou de classification.

Le modèle LSTM est une architecture récurrente de réseau de neurones qui est capable de mémoriser des informations à long terme, et dispose de librairie Python permettant un accès facile ce qui en fait un choix populaire pour la prédiction de séries temporelles.

La prédiction des prix du Bitcoin est une tâche complexe en raison de la nature volatile et non linéaire de cette cryptomonnaie. Les modèles de prédiction doivent prendre en compte à la fois les tendances historiques des prix et les facteurs externes qui peuvent influencer ces prix.

L'intégration de données externes dans les modèles de prédiction vise à incorporer des informations supplémentaires provenant de sources variées pour améliorer la précision des prédictions. Cela permet de tenir compte des certaines informations disponible que détienne les investisseurs au moment d’intervenir sur les marchés qui peuvent avoir un impact sur les prix du Bitcoin.

Cette recherche vise à utiliser le modèle LSTM et à intégrer des données externes pour améliorer la prédiction des prix journaliers du Bitcoin. Les résultats de cette étude pourraient contribuer à une meilleure compréhension des mécanismes.

## Orientation méthodologique

### Type de méthode utilisée

#### Pertinence de l'approche quantitative pour la prédiction des prix du Bitcoin

Dans le contexte de la prédiction des prix du Bitcoin, l'approche quantitative se révèle pertinente et adaptée. En effet, les données relatives aux prix du Bitcoin sont essentiellement de nature quantitative, consistant en des séries chronologiques de valeurs numériques représentant les prix à différents moments dans le temps. Les méthodes quantitatives permettent d'analyser ces données chiffrées, d'extraire des modèles mathématiques et statistiques, et d'effectuer des prédictions basées sur ces modèles.

La prédiction des prix du Bitcoin repose sur des caractéristiques numériques, telles que les prix passés, les volumes de transactions, la volatilité, et des informations sur le prix. Les méthodes quantitatives offrent des outils appropriés pour modéliser ces données, identifier des tendances et des motifs, et construire des modèles de prédiction précis.

#### Avantages des méthodes quantitatives en deep learning

Le recours aux méthodes quantitatives, en particulier le deep learning, présente plusieurs avantages pour la prédiction des prix du Bitcoin :

* Capacité à capturer des relations complexes : Le deep learning permet de modéliser des relations non linéaires entre les variables, ce qui est particulièrement important pour la prédiction des prix du Bitcoin qui peut être influencée par une multitude de facteurs complexes et interconnectés.
* Apprentissage à partir de grandes quantités de données : Les modèles de deep learning, tels que le modèle LSTM, sont capables d'apprendre à partir de vastes ensembles de données historiques, ce qui permet de mieux saisir les schémas et les dynamiques temporelles du marché du Bitcoin.
* Généralisation et adaptabilité : Les modèles de deep learning ont la capacité de généraliser à partir des données d'entraînement et d'adapter leurs prédictions aux nouvelles informations, ce qui est crucial compte tenu de la volatilité et de l'évolution rapide du marché du Bitcoin.
* Automatisation du processus de prédiction : Les méthodes quantitatives basées sur le deep learning permettent d'automatiser le processus de prédiction, réduisant ainsi la dépendance à l'égard des jugements subjectifs et des biais humains.

L’approche quantitative, en utilisant le modèle LSTM, est justifiée pour la prédiction des prix du Bitcoin en raison de sa pertinence pour les données quantitatives et des avantages qu'elle offre en termes de modélisation des relations complexes, d'apprentissage à partir de grandes quantités de données et d'adaptabilité aux conditions changeantes du marché.

## Terrain de recherche

### Présentation du cadre de recherche

#### Choix de la plateforme d'échange Binance et l'utilisation des données de investing.com

Dans notre recherche sur la prédiction des prix journaliers du Bitcoin et de l’Ethereum, nous avons choisi d'utiliser les prix du Bitcoin provenant de la plateforme d'échange Binance, qui sont répertoriés sur investing.com. Binance est l'une des plus grandes et des plus réputées plateformes d'échange de cryptomonnaies au monde, offrant une liquidité élevée et un volume important de transactions pour le Bitcoin.

Pour collecter les données nécessaires à notre recherche, nous avons utilisé investing.com en raison de sa disponibilité d'accès aux données historiques des prix du Bitcoin sur l’exchange Binance. Nous avons extrait les données de prix du Bitcoin de la plateforme Binance via investing.com, lesquelles étaient accessibles et téléchargeables sous format CSV.

#### Justification de l'utilisation des données de Binance via investing.com

Les raisons de l'utilisation des données provenant de Binance via investing.com sont les suivantes:

* + Disponibilité des données historiques : investing.com fournit un accès pratique aux données historiques des prix du Bitcoin sur une période significative, ce qui permet d'analyser les tendances passées et de construire des modèles prédictifs basés sur des données fiables et complètes.
  + Réputation et fiabilité des données de Binance : Binance est une plateforme d'échange de confiance, et les données de prix du Bitcoin qu'elle fournit sont considérées comme précises et vérifiables. En utilisant les données de Binance via investing.com, nous nous assurons de disposer de données fiables et représentatives du marché réel du Bitcoin.
  + Extraction facile : investing.com permet de télécharger les données historiques des prix du Bitcoin sous format CSV, ce qui facilite l'extraction et le traitement des données pour l'entraînement et l'analyse du modèle LSTM.

En utilisant les prix du Bitcoin de Binance disponibles sur investing.com, nous avons accès à des données historiques fiables et cohérentes, provenant d'une plateforme d'échange réputée. Ces données nous permettent de développer des modèles de prédiction précis et d'évaluer l'efficacité du modèle LSTM dans la prédiction des prix journaliers du Bitcoin en intégrant des données externes.

Nous avons sélectionné les prix du Bitcoin de la plateforme d'échange Binance, répertoriés sur investing.com, et avons extrait ces données au format CSV pour notre recherche. Cette approche nous permet d'utiliser des données fiables et représentatives du marché réel du Bitcoin, facilitant ainsi l'entraînement et l'analyse du modèle LSTM pour la prédiction des prix journaliers.

## Justification du choix de la période d'étude

### Raisons pour sélectionner une période spécifique

Nous avons choisi de focaliser notre étude sur la période allant de février 2018 à septembre 2022 pour plusieurs raisons :

* + Couverture d'une période significative : La période sélectionnée couvre une période de près de quatre ans, ce qui offre une base de données historiques suffisamment étendue pour analyser les tendances et les schémas du marché du Bitcoin. Cette durée permet d'obtenir des résultats plus robustes et d'évaluer l'efficacité du modèle LSTM sur une période significative.
  + Inclusion de l'événement de 2018 : La période choisie inclut l'impact de l'événement majeur de 2018, où le prix du Bitcoin a connu une chute importante après avoir atteint des sommets historiques. L'inclusion de cet événement permet d'évaluer la capacité du modèle LSTM à détecter et à prédire les fluctuations brusques du marché.
  + Prise en compte de la pandémie de COVID-19 : La période choisie comprend également l'impact de la pandémie de COVID-19, qui a eu des répercussions significatives sur l'économie mondiale et les marchés financiers. Étant donné que le marché du Bitcoin est influencé par les événements économiques mondiaux, il est important d'évaluer la capacité du modèle LSTM à prédire les prix du Bitcoin pendant cette période de perturbations.

### Explication de l'importance de cette période pour la recherche

La période allant de février 2018 à septembre 2022 revêt une importance particulière pour notre recherche pour les raisons suivantes :

* Pertinence pour les revues académiques : La sélection de cette période s'aligne avec la décision de nombreux revues académiques de ne pas inclure de données postérieures à 2018. En respectant cette limitation, nous nous assurons de pouvoir comparer nos résultats avec d'autres travaux de recherche existants et de contribuer à la littérature académique dans ce domaine spécifique.
* Analyse des tendances à long terme : En couvrant une période de près de quatre ans, notre étude permet d'analyser les tendances à long terme du marché du Bitcoin. Cela inclut des cycles de hausse et de baisse, des périodes de consolidation, ainsi que des événements économiques et géopolitiques qui ont influencé le marché. Une analyse sur une période plus étendue permet d'obtenir une perspective plus complète des variations de prix et d'évaluer la capacité du modèle LSTM à prévoir ces mouvements.
* Évaluation de la robustesse du modèle LSTM : En incluant des périodes de volatilité et d'incertitude, notre choix de période permet d'évaluer la robustesse du modèle LSTM dans des conditions de marché variées. Cela nous permet de déterminer si le modèle est capable de s'adapter à différents contextes et de fournir des prédictions précises même en présence de fortes fluctuations de prix.

La sélection de la période allant de février 2018 à septembre 2022 pour notre étude sur la prédiction des prix journaliers du Bitcoin est justifiée pour plusieurs raisons. Tout d'abord, cette période couvre une durée significative, ce qui nous permet d'analyser les tendances et les schémas du marché du Bitcoin sur une période plus longue. Cela nous donne une perspective plus complète et nous permet d'évaluer l'efficacité du modèle LSTM dans des conditions de marché variées.

L'inclusion de l'événement de 2018 et de la pandémie de COVID-19 est d'une importance cruciale. L'événement de 2018, où le prix du Bitcoin a connu une chute importante après avoir atteint des sommets historiques, permet d'évaluer la capacité du modèle LSTM à détecter et à prédire les fluctuations brusques du marché. De même, la période de la pandémie de COVID-19 a entraîné des perturbations économiques mondiales, ce qui peut avoir un impact significatif sur le marché du Bitcoin. Il est donc essentiel de comprendre comment le modèle LSTM se comporte pendant ces périodes de perturbations. Notre choix de période est en accord avec les limites temporelles souvent établies dans les revues académiques, ce qui nous permet de comparer nos résultats avec d'autres travaux de recherche existants et de contribuer à la littérature académique dans ce domaine spécifique.

## Collecte des données

### Échantillonnage des données

#### Méthode d'échantillonnage utilisée

Dans notre étude, nous avons utilisé des données déjà disponibles sur les prix journaliers du Bitcoin, ainsi que d'autres variables telles que le sentiment, le prix de l'Ethereum (ETH), le volume de transactions, les plus hauts et les plus bas du Bitcoin la variation du BTC et la volatilité du BTC en journalier. Ces données ont été collectées chaque jour car le marché du bitcoin est ouvert 24H/24 et 7J/7.

#### Justification du choix de la méthode d'échantillonnage

Dans le contexte de notre recherche, nous avons utilisé des données historiques :

* + Disponibilité des données : Les données que nous avons utilisées étaient déjà disponibles et accessibles, notamment les prix du Bitcoin et les autres variables associées.
  + Utilisation de données historiques : Dans notre recherche, nous nous sommes intéressés à l'analyse des tendances et à la prédiction des prix du Bitcoin en utilisant l’approche LSTM. Par conséquent, l'utilisation de données historiques existantes était pertinente pour notre étude, car elle nous permettait d'analyser les comportements passés du marché et de développer des modèles prédictifs.
  + Facilité d'accès aux données : Les données historiques sur les prix du Bitcoin, ainsi que d'autres variables associées, sont souvent disponibles auprès de diverses sources en ligne, telles que des plateformes d'échange de cryptomonnaies ou des agrégateurs de données financières. L'utilisation de ces données préexistantes nous a permis de collecter rapidement et facilement les informations nécessaires à notre recherche.

Il est important de noter que, bien que nous ayons utilisé des données historiques existantes, nous avons veillé à choisir une période de temps spécifique et à respecter des critères de qualité et de fiabilité des données. Cela nous permet d'obtenir des résultats cohérents et significatifs dans notre étude sur la prédiction des prix journaliers du Bitcoin.

### Modes de collecte des données

#### Sources de données externes utilisées

Pour enrichir notre modèle de prédiction du prix du Bitcoin, nous avons utilisé diverses sources de données externes afin d'avoir une vue plus large et plus complète du marché. Ces sources de données nous ont permis d'intégrer plusieurs variables dans notre modèle, augmentant ainsi la complexité et la pertinence de notre analyse. Voici les sources de données externes que nous avons utilisées :

* Sentiment de marché Crypto Fear and Greed Index: Cette source fournit des données sur le sentiment du marché, évalué sur une échelle allant de "Extreme Fear" à "Extreme Greed". Cet indice est calculé en prenant en compte divers facteurs tels que la volatilité du marché, le volume des transactions, le momentum du marché, les enquêtes sur le sentiment des investisseurs, les tendances sur les réseaux sociaux, et la domination du Bitcoin sur le marché. Ces données nous permettent d'incorporer le sentiment général du marché dans notre modèle, ce qui peut avoir un impact significatif sur les prix.
* Données "Close" :Ces données, généralement obtenues à partir de sources financières comme Yahoo Finance ou d'autres plateformes d'échange de cryptomonnaies, représentent le prix de clôture du Bitcoin pour chaque jour. Cette variable est un élément essentiel de notre modèle, car elle fournit la valeur du Bitcoin à la fin de chaque journée de trading. C’est la données que l’on souhaite prédire.
* Données ETH (Ethereum) : Ces données, qui représentent le prix de l'Ethereum, une autre cryptomonnaie populaire, sont également intégrées dans notre modèle. Nous supposons que les mouvements de prix des autres cryptomonnaies majeures peuvent influencer le prix du Bitcoin, d'où l'inclusion de ces données.
* Données de Volume (Vol) : Ces données représentent le volume total de Bitcoin échangé pendant une journée donnée. Une augmentation du volume peut indiquer une hausse de l'activité du marché, qui peut à son tour influencer le prix du Bitcoin.
* Données "Haut" et "Bas" : Ces données représentent respectivement les prix les plus élevés et les plus bas atteints par le Bitcoin dans une journée donnée. Ces valeurs peuvent fournir des informations sur la volatilité du prix du Bitcoin dans une journée.
* Données de Variation : Ces données représentent la différence entre le prix de clôture et le prix d'ouverture du Bitcoin pour une journée donnée. Elles fournissent des informations sur la direction et l'ampleur des mouvements de prix du Bitcoin dans une journée.
* Données de Volatilité : Ces données, généralement calculées à partir des prix haut, bas et de clôture, fournissent une mesure de l'incertitude ou du risque associé au prix du Bitcoin. Une Volatilité élevée peut indiquer des fluctuations de prix importantes, tandis qu'une volatilité plus faible peut indiquer une stabilité relative.

### Justification de l'adéquation des données collectées pour l'entraînement du modèle LSTM

#### Cohérence des données externes avec la prédiction des prix du Bitcoin

Les données externes que nous avons collectées sont cohérentes avec la prédiction des prix du Bitcoin. L'inclusion de ces données externes dans notre modèle LSTM permet de prendre en compte des facteurs supplémentaires qui peuvent influencer les prix du Bitcoin.

Le sentiment reflète les opinions et les émotions des utilisateurs à l'égard du Bitcoin. Il peut s'agir de l'opinion générale des investisseurs, du sentiment des médias sociaux ou d'autres indicateurs de confiance ou de pessimisme. En intégrant le sentiment dans notre modèle, nous tenons compte de la dimension sociale et psychologique qui peut influencer la demande et l'offre de Bitcoin. Nous avons récupéré les données du crypto fear and greed index qui est la référence en termes de sentiment de marché dans cette industrie.

La volatilité du Bitcoin est un indicateur de la rapidité et de l'ampleur des variations de prix. Elle est souvent utilisée pour mesurer le niveau de risque associé au Bitcoin. En incluant la volatilité dans notre modèle, nous prenons en compte la nature volatile du Bitcoin, ce qui peut aider à mieux prédire les mouvements de prix et à évaluer les risques associés. Pour la volatilité du Bitcoin, plusieurs possibilités s’offraient, l’index CVI qui est un indice de la finance décentralisé qui est un indicateur de volatilité des crypto-monnaies. Nous pouvions également opter le VIX mais est plus représentatif du marché traditionnel. Les données de l’index CVI n’étant pas récupérable. Nous avons opté pour réaliser notre propre calcul de la volatilité du Bitcoin comme suivant :

Volatilité = √(σ²)

Avec : l'écart type (σ)

Les données sur le #BTC ou #Bitcoin provenant de Twitter étaient une approche que je souhaitais apporter à mon modèle et auraient pu nous permettre d’enrichir notre modèle avec des données justifiant l’évolution du prix du Bitcoin. Cependant, l’utilisation de ces données s’est avérée être une option extrêmement longue ou coûteuse. Après avoir récupéré l’API Twitter, nous nous sommes rendu compte que le nombre de requêtes était limité par jour le nombre de # sur ce sujet peut facilement représenter plus de 400 000 # par jours) ou payant, ce qui aurait entraîné des coûts significatifs, voire des milliers d’euros. En raison de ces contraintes, l’intégration des données Twitter n’a pas été réalisable pour notre étude. Toutefois, il convient de souligner que l’incorporation des données Twitter reste un axe de développement intéressant à explorer pour de futures recherches.

#### Évaluation de la qualité et de la fiabilité des données collectées

Pour assurer la qualité et la fiabilité des données collectées, nous avons suivi les étapes suivantes :

* + Sélection de sources de données réputées : Nous avons choisi des sources de données externes fiables, telles que des agrégateurs de données financières ou des sources spécialisées dans l’analyse du sentiment. Ces sources jouissent d’une solide réputation et fournissent des données de qualité.
  + Vérification de la cohérence et de l’intégrité des données : Nous avons effectué des contrôles de cohérence pour nous assurer que les données étaient complètes, précises et exemptes d’erreurs. Nous avons vérifié les formats, les unités de mesure, et nous nous sommes assurés que les données étaient conformes à nos attentes.
  + Validation croisée des données : Nous avons comparé les données collectées avec d’autres sources fiables pour valider leur exactitude et leur cohérence. Nous avons vérifié les tendances, les valeurs extrêmes et les variations pour nous assurer que les données correspondent aux attentes du marché.
  + Prise en compte de la période d’étude : Nous avons évalué la pertinence de la période d’étude sélectionnée (février 2018 à septembre 2022) pour notre recherche. Nous avons vérifié que cette période couvrait des événements et des tendances significatives dans le marché du Bitcoin, ce qui renforce la pertinence des données collectées.

En évaluant la qualité et la fiabilité des données collectées, nous nous assurons que les informations externes que nous utilisons pour l’entraînement de notre modèle LSTM sont valides et représentatives de la réalité du marché du Bitcoin. Cela renforce la crédibilité de notre étude et la robustesse de notre modèle de prédiction.

## Traitement et analyse des données

### Prétraitement des données

Le prétraitement des données est une étape cruciale pour garantir la qualité et la pertinence des données utilisées dans notre modèle LSTM pour la prédiction des prix journaliers du Bitcoin. Dans cette partie, nous détaillons les étapes de prétraitement que nous avons appliquées à nos données.

#### Normalisation des données

La normalisation des données est une étape importante pour mettre toutes les variables sur une même échelle, ce qui facilite l’apprentissage du modèle LSTM. Nous avons utilisé la technique de la mise à l’échelle Min-Max avec la fonction MinMaxscaler(feature\_range=(0,1), qui transforme les données en ajustant les valeurs entre une plage spécifique, généralement de 0 à 1. Cette normalisation permet d'éviter que des variables avec des valeurs de grande amplitude ne dominent les autres variables.

### Entraînement du modèle LSTM

L'étape d'entraînement du modèle LSTM est cruciale pour apprendre les relations complexes entre les données d'entrée et les prix journaliers du Bitcoin. Voici les éléments clés de cette étape :

#### Description de la méthode d'entraînement utilisée

Pour entraîner notre modèle LSTM, nous avons utilisé une méthode d'apprentissage supervisé. Nous avons divisé nos données en ensembles d'entraînement, de validation et de test. Voici les étapes clés de la méthode d'entraînement que nous avons suivie :

* + Ensemble d'entraînement : Nous avons utilisé l'ensemble d'entraînement pour ajuster les poids et les biais du modèle LSTM en minimisant une fonction de perte, telle que l'erreur quadratique moyenne (MSE) ou l'erreur absolue moyenne (MAE). Nous avons utilisé l’algorithme d'optimisation Adam, pour ajuster les paramètres du modèle.
  + Ensemble de validation : Nous avons utilisé l'ensemble de validation pour évaluer les performances du modèle à différentes étapes de l'entraînement. Cela nous a permis de surveiller l'ajustement du modèle, de sélectionner les meilleurs hyperparamètres et d'éviter le surapprentissage.
  + Ensemble de test : Enfin, nous avons utilisé l'ensemble de test pour évaluer les performances finales du modèle LSTM. Nous avons fait des prédictions sur cet ensemble de données et comparé les résultats avec les valeurs réelles pour mesurer l'efficacité du modèle.

L'entraînement du modèle LSTM a nécessité plusieurs itérations, ajustements d'hyperparamètres et évaluations pour obtenir les meilleures performances de prédiction possibles.

Lors de l'entraînement du modèle LSTM, nous avons configuré son architecture et réglé ses hyperparamètres. Nous avons utilisé une méthode d'apprentissage supervisé avec des ensembles d'entraînements et de validation pour ajuster les poids et les biais du modèle. Nous avons utilisé l’optimiseur Adam pour minimiser la fonction de perte et mettre à jour les paramètres du modèle lors de chaque itération.

L'ajustement des hyperparamètres a été réalisé en expérimentant différentes valeurs pour des paramètres tels que le taux d'apprentissage, le nombre de couches LSTM, le nombre de neurones par couche.

Après chaque itération d'entraînement, nous avons évalué les performances du modèle LSTM en utilisant des métriques d'évaluation telles que l'erreur quadratique moyenne (MSE), la racine carrée de l'erreur quadratique moyenne (RMSE) ou le coefficient de détermination (R2). Ces métriques nous ont permis de mesurer la précision du modèle dans la prédiction des prix journaliers du Bitcoin.

En interprétant les résultats, nous avons analysé les variables qui ont le plus influencé les prédictions du modèle LSTM. Nous avons examiné l’influence des données et paramètres sur la prédiction du prix par ce modèle.

Enfin, nous avons discuté des limites et des possibilités d'amélioration du modèle LSTM. Nous avons identifié les éventuelles contraintes ou biais liés aux données, ainsi que les aspects qui pourraient être améliorés dans la configuration du modèle ou dans la collecte des données externes.

Le processus de traitement et d'analyse des données a permis de préparer les données pour l'entraînement du modèle LSTM, d'ajuster les hyperparamètres, d'évaluer la performance du modèle et d'interpréter les résultats obtenus. Ces étapes ont été essentielles pour construire un modèle LSTM performant dans la prédiction des prix journaliers du Bitcoin.

# Analyse et Résultats

## Mise en œuvre du modèle LSTM

### Configuration et entraînement du modèle

Notre première étape a été la préparation et la normalisation des données. Nous avons utilisé les données du Bitcoin extraites d'un fichier CSV, qui comprenait des informations telles que la date, le prix de clôture, la variation, le sentiment, la valeur du Ethereum (ETH), le volume (Vol), le prix le plus haut (Haut), le prix le plus bas (Bas) et la volatilité. Chacune de ces données a été normalisée à l'aide du MinMaxScaler pour convertir les valeurs dans un intervalle entre 0 et 1, ce qui facilite l'apprentissage du modèle.

Pour former notre modèle LSTM, nous avons d'abord divisé les données en un ensemble d'entraînement et un ensemble de test. Ensuite, nous avons créé une fonction `create\_dataset` pour transformer nos données en un format approprié pour le modèle LSTM. Cette fonction prend en entrée un ensemble de données et un paramètre `look\_back` qui détermine combien de pas de temps précédents le modèle doit prendre en compte pour faire une prédiction. Dans notre cas, nous avons choisi `look\_back = 100`, ce qui signifie que le modèle utilise les 100 dernières observations pour prédire la prochaine valeur.

Le modèle LSTM que nous avons utilisé est assez simple, avec une seule couche LSTM suivie d'une couche Dense. La couche LSTM a 50 neurones et prend en entrée des données de forme `(look\_back, 8)`, où 8 correspond au nombre de caractéristiques utilisées pour l'entraînement (`Sentiment normalisé`, `Close normalisé`, `Variation normalisée`, `ETH normalisé`, `Vol normalisé`, `Haut normalisé`, `Bas normalisé`, `Volatilite normalisée`). La couche Dense a une seule unité, qui est la sortie du modèle (la prédiction du prix de clôture du Bitcoin). Nous avons utilisé l'erreur quadratique moyenne comme fonction de perte et Adam comme optimiseur.

Nous avons ensuite formé le modèle en utilisant l'ensemble d'entraînement pendant 100 époques. Une époque correspond à une passe complète à travers l'ensemble des données d'entraînement. La taille du lot a été définie à 1, ce qui signifie que le modèle est mis à jour après chaque exemple d'entraînement. Cette configuration peut aider à améliorer la performance du modèle, mais elle peut également rendre l'entraînement plus lent.

Dans l'ensemble, notre approche a été guidée par l'idée que la prédiction du prix du Bitcoin peut être améliorée en intégrant diverses sources d'informations, et pas seulement en se basant sur les prix passés. C'est pourquoi nous avons inclus plusieurs caractéristiques dans notre modèle, comme le sentiment, la variation, le prix du Ethereum, le volume, le prix le plus haut et le plus bas, et la volatilité.

### Importance des paramètres :

Dans cette section, nous allons nous concentrer sur l'importance des paramètres lors de la configuration et de l'entraînement du modèle LSTM. Ces paramètres jouent un rôle crucial dans la performance du modèle et peuvent influencer la précision des prédictions.

Lors de la configuration du modèle LSTM, plusieurs paramètres doivent être pris en compte pour obtenir des performances optimales.

* + Le nombre de couches cachées : Le choix du nombre de couches cachées est essentiel car il détermine la capacité du modèle à capturer les relations temporelles complexes dans les données. Plus le nombre de couches cachées est élevé, plus le modèle peut apprendre des motifs et des dépendances complexes. Cependant, un nombre excessif de couches cachées peut entraîner un surapprentissage.

Dans notre code nous utilisant 50 neurones, la ligne correspondante est la suivante :

model.add(LSTM(50, input\_shape=(look\_back, 8)))

* + Ensembles d'entraînement et de test : La variable **train\_size** est calculée en multipliant la longueur totale du dataframe **df** par 0.8 (80%). Cela permet de déterminer la taille de l'ensemble d'entraînement, qui sera égale à 80% des données. Ensuite, les données sont divisées en deux ensembles : **train\_data** et **test\_data**. La notation **df[:train\_size]** sélectionne les lignes du dataframe **df** depuis le début jusqu'à l'index **train\_size**, ce qui correspond à l'ensemble d'entraînement. La notation **df[train\_size:]** sélectionne les lignes du dataframe à partir de l'index **train\_size** jusqu'à la fin, ce qui correspond à l'ensemble de test.

Dans notre code nous utilisons différents dataframe : les lignes de code correspondantes sont les suivants :

train\_size = int(len(df) \* 0.8)

train\_data, test\_data = df[:train\_size], df[train\_size:]

* + La taille du batch : La taille du batch fait référence au nombre d'exemples d'entraînement utilisés pour mettre à jour les poids du modèle lors de chaque itération. Un batch plus grand peut permettre une convergence plus rapide du modèle, mais il peut également nécessiter plus de mémoire et ralentir le processus d'entraînement. D'autre part, un batch plus petit peut réduire la quantité de bruit dans les mises à jour de poids, mais il peut nécessiter plus d'itérations pour atteindre la convergence.

Dans notre code nous utilisons une taille de 1 pour le batch, nous avons utilisé tout au long des analyses.

model.fit(train\_X, train\_Y, epochs=100, batch\_size=1, verbose=2)

* + Le nombre d'époques : Le nombre d'époques représente le nombre de fois que le modèle parcourt l'ensemble d'entraînement complet. Il est essentiel pour que le modèle puisse apprendre à partir des données et ajuster ses poids de manière itérative. Un nombre insuffisant d'époques peut conduire à un sous-apprentissage, tandis qu'un nombre excessif d'époques peut entraîner un surapprentissage.

Dans notre code nous utilisons 100 epoques nous avons utilisé tout au long des analyses

model.fit(train\_X, train\_Y, epochs=100, batch\_size=1, verbose=2)

* + La fonction de coût : La fonction de coût, également appelée fonction d'erreur, est utilisée pour mesurer la différence entre les prédictions du modèle et les valeurs réelles. Le choix de la fonction de coût dépend du problème de prédiction spécifique et de la nature des données. Par exemple, la perte quadratique moyenne (MSE) est couramment utilisée pour les problèmes de régression, tandis que l'entropie croisée est utilisée pour les problèmes de classification.

Dans notre code la fonction de cout est la suivante :

model.compile(loss='mean\_squared\_error', optimizer='adam')

* + L'optimiseur : L'optimiseur est responsable de la mise à jour des poids du modèle pendant l'entraînement. Différents optimiseurs, tels que Adam, RMSprop ou SGD, peuvent être utilisés. Le choix de l'optimiseur peut influencer la vitesse de convergence et la qualité des résultats obtenus. L’optimiseur utilisé est Adam, qui est un optimiseur populaire en raison de sa capacité à adapter le taux d'apprentissage de manière adaptative en fonction des gradients calculés. Cela permet un ajustement plus précis des poids et une convergence plus rapide du modèle.

Dans notre code la fonction d’optimiseur est la suivante :

model.compile(loss='mean\_squared\_error', optimizer='adam')

### Validation et test du modèle

Après l'entraînement du modèle, nous sommes passés à la phase de test. Pour cela, nous avons utilisé notre ensemble de test pour évaluer la performance du modèle LSTM. Le modèle a généré des prédictions pour le prix de clôture du Bitcoin, qui ont été comparées aux vraies valeurs.

Pour inverser la normalisation appliquée précédemment et obtenir des prédictions dans l'échelle originale des prix, nous avons utilisé la méthode `inverse\_transform` du scaler de la fonction `price`. De cette façon, nous avons pu comparer directement les prédictions du modèle aux vraies valeurs.

Pour évaluer la qualité de ces prédictions, nous avons calculé trois métriques d'erreur :

* + L'erreur absolue moyenne (MAE), qui mesure la différence moyenne en valeur absolue entre les prédictions et les vraies valeurs. C'est une mesure d'erreur facile à comprendre, car elle est dans les mêmes unités que les données.
  + L'erreur quadratique moyenne (RMSE), qui mesure la racine carrée de la différence moyenne au carré entre les prédictions et les vraies valeurs. Cette mesure pénalise plus fortement les erreurs importantes.
  + Le coefficient de détermination (R2), qui mesure la proportion de la variance des vraies valeurs qui est prédite par le modèle. Un R2 de 1 indique que le modèle prédit parfaitement les vraies valeurs, tandis qu'un R2 de 0 indique que le modèle ne prédit pas mieux que simplement prendre la moyenne des vraies valeurs.

Ces mesures nous ont permis d'évaluer objectivement la performance de notre modèle.

De plus, nous avons réalisé des graphiques pour visualiser les prédictions du modèle par rapport aux vraies valeurs. Le premier graphique montre les prédictions et les vraies valeurs en fonction du temps, ce qui permet de voir comment les prédictions suivent les tendances des vraies valeurs. Le deuxième graphique est un nuage de points qui montre la relation entre les prédictions et les vraies valeurs, ce qui permet de voir si les prédictions sont biaisées ou si elles ont tendance à sous-estimer ou à surestimer les vraies valeurs.

### Analyse des résultats

#### Interprétation des performances du modèle

Les résultats obtenus lors des tests du modèle LSTM démontrent de manière convaincante sa performance exceptionnelle dans l'explication de la variance des données. Avec un taux de variance expliquée (R2 score) supérieure à 90% pour toutes les combinaisons de paramètres de données et de plage temporelle, il est clair que ce modèle est robuste et efficace.

Un R2 score supérieur à 90% indique que plus de 90% de la variabilité des données est capturée et expliquée par le modèle LSTM. Ce niveau élevé de précision dans les prédictions est crucial pour de nombreuses tâches, car il témoigne de la capacité du modèle à saisir les relations complexes entre les variables et à fournir des résultats fiables.

Ces résultats solides suggèrent que ce modèle LSTM est capable de généraliser et de bien s'adapter à différentes tailles d'ensemble d'entraînement. La taille d'ensemble d'entraînement de 80% a été identifiée comme la meilleure configuration après avoir tester toutes les dizaines de pourcentage, ce qui souligne sa capacité à tirer le meilleur parti des données disponibles tout en évitant le surapprentissage.

En conclusion, les performances remarquables du modèle LSTM, avec un R2 score dépassant les 90% pour toutes les conditions de test, témoignent de sa puissance et de sa capacité à expliquer la variance des données. Ces résultats sont encourageants pour une utilisation future de ce modèle dans des tâches prédictives, fournissant ainsi des bases solides pour prendre des décisions éclairées.

Nous avons entrepris des tests initiaux pour évaluer les performances de notre modèle en utilisant uniquement le prix du Bitcoin (BTC) comme donnée d'entrée. Pour affiner notre recherche, nous avons exploré différentes configurations en ajustant le nombre de neurones dans le LSTM.

En utilisant un LSTM de 50 neurones, nous avons obtenu une erreur absolue moyenne (MAE) de 1475.04. Cela signifie que, en moyenne, les prédictions de notre modèle diffèrent du prix réel de clôture du BTC d'environ 1475.04 unités. Nous avons également calculé la racine carrée de l'erreur quadratique moyenne (RMSE), qui s'est élevée à 1877.64. Un RMSE plus faible indiquerait une meilleure précision du modèle. De plus, le score R2 de 0.9112 indique que notre modèle explique environ 91.12% de la variance des données du BTC.

En revanche, en utilisant un LSTM de 150 neurones, nous avons observé une légère amélioration des performances. La MAE obtenue était de 1454.52, ce qui indique une réduction de l'écart moyen entre les prédictions et les prix réels de clôture du BTC par rapport au modèle précédent. Le RMSE pour ce modèle était de 1853.45, et le score R2 était de 0.9307, ce qui suggère une amélioration globale des performances.

Ces résultats montrent que l'ajout de neurones dans le LSTM a permis d'améliorer légèrement les performances du modèle, avec une MAE et un RMSE plus bas pour le LSTM de 150 neurones par rapport au LSTM de 50 neurones. De plus, le score R2 plus élevé pour le modèle avec 150 neurones indique une meilleure capacité à expliquer la variance des données du BTC.

Il est important de noter que malgré ces améliorations, il reste encore une marge d'amélioration pour réduire davantage les écarts entre les prédictions et les prix réels de clôture. Nous continuerons à explorer d'autres caractéristiques et configurations afin de maximiser la précision de notre modèle et d'obtenir des prédictions plus précises pour le prix du BTC. Nous pouvons conclure que les paramètres permettent d’aider le modèle à avoir une prédiction plus juste.

Nous avons élargi notre approche en intégrant une donnée supplémentaire, à savoir le sentiment de marché, en plus du prix du Bitcoin (BTC). Cette extension a été appliquée à trois configurations différentes de notre modèle LSTM, utilisant respectivement 50, 150 et 300 neurones.

Pour le LSTM de 50 neurones, nous avons obtenu une erreur absolue moyenne (MAE) de 1498.81 et un RMSE de 1906.32. Cela signifie que, en moyenne, les prédictions de notre modèle diffèrent du prix réel de clôture du BTC d'environ 1498.81 unités. Le score R2 de 0.8897 nous informe que notre modèle explique environ 88.97% de la variance des données du BTC.

En utilisant un LSTM de 150 neurones, nous avons observé une amélioration des performances. La MAE obtenue était de 1459.57, ce qui suggère une réduction de l'écart moyen entre les prédictions et les prix réels de clôture du BTC par rapport au modèle précédent. Le RMSE pour ce modèle était de 1863.65 et le score R2 était de 0.9184.

Nous avons également testé un LSTM de 300 neurones, pour lequel nous avons obtenu une MAE de 1434.49, un RMSE de 1878.36 et un score R2 de 0.9177.

Ces résultats indiquent que l'ajout du sentiment de marché à notre modèle LSTM a conduit à une amélioration progressive des performances. Les modèles avec plus de neurones ont généralement obtenu de meilleurs résultats, avec des MAE, RMSE et scores R2 plus favorables. Cela suggère que le sentiment de marché est une variable pertinente pour améliorer les prédictions du modèle.

Cependant, il est important de noter qu'il subsiste encore des écarts entre les prédictions et les prix réels de clôture du BTC, même avec ces améliorations. Nous continuons à explorer d'autres caractéristiques et ajustements de notre modèle afin de maximiser sa précision et d'obtenir des prédictions encore plus précises pour le prix du BTC, en tenant compte du sentiment de marché.

Nous avons maintenant incorporé la volatilité en plus du prix du Bitcoin (BTC) dans notre modèle LSTM. Cette approche a été testée avec deux configurations différentes : un LSTM de 50 neurones et un LSTM de 150 neurones.

Pour le LSTM de 50 neurones, nous avons obtenu une erreur absolue moyenne (MAE) de 1465.43. Cela signifie que, en moyenne, les prédictions de notre modèle diffèrent du prix réel de clôture du BTC d'environ 1465.43 unités. Le calcul de la racine carrée de l'erreur quadratique moyenne (RMSE) nous a donné une valeur de 1867.37. Un RMSE plus faible indiquerait une meilleure précision du modèle. Le score R2 de 0.9408 indique que notre modèle explique environ 94.08% de la variance des données du BTC.

En ce qui concerne le LSTM de 150 neurones, nous avons observé une amélioration des performances. La MAE obtenue était de 1447.71, ce qui suggère une réduction de l'écart moyen entre les prédictions et les prix réels de clôture du BTC par rapport au modèle précédent. Le RMSE pour ce modèle était de 1841.65, et le score R2 était de 0.9226, ce qui indique une meilleure capacité du modèle à expliquer la variance des données du BTC.

L'ajout de la volatilité à notre modèle LSTM a conduit à une amélioration globale des performances, avec des valeurs de MAE, RMSE et score R2 plus favorables pour le LSTM de 150 neurones par rapport au LSTM de 50 neurones. Cela suggère que la volatilité est une variable pertinente pour améliorer les prédictions du modèle.

Cependant, malgré ces améliorations, il reste encore une marge d'amélioration pour réduire davantage les écarts entre les prédictions et les prix réels de clôture du BTC. Nous poursuivrons nos efforts pour affiner notre modèle en explorant d'autres caractéristiques et ajustements afin d'obtenir des prédictions encore plus précises pour le prix du BTC en tenant compte de la volatilité.

Nous observons que plus nous mettons de neurones plus le modèle prend du temps à prédire allant jusqu’à 40 secondes pour une époque ce qui peut être un inconvénient si nous travaillons sur des périodes de temps plus courts de quelques minutes. La prédiction s’améliore quand nous ajoutons plus de neurones mais n’augmentent pas toujours la variance des données.

Désormais, notre objectif est de maximiser les performances de notre modèle en travaillant avec toutes les données disponibles simultanément. Nous avons inclus les informations telles que "Date", "Sentiment", "Close", "ETH", "volume", "Haut", "Bas", "Variation" et "Volatilité" pour bénéficier d'un ensemble riche en informations. Cela nous permet d'explorer de manière plus approfondie les relations entre ces variables et d'améliorer la précision de nos prédictions. De plus, nous sommes conscients de l'importance de la sélection des paramètres pour optimiser notre modèle LSTM.

Nous avons réalisé un entraînement de notre modèle en utilisant toutes les données disponibles. Les résultats obtenus pour différentes configurations de notre modèle LSTM sont les suivants :

Pour le modèle LSTM avec 150 neurones, nous avons obtenu une MAE de 1501,9, un RMSE de 1911,5 et un score R2 de 0,9946. Ces mesures indiquent que, en moyenne, les prédictions de notre modèle diffèrent du prix réel de clôture du BTC d'environ 1501,9 unités, avec une racine carrée de l'erreur quadratique moyenne de 1911,5. Le score R2 élevé de 0,9946 suggère que notre modèle explique 99,46% de la variance des données, ce qui indique une excellente performance.

Pour le modèle LSTM avec 100 neurones, nous avons obtenu une MAE de 1482,01, un RMSE de 1889,50 et un score R2 de 0,99. Ces résultats montrent une légère amélioration par rapport au modèle précédent, avec une MAE plus basse, ce qui indique des prédictions plus précises avec une différence moyenne de 1482,01 unités par rapport aux prix réels de clôture. Le score R2 de 0,99 indique que notre modèle explique 99% de la variance des données, confirmant sa capacité à capturer les variations du BTC.

Pour le modèle LSTM avec 50 neurones, nous avons obtenu une MAE de 1493,63, un RMSE de 1900,47 et un score R2 de 0,98. Bien que légèrement inférieurs aux modèles précédents, ces résultats témoignent toujours d'une performance solide. La MAE de 1493,63 indique une différence moyenne entre les prédictions et les prix réels de clôture du BTC, tandis que le score R2 de 0,98 indique que notre modèle explique 98% de la variance des données.

En analysant ces résultats, nous constatons que le modèle LSTM avec 150 neurones présente la MAE la plus élevée, indiquant des prédictions légèrement moins précises que les autres modèles. Cependant, tous les modèles ont obtenu des scores R2 élevés, démontrant leur capacité à expliquer une grande partie de la variance des données

## Synthèse des résultats

Les résultats démontrent une performance exceptionnelle du modèle LSTM pour expliquer la variance des données, avec un R2 score supérieur à 90%. Plus le nombre de neurones dans le LSTM est élevé, plus les performances du modèle s'améliorent, comme en témoignent les valeurs de MAE et de RMSE plus faibles, ainsi que les scores R2 plus élevés. Cependant, l'augmentation du nombre de neurones entraîne également un temps de prédiction plus long, ce qui peut être un inconvénient pour des périodes de temps plus courtes.

L'incorporation de données supplémentaires, comme le sentiment de marché et la volatilité, a généralement conduit à une amélioration des performances du modèle. Cependant, il subsiste toujours des écarts entre les prédictions et les prix réels de clôture du BTC, ce qui suggère que d'autres améliorations sont possibles.

Lors de l'entraînement du modèle avec toutes les données disponibles, les performances du modèle étaient encore excellentes, avec un R2 score atteignant jusqu'à 0,9946. Cela indique que le modèle est capable de saisir efficacement les relations complexes entre un ensemble riche de variables.

Le R2 score est une mesure très révélatrice de la performance d'un modèle. Un score supérieur à 90% indique que le modèle est capable d'expliquer une grande partie de la variance des données, ce qui est une indication de sa robustesse. Cependant, il est important de noter que le R2 score n'est pas la seule mesure de performance. Des mesures comme la MAE et le RMSE sont également cruciales pour évaluer à quel point les prédictions du modèle sont proches des valeurs réelles.

L'ajout de neurones au LSTM a amélioré les performances du modèle, bien que l'amélioration n'ait pas toujours été linéaire. Cela peut s'expliquer par le fait que l'ajout de neurones augmente la complexité du modèle, ce qui peut aider à capturer des relations plus complexes dans les données, mais peut aussi conduire à un surapprentissage si le nombre de neurones est trop élevé.

L'intégration de données supplémentaires, comme le sentiment de marché et la volatilité, a amélioré les performances du modèle, ce qui suggère que ces variables sont pertinentes pour la prédiction du prix du BTC. Cependant, il est important de souligner que l'ajout de données supplémentaires augmente la complexité du modèle et peut également augmenter le risque de surapprentissage. Par conséquent, une approche équilibrée est nécessaire pour déterminer quelles données ajouter et comment les intégrer.

L'entraînement du modèle avec toutes les données disponibles a conduit à des performances exceptionnelles, ce qui indique que le modèle est capable d'exploiter efficacement un ensemble riche d'informations. Cependant, il est important de garder à l'esprit que l'entraînement d'un modèle avec un ensemble de données plus large nécessite plus de ressources computationnelles et peut augmenter le temps nécessaire pour entraîner le modèle.

Dans l'ensemble, les résultats démontrent clairement la pertinence du modèle LSTM pour la prédiction du prix du Bitcoin (BTC). Il est particulièrement impressionnant de constater que la variabilité des données est expliquée à plus de 90% dans tous les scénarios, ce qui témoigne de la robustesse du modèle.

De même, l'exploitation de toutes les données disponibles, y compris "Date", "Sentiment", "Close", "ETH", "volume", "Haut", "Bas", "Variation" et "Volatilité", permet une amélioration significative des prédictions. Cela souligne l'importance d'une approche multifactorielle pour prédire les prix du BTC.

Cependant, malgré ces progrès, il existe toujours une marge d'amélioration. Les erreurs absolues moyennes (MAE) et la racine carrée des erreurs quadratiques moyennes (RMSE) restent relativement élevées, indiquant que les prédictions du modèle peuvent encore être affinées.

Alors que l'augmentation du nombre de neurones dans le LSTM conduit à des améliorations, cela se fait au détriment du temps de calcul, qui augmente également. Il est donc nécessaire de trouver un équilibre entre précision et efficacité.

Pour l'avenir, l’exploration d'autres facteurs susceptibles d'influencer le prix du BTC, tels que des données venues de twitter, des variables macroéconomiques, des indicateurs techniques ou des indicateurs qui permettent de contrecarrer la volatilité. De plus, il pourrait être intéressant d'expérimenter d'autres types de réseaux de neurones, tels que les réseaux de neurones convolutifs (CNN) ou les réseaux de neurones récurrents (RNN), pour voir s'ils peuvent offrir de meilleures performances.

Il serait utile d'analyser plus en détail pourquoi certaines prédictions sont plus éloignées de la réalité que d'autres. Cela pourrait aider à comprendre quels aspects du modèle ou des données doivent être améliorés pour obtenir des prédictions plus précises.

En conclusion, bien que des progrès importants aient été réalisés dans la prédiction des prix du BTC à l'aide de modèles LSTM, des travaux supplémentaires sont nécessaires pour améliorer encore la précision de ces prédictions. Cependant, les résultats obtenus jusqu'à présent sont prometteurs et offrent une base solide sur laquelle continuer à construire.

#### Limitations et perspectives d’amélioration

Bien que notre modèle LSTM ait démontré une certaine capacité à prédire les prix journaliers du Bitcoin, il présente plusieurs limitations qui pourraient être abordées dans des travaux futurs.

Premièrement, notre modèle est entièrement basé sur des informations historiques et ne peut donc pas prendre en compte des événements futurs imprévus qui pourraient affecter les prix du Bitcoin. Par exemple, des changements dans la réglementation, des développements technologiques ou des crises économiques pourraient avoir un impact majeur sur les prix du Bitcoin et ne sont pas pris en compte par notre modèle. Une perspective d'amélioration serait de combiner notre modèle LSTM avec d'autres types de modèles capables de prendre en compte ces informations, comme des modèles basés sur des news ou des analyses fondamentales.

Deuxièmement, notre modèle utilise plusieurs variables externes en plus des données de prix pour faire ses prédictions. Bien que ces variables aient été choisie sur la base d'études précédentes montrant son impact sur les prix des cryptomonnaies, il est possible que d'autres variables puissent également être utiles. Des travaux futurs pourraient explorer l'inclusion d'autres variables externes, comme le volume de transactions, le taux d'intérêt ou les données macroéconomiques.

Troisièmement, notre modèle utilise une fenêtre temporelle fixe (look-back) pour faire ses prédictions. Il est possible que l'utilisation d'une fenêtre temporelle dynamique, qui s'adapte aux conditions du marché, puisse améliorer la précision des prédictions.

Enfin, il convient de noter que notre modèle a été conçu pour prédire les prix journaliers du Bitcoin et doit être adapter pour des prédictions à plus long terme ou à d'autres types de cryptomonnaies. Des travaux futurs pourraient explorer l'adaptation de notre modèle à d'autres types de prédictions ou à d'autres cryptomonnaies.

Malgré les performances encourageantes de notre modèle LSTM, il existe de nombreuses possibilités d'amélioration et de développement futur.

## Importance des données externes

Dans notre modèle, nous avons inclus plusieurs types de données externes. Pour comprendre l'importance relative de ces différentes données, nous avons effectué une analyse de l'importance des caractéristiques. Cette analyse permet de mesurer l'impact de chaque caractéristique sur les prédictions de notre modèle.

#### Données de sentiment

Comme mentionné précédemment, les données de sentiment ont eu un impact significatif sur les prédictions de notre modèle. Les jours où le sentiment était particulièrement positif ou négatif, nous avons observé une correspondance notable avec les mouvements de prix du Bitcoin. Cela suggère que le sentiment du marché est un indicateur important de la dynamique des prix des cryptomonnaies.

#### Prix de l'Ethereum (ETH)

Nous avons également inclus le prix de l'Ethereum comme une caractéristique externe, en raison de la corrélation souvent observée entre les prix des différentes cryptomonnaies. Notre analyse a montré que cette caractéristique a également eu un impact notable sur les prédictions de notre modèle. Nous voulions capturer les mouvements de prix de la deuxième crypto-monnaies du marché malgré que ce soit une relation inverse, c’est-à-dire que l’évolution de prix du Bitcoin est le moteur de tendance et d’intensité. Il serait plus intéressant de l’intégrer si le mémoire traitait de la prédiction du prix de l’Ethereum.

#### Volatilité du Bitcoin

La volatilité du Bitcoin, mesurée par la variation quotidienne du prix, a également été une caractéristique importante dans notre modèle. Les jours de forte volatilité ont souvent été associés à des changements de prix plus importants, ce qui a été capturé par notre modèle. J’ai réalisé un indice représentant la volatilité du Bitcoin, cette métrique pourrait être largement améliorer par un calcul plus précis, ou utiliser l’indice CVI qui est un indice de volatilité du marché des crypto-monnaies dans la finance décentralisé.

#### Plus haut et plus bas du Bitcoin

L'inclusion du plus haut et du plus bas du prix du Bitcoin en tant que caractéristiques externes a montré une importance significative dans nos prédictions. Ces informations permettent de capturer les niveaux extrêmes atteints par le prix du Bitcoin au cours d'une période donnée. Les valeurs élevées du plus haut et les valeurs basses du plus bas peuvent être des points de référence importants pour analyser les tendances du marché et les mouvements de prix futurs.

Nous avons constaté que les prédictions du modèle LSTM étaient influencées par ces caractéristiques. Les variations du plus haut et du plus bas du prix du Bitcoin ont été prises en compte dans la dynamique du modèle, contribuant ainsi à une meilleure compréhension des variations des prix et à des prédictions plus précises.

#### Variation du prix du Bitcoin

La variation du prix du Bitcoin, mesurée par l'écart entre les prix de clôture successifs, a également joué un rôle significatif dans notre modèle. Cette caractéristique permet de capturer les changements de prix à court terme et de fournir des informations sur la volatilité du Bitcoin. Les variations importantes du prix peuvent indiquer des périodes de forte activité sur le marché et des opportunités potentielles de trading.

En incluant la variation du prix du Bitcoin dans notre modèle, nous avons pu prendre en compte les changements rapides et les tendances à court terme du marché. Cela a contribué à améliorer la capacité de prédiction du modèle et à capturer les mouvements de prix plus précisément.

En conclusion, notre analyse a montré que plusieurs types de données externes peuvent avoir un impact significatif sur la précision de la prédiction des prix du Bitcoin. Cependant, l'importance de ces données peut varier en fonction du contexte du marché et d'autres facteurs, ce qui souligne l'importance de l'adaptation et de la mise à jour continuelle de notre modèle pour refléter les conditions changeantes du marché.

# Discussion

## 1.Interprétation des résultats

L'interprétation des résultats de notre recherche se concentre principalement sur l'identification des facteurs qui influencent le prix du Bitcoin. Notre modèle a révélé plusieurs relations significatives.

* Volume des transactions (Vol) : Nous avons constaté une corrélation positive entre le volume des transactions Bitcoin et son prix. Cela signifie que lorsque le nombre de Bitcoin échangés augmente, le prix du Bitcoin a tendance à augmenter également. Cette relation peut être expliquée par la loi fondamentale de l'offre et de la demande : lorsque la demande pour un bien (dans ce cas, le Bitcoin) dépasse son offre, le prix augmente. Il est intéressant de noter que cela indique que le marché du Bitcoin répond aux forces du marché de la même manière que les marchés financiers plus traditionnels.
* Sentiment de marché (Crypto Fear and Greed Index) : Notre étude a révélé une corrélation entre le sentiment du marché, tel que mesuré par l'indice Crypto Fear and Greed, et le prix du Bitcoin. Lorsque l'indice montre une "peur" dominante, les prix ont tendance à baisser. À l'inverse, lorsque l'indice montre une "avidité" dominante, les prix ont tendance à augmenter. Cela peut s'expliquer par le comportement des investisseurs qui sont plus enclins à acheter des Bitcoin lorsque le sentiment général est positif (avidité) et à vendre lorsque le sentiment général est négatif (peur). Cela démontre l'importance des facteurs psychologiques dans la détermination des prix sur le marché des cryptomonnaies.
* Prix de l'Ethereum (ETH) : Nous avons également trouvé une corrélation positive entre le prix de l'Ethereum et celui du Bitcoin. Cela suggère que lorsque le prix de l'Ethereum augmente, le prix du Bitcoin a également tendance à augmenter, et vice versa. Cette relation peut être due à plusieurs facteurs. Par exemple, Bitcoin et Ethereum sont les deux plus grandes cryptomonnaies en termes de capitalisation boursière et ont souvent été perçues comme des actifs numériques de premier plan. Par conséquent, les mouvements de prix dans l'un peuvent influencer les perceptions et les comportements des investisseurs dans l'autre.

Ces résultats fournissent des informations précieuses sur les facteurs qui influencent le prix du Bitcoin. Cependant, il est important de noter que, bien que ces variables aient une corrélation avec le prix du Bitcoin, elles ne garantissent pas nécessairement une prédiction précise du prix futur. Comme pour tout modèle prédictif, il est crucial de comprendre que la corrélation ne signifie pas nécessairement causalité.

## 

## Limitations de l'étude

Bien que notre étude offre des insights significatifs sur les facteurs qui influencent le prix du Bitcoin, il est important de reconnaître ses limitations.

* Complexité du marché des cryptomonnaies : Le marché des cryptomonnaies est extrêmement complexe et volatile, et il est influencé par une multitude de facteurs, y compris des événements mondiaux, des politiques gouvernementales, des innovations technologiques, et plus encore. Bien que notre modèle intègre plusieurs facteurs clés, il ne couvre pas tous les facteurs possibles qui pourraient influencer le prix du Bitcoin.
* Manque de données historiques : Les cryptomonnaies sont relativement nouvelles et manquent de données historiques à long terme par rapport à d'autres classes d'actifs, ce qui limite notre capacité à généraliser nos résultats.
* Corrélation ne signifie pas causalité : Comme mentionné précédemment, bien que notre modèle identifie des corrélations entre certaines variables et le prix du Bitcoin, cela ne signifie pas nécessairement qu'il y a une relation de cause à effet. Par exemple, bien que le prix de l'Ethereum et celui du Bitcoin soient souvent corrélés, d'autres facteurs non inclus dans notre modèle pourraient expliquer cette relation.
* Risque de surajustement : Dans notre quête pour créer un modèle prédictif précis, il y a un risque de surajustement. Cela signifie que notre modèle pourrait être trop spécifique à notre ensemble de données et ne pas bien se généraliser à de nouvelles données.
* Dépendance aux données externes : Notre modèle repose sur des données externes provenant de diverses sources. Si ces sources devaient disparaître ou changer leur méthode de collecte de données, cela pourrait affecter la validité de notre modèle.

Ces limitations ne remettent pas en cause la valeur de notre recherche, mais elles soulignent l'importance de la prudence dans l'interprétation et l'application de nos résultats. Comme toujours, il est recommandé de combiner les résultats de notre modèle avec une analyse approfondie du marché et une connaissance du secteur avant de prendre des décisions d'investissement.

## Implications pour les traders et les investisseurs

Notre étude a plusieurs implications pour les traders et les investisseurs en cryptomonnaies, spécifiquement ceux qui s'intéressent au Bitcoin.

* Compréhension des facteurs de prix : Notre modèle souligne l'importance de divers facteurs, tels que le sentiment du marché, le prix de l'Ethereum, et la volatilité dans la prédiction du prix du Bitcoin. Les traders et les investisseurs peuvent utiliser ces informations pour mieux comprendre les mouvements de prix et affiner leurs stratégies de trading.
* Importance du sentiment du marché : Le sentiment du marché, tel que mesuré par l'indice de peur et de cupidité, joue un rôle crucial dans notre modèle. Cela suggère que les traders et les investisseurs devraient prêter une attention particulière à l'humeur générale du marché lorsqu'ils prennent des décisions d'investissement.
* Corrélation avec l'Ethereum : La corrélation significative entre le prix du Bitcoin et celui de l'Ethereum indiquée par notre modèle pourrait offrir des opportunités pour des stratégies d'arbitrage ou de diversification.
* Préparation pour la volatilité : Notre modèle montre également que la volatilité est un facteur important dans la prédiction du prix du Bitcoin. Les traders et les investisseurs doivent être conscients de ce facteur et préparés à gérer les risques associés à la volatilité du marché des cryptomonnaies.

Il convient de noter que, malgré les précieuses informations que notre modèle peut fournir, il ne garantit pas le succès et il doit être utilisé comme un outil parmi d'autres dans le cadre d'une approche d'investissement bien réfléchie et diversifiée.

## Suggestions pour les recherches futures

Bien que notre étude ait apporté des informations précieuses sur la prédiction des prix journaliers du Bitcoin, il existe plusieurs pistes intéressantes pour les recherches futures :

* Inclusion d'autres indicateurs de marché : Notre modèle a utilisé un ensemble de facteurs spécifiques, mais il existe de nombreux autres indicateurs de marché qui pourraient avoir un impact sur le prix du Bitcoin. Les futures recherches pourraient envisager d'inclure d'autres indicateurs, tels que l'indice de dominance du Bitcoin, l'activité du réseau (par exemple, le nombre de transactions ou le hashrate), ou les données macroéconomiques globales.
* Examiner d'autres cryptomonnaies : Nous nous sommes concentrés sur le Bitcoin dans cette étude, mais il serait intéressant d'examiner si les mêmes facteurs jouent un rôle dans la prédiction des prix d'autres cryptomonnaies.
* Analyse de sous-périodes : Les marchés de cryptomonnaies peuvent être volatils et influencés par divers événements. L'analyse des sous-périodes pourrait aider à comprendre comment différents facteurs affectent le prix du Bitcoin dans différentes conditions de marché.
* Utilisation de modèles plus complexes : Bien que notre modèle basé sur la régression linéaire multiple ait montré une capacité notable à prédire le prix du Bitcoin, les méthodes plus avancées, comme les réseaux de neurones artificiels, les machines à vecteurs de support ou les modèles de forêt aléatoire, pourraient améliorer la précision de la prédiction.
* Implication des médias sociaux : Les médias sociaux jouent un rôle de plus en plus important dans la cryptosphère. L'analyse du sentiment des médias sociaux et son impact sur le prix du Bitcoin pourrait être une piste intéressante pour les futures recherches.

Notre étude ait fourni des aperçus utiles, le domaine de la prédiction des prix de la cryptomonnaie reste vaste et riche en possibilités pour la recherche future.

# Conclusion

## Résumé des résultats clés

Notre analyse et expérimentation avec le modèle LSTM ont produit plusieurs résultats clés concernant la prédiction des prix du Bitcoin. Voici un résumé de ces découvertes :

* Performance du LSTM : Le modèle LSTM a démontré une capacité impressionnante à capter les tendances et les fluctuations du prix du Bitcoin. En raison de sa capacité à "se souvenir" des informations sur une longue période, le LSTM a réussi à prédire avec précision les mouvements de prix, en tenant compte de la volatilité et des tendances complexes du marché du Bitcoin.
* Influence des données externes : L'intégration de données externes a amélioré la performance de notre modèle LSTM. Les indicateurs tels que le sentiment du marché Crypto Fear and Greed, le volume des échanges, les prix d'autres cryptomonnaies (en particulier l'ETH), et d'autres variables de marché se sont avérés précieux pour affiner nos prédictions.
* Volatilité du Bitcoin : Le modèle LSTM a bien réussi à prendre en compte la volatilité inhérente au Bitcoin. Cela suggère que ce type de modèle est particulièrement utile pour analyser des actifs financiers volatils comme les cryptomonnaies.
* Importance de la sélection des variables : Le choix des variables externes à intégrer dans le modèle a joué un rôle crucial dans la précision de la prédiction. Cela souligne l'importance d'une compréhension approfondie du marché des cryptomonnaies pour la construction d'un modèle de prédiction efficace.

Nos résultats indiquent que le LSTM, lorsqu'il est enrichi de données externes pertinentes, peut être un outil puissant pour prédire les prix du Bitcoin. Cependant, comme tout modèle prédictif, il ne peut pas éliminer entièrement l'incertitude et la volatilité inhérentes au marché des cryptomonnaies.

## Réponse à la question de recherche

La question de recherche initiale était de déterminer si un modèle LSTM, enrichi de diverses sources de données externes, pouvait prédire efficacement les fluctuations des prix journaliers du Bitcoin. Notre étude a mis en évidence une réponse positive à cette question.

En appliquant un modèle LSTM aux données, nous avons pu capter les tendances de fond et les fluctuations à court terme du prix du Bitcoin. Le modèle LSTM, en exploitant la capacité de mémoriser les informations sur une longue période, a réussi à appréhender les dynamiques complexes et volatiles du marché des cryptomonnaies.

L'incorporation de diverses sources de données externes a amélioré la performance de notre modèle. En particulier, le sentiment du marché Crypto Fear and Greed, le volume des échanges, le prix d'autres cryptomonnaies comme l'ETH, et divers autres indicateurs de marché ont été des ajouts précieux qui ont permis d'augmenter la précision de nos prédictions.

Il convient de noter que bien que le modèle LSTM ait montré une capacité prometteuse à prédire les prix du Bitcoin, il n'est pas à l'abri des incertitudes inhérentes au marché des cryptomonnaies. Le marché du Bitcoin est fortement influencé par une multitude de facteurs imprévisibles, allant des annonces politiques aux innovations technologiques, qui peuvent tous déclencher des fluctuations de prix importantes et imprévisibles. En conséquence, même avec un modèle de prédiction sophistiqué comme le LSTM, il est impossible de prédire avec précision tous les mouvements de prix du Bitcoin.

L'avenir de l'utilisation des modèles LSTM dans la prédiction des prix du Bitcoin est plein de promesses, mais nous sommes encore dans les premières étapes de ce voyage exploratoire.

Il existe un potentiel considérable pour améliorer les modèles LSTM existants. Les LSTM ont démontré leur habileté à modéliser et à prédire les prix du Bitcoin avec une précision impressionnante. Cependant, nous pourrions explorer des variantes de LSTM, comme les réseaux Gated Recurrent Unit (GRU) ou les LSTM à attention, pour améliorer encore plus la précision de nos prédictions.

Par ailleurs, notre étude a révélé l'importance cruciale de l'intégration de données externes pour améliorer la performance des modèles LSTM. À l'avenir, nous pourrions considérer l'inclusion d'autres types de données, comme les signaux sociaux (tweets, posts de blogs, etc.), les données macroéconomiques ou même les actualités de l'industrie cryptographique. L'inclusion de ces données supplémentaires pourrait affiner davantage nos modèles et améliorer la précision de nos prédictions.

L'avenir pourrait également voir l'émergence de modèles hybrides, combinant des LSTM avec d'autres techniques de machine learning. Des modèles de régression, des arbres de décision ou des méthodes d'ensemble pourraient être utilisés conjointement avec les LSTM pour créer des prédictions plus robustes et précises.

Un autre domaine d'intérêt pour l'avenir est la recherche d'outils et de techniques pour interpréter les prédictions des modèles LSTM. Comprendre pourquoi ces modèles font certaines prédictions reste un défi, et le développement d'outils pour interpréter ces prédictions pourrait être un domaine de recherche précieux.

Enfin, bien que notre étude ait été menée dans un cadre de recherche, il serait intéressant de tester ces modèles LSTM dans un environnement de trading en temps réel. Cela nécessiterait la mise en place d'une infrastructure de données en temps réel et d'un système de backtesting robuste pour évaluer les performances du modèle.

L'accès à des données de qualité est une autre dimension cruciale pour l'avenir de l'utilisation des LSTM dans la prédiction des prix du Bitcoin. Le monde des cryptomonnaies est en constante évolution, avec de nouvelles informations et données disponibles chaque jour. Pour que les modèles LSTM soient efficaces, ils doivent être formés sur des données actuelles et précises. Cela soulève des défis en matière d'acquisition de données, car toutes les sources de données ne sont pas créées égales. L'accès à des données de qualité supérieure, fiables et actualisées est essentiel pour obtenir des prédictions précises.

La qualité des données introduites dans un modèle LSTM est un autre aspect important. Les données incomplètes peuvent entraîner des prédictions inexactes. Cela met en évidence la nécessité de méthodes de nettoyage de données robustes et d'une compréhension approfondie des données utilisées. Par exemple, comprendre les tendances saisonnières dans les données ou être conscient des événements externes qui peuvent influencer les prix du Bitcoin peut être crucial pour interpréter correctement les résultats.

Quant à l'utilisation et à l'interprétation des prédictions des modèles LSTM, il est important de comprendre que ces prédictions sont probabilistes et non déterministes. Cela signifie que, bien qu'elles puissent donner une indication de la direction probable des prix, elles ne garantissent pas un résultat spécifique. Les traders et les investisseurs doivent être prudents lorsqu'ils se basent sur ces prédictions pour prendre des décisions de trading. Il est toujours important d'utiliser une gamme d'outils et d'informations lors de la prise de décisions d'investissement, et de comprendre que même les modèles les plus sophistiqués ne peuvent pas prédire l'avenir avec une certitude absolue.

L’'accès à des données de qualité, leur utilisation adéquate et une interprétation prudente des prédictions sont autant d'éléments clés pour l'avenir de l'utilisation des LSTM dans la prédiction des prix du Bitcoin.

L'avenir de l'application des Réseaux de Neurones à Long Terme (LSTM) dans la prédiction des prix du Bitcoin semble prometteur, mais il est également parsemé de défis intéressants et d'opportunités significatives. L'expansion continue de la technologie, l'augmentation exponentielle de la disponibilité des données et l'approfondissement de notre compréhension des cryptomonnaies suggèrent que les LSTM pourraient jouer un rôle de plus en plus central dans les stratégies d'investissement en Bitcoin.

L'une des clés de cette évolution réside dans la qualité et l'accessibilité des données. Comme le marché des cryptomonnaies est dynamique et en constante évolution, l'accès à des informations précises, à jour et pertinentes est crucial. C'est un défi permanent, car toutes les sources de données ne sont pas équivalentes. Les futures avancées dans l'acquisition de données fiables, actualisées et de haute qualité seront sans doute un élément catalyseur pour améliorer la précision des prédictions LSTM.

La qualité des données fournies aux modèles LSTM est un autre facteur déterminant. Des données bruitées, inexactes ou incomplètes peuvent conduire à des prédictions trompeuses. Cela souligne l'importance d'avoir des procédures robustes de nettoyage et de prétraitement des données, ainsi qu'une compréhension approfondie des spécificités des données utilisées. Par exemple, la prise en compte des tendances saisonnières ou des événements externes susceptibles d'influencer les prix du Bitcoin peut être déterminante pour l'interprétation correcte des prédictions.

En conclusion, l'avenir de l'utilisation des LSTM dans la prédiction des prix du Bitcoin est prometteur, mais il nécessite une attention constante à la qualité des données, à leur traitement et à une interprétation prudente des prédictions. Avec le développement continu de ces domaines, nous sommes impatients de voir comment ces modèles vont continuer à évoluer et à offrir des outils de plus en plus précis et utiles pour les investisseurs et les traders dans le domaine des cryptomonnaies.

# 

# Références

Patrick Jaquart, David Dann, Christof Weinhardt. 8 mars 2021 Short-term bitcoin market prediction via machine learning. Ke AI advancing research evolving science

Ahmed Ibrahim , Rasha Kashef , Liam Corrigan . 6 novemvre 2020. Predicting market movement direction for bitcoin: A comparison of time series modeling methods. Computers and Electrical Engineering 89 (2021) 106905

Poongodi M. a, Tu N. Nguyen b, Mounir Hamdi a, Korhan Cengiz. 27 juillet 2021. Global cryptocurrency trend prediction using social media. Information Processing and Management 58 (2021) 102708

Azeez A. Oyedele a, Anuoluwapo O. Ajayi b, \*, Lukumon O. Oyedele b, Sururah A. Bello b, Kudirat O. Jimoh. 4 novembre 2022. Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction. Expert Systems With Applications 213 (2023) 119233

Chen Wang a, Dehua Shen b, \*, Youwei Li. 8 juillet 2022. Aggregate Investor Attention and Bitcoin Return: The Long Short-term Memory Networks Perspective. Finance Research Letters 49 (2022) 103143

Samiksha Marne, Shweta Churi, Delisa Correia, Joanne Gomes. 2021. Predicting Price of Cryptocurrency - A Deep Learning Approach. International Journal of Engineering Research & Technology

Sushree Das, Ranjan Kumar Behera, Mukesh Kumar, Santanu Kumar Rath. 2018. Real-Time Sentiment Analysis of Twitter Streaming data for Stock. Procedia computer science

Liu Keyan ∗, Zhou Jianan, Dong Dayong. 12 avril 2021. Improving stock price prediction using the long short-term memory model combined with online social networks. Journal of Behavioral and Experimental Finance 30 (2021) 100507

Hadi Rezaei \*, Hamidreza Faaljou , Gholamreza Mansourfar. 14 novembre 2020. Stock price prediction using deep learning and frequency decomposition. Expert Systems With Applications 169 (2021) 114332

Yaohao Peng, Pedro Henrique Melo Albuquerque∗, Jader Martins Camboim de Sá, Ana Julia Akaishi Padula, Mariana Rosa Montenegro. 2 decembre 2017. The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression. Expert Systems With Applications 97 (2018) 177–192

Dennys C.A. Mallqui a, Ricardo A.S. Fernandes. 1 decembre 2018. Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. Applied Soft Computing Journal 75 (2019) 596–606

Steven L. Heston and Nitish R. Sinha. 2016. News versus Sentiment: Predicting Stock Returns from News Stories. Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C.

Rajat Kumar Rathore a, Deepti Mishra a, Pawan Singh Mehra b, \*, Om Pal, AHMAD SOBRI HASHIM d, Azrulhizam Shapi’I, T. Ciano f, Meshal Shutaywi. 199. Mai 2022. Real-world model for bitcoin price prediction. Information Processing and Management 59 (2022) 102968

Dirk F. Gerritsen a, Rick A.C. Lugtigheid a, Thomas Walther. 24 juin 2021. Can Bitcoin Investors Profit from Predictions by Crypto Experts?. Finance Research Letters 46 (2022) 102266

Nasirtafreshi. 16 mars 2022. Forecasting cryptocurrency prices using Recurrent Neural Network and Long Short-term Memory. Data & Knowledge Engineering 139 (2022) 102009

Poongodi M.a,1, Ashutosh Sharmab,∗, Vijayakumar V.c, Vaibhav Bhardwajc, Abhinav Parkash Sharmac, Razi Iqbald, Rajiv Kumare. 25 novembre 2019. Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. Computers and Electrical Engineering 81 (2020) 106527

Svitlana Galeshchuk. 5 aout 2015. Neural networks performance in exchange rate prediction. Neurocomputing 172 (2016) 446–452

Salim Lahmiria, Stelios Bekiros. 19 janvier 2020. Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market. Chaos, Solitons and Fractals 133 (2020) 109641

Syed Abul Basher a, Perry Sadorsky. 5 juin 2022. Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?. Machine Learning with Applications 9 (2022) 100355

Han-Min Kim a, Gee-Woo Bock a, \*, Gunwoong Lee. 23 juin 2021. Predicting Ethereum prices with machine learning based on Blockchain information. Expert Systems With Applications 184 (2021) 115480

Noella Nazareth , Yeruva Venkata Ramana Reddy. 27 janvier 2023. Financial applications of machine learning: A literature review. Expert Systems With Applications 219 (2023) 119640

Pavan Kumar Nagula, Christos Alexakis . 24 aout 2022. A new hybrid machine learning model for predicting the bitcoin (BTC-USD) price. Science direct

# Annexe 1

Une image contenant texte, logiciel, Logiciel multimédia, Logiciel de graphisme

Description générée automatiquement

# Annexe 2

Une image contenant texte, logiciel, Logiciel multimédia, Logiciel de graphisme

Description générée automatiquement

# Annexe 3

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Description générée automatiquement

# Annexe 4

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Description générée automatiquement

# Annexe 4

Une image contenant texte, capture d’écran, logiciel

Description générée automatiquement

# Annexe 5

Une image contenant texte, capture d’écran, logiciel

Description générée automatiquement

# Annexe 6

Une image contenant capture d’écran, logiciel, Logiciel multimédia, texte

Description générée automatiquement

# Annexe 7

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Description générée automatiquement

# Annexe 8

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

Description générée automatiquement

# Annexe 10

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

Description générée automatiquement