

الجمهورية الجزائرية الديمقراطية الشعبية  
RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE  
وزارة التعليم العالي و البحث العلمي  
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE  
جامعة عبد الرحمن ميرة بجاية  
UNIVERSITÉ ABDERRAHMANE MIRA - BÉJAÏA  
FACULTÉ DES SCIENCES EXACTES



MÉMOIRE DE FIN D'ÉTUDE  
EN VUE DE L'OBTENTION DU DIPLÔME DE MASTER  
EN MATHÉMATIQUES APPLIQUÉES.

Présenté par : Hamoudi karima Et Hamma Sara

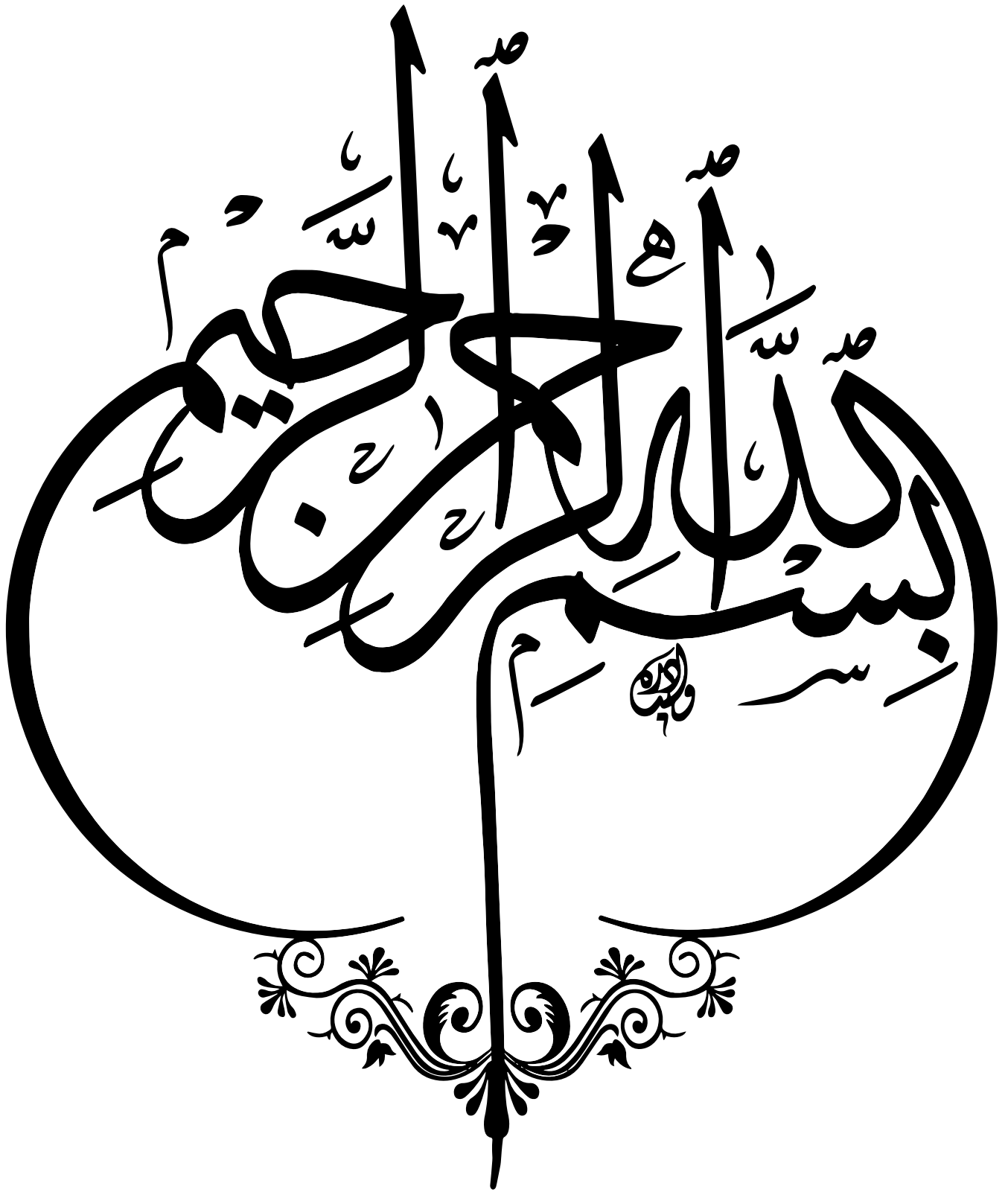
Thème

Machine Learning pour l'estimation de la volatilité

Soutenu le 12 Septembre 2023, devant le jury composé de :

M.	K. Abbas	Pr	Université de Béjaïa	Président
Mme.	B. Takhedmit	MCA	Université de Béjaïa	Rapporteur
Mme.	B. Barache	MCB	Université de Béjaïa	Examinatrice
M.	M. Soufit	MCB	Université de Béjaïa	Examineur

Année Universitaire 2022/2023



# Dedications

À mes chers parents, frères et sœurs,

Sara HAMMA.

À mon très cher père,  
À ma très chère mère,  
À mes frères et sœurs,  
À mon fiancé,  
À ma binôme Sara qui m'a soutenue jusqu'au bout,  
À toute ma famille et mes amies,

Karima Hamoudi.

# Remerciements

Toutes nos gratitude, grâce et remerciements vont à **Dieu le Tout-Puissant** qui nous a donné la force, la patience, le courage et la volonté pour élaborer ce modeste travail.

Nous tenons à remercier notre encadrante, **Mme TAKHDMIT BAYA**, pour son aide et ses précieux conseils durant toute la période de la réalisation de ce travail.

Nous exprimons notre profonde gratitude à **M. Karim Abbas**, le président du jury ainsi qu'aux examinateurs **Mme. B. Barache** et **M. M. Soufit** pour leur dévouement et leur soutien tout au long de ce processus d'examen.

On n'oubliera pas, bien évidemment, nos enseignants qu'on remercie chaleureusement pour tous ces agréables moments passés ensemble et pour l'intérêt qu'ils nous ont porté tout au long de notre cursus universitaire, ainsi que pour leur aide.

***Septembre 2023***

# Contents

<b>Dedications</b>	<b>i</b>
<b>Remerciements</b>	<b>iii</b>
<b>Table des Matières</b>	<b>iv</b>
<b>Table des Figures</b>	<b>1</b>
<b>Introduction Générale</b>	<b>1</b>
<b>1 Estimation de la volatilité</b>	<b>4</b>
1.1 Introduction . . . . .	4
1.2 Revue de la littérature . . . . .	4
1.2.1 Recherche académique sur l'estimation de la volatilité . . . . .	4
1.2.2 Pratiques de l'industrie dans l'estimation de la volatilité . . . . .	5
1.3 Quelques notions financières . . . . .	5
1.3.1 Marché financier . . . . .	6
1.3.2 Option financière . . . . .	6
1.3.3 Risque financier . . . . .	6
1.3.4 Actifs financiers . . . . .	7
1.3.5 Portefeuille . . . . .	7
1.3.6 Actions financières . . . . .	7
1.4 La volatilité financière . . . . .	7
1.4.1 Définition . . . . .	7
1.4.2 L'impact de la volatilité en finance . . . . .	7
1.5 Aperçu des méthodes traditionnelles d'estimation . . . . .	8
1.6 Fondements mathématiques et concepts associés . . . . .	10
1.7 Conclusion . . . . .	10
<b>2 Machine Learning</b>	<b>11</b>
2.1 Introduction . . . . .	11
2.2 Signification 'Modèle' en apprentissage automatique . . . . .	11

2.3	Revue de la littérature sur les techniques d'apprentissage automatique en finance . . . . .	12
2.4	Application de l'apprentissage automatique en finance . . . . .	13
2.5	Conclusion . . . . .	14
<b>3</b>	<b>Contribution</b>	<b>15</b>
3.1	Introduction . . . . .	15
3.2	Régression Linéaire . . . . .	16
3.3	Outils utilisés . . . . .	18
3.3.1	Python . . . . .	18
3.3.2	Google Colab . . . . .	20
3.4	Étapes clés du processus de développement . . . . .	20
3.4.1	Collecte de données et prétraitement . . . . .	20
3.4.2	Calcul de predicteurs . . . . .	22
3.4.3	Entraînement du modèle . . . . .	27
3.4.4	Tester le modèle . . . . .	28
3.4.5	Evaluation du modèle . . . . .	28
3.4.6	Analyse et Interpretation des resultats . . . . .	29
3.5	Conclusion . . . . .	31
	<b>Conclusion Générale</b>	<b>33</b>
	<b>Bibliographie</b>	<b>33</b>
	<b>Résumé</b>	<b>35</b>

# Table des Figures

3.1	Illustration d'un exemple de régression linéaire . . . . .	18
3.2	Importer les bibliothèques en Python. . . . .	19
3.3	Google Colab. . . . .	20
3.4	Collecte de données et prétraitement. . . . .	22
3.5	Le calcul de prédicteurs en Python. . . . .	22
3.6	Code Python pour afficher un graphique des prix de clôture. . . . .	23
3.7	Les prix de clôture de l'actif financier AAPL. . . . .	24
3.8	Code Python pour calculer la volatilité historique. . . . .	25
3.9	Le graphe de la volatilité historique. . . . .	26
3.10	Entraînement du modèle. . . . .	27
3.11	La volatilité observé vs prédit. . . . .	30



# Introduction Générale

Dans le monde d'aujourd'hui, la finance joue un rôle extrêmement important et peut parfois être à l'origine de crises mondiales. Il est donc primordial que la finance repose sur des modèles solides permettant de gérer les risques et les prix. De ce fait, le modèle et la formule de Black-Scholes se sont imposés comme référence depuis 1973 [1] dans le domaine du calcul des options. Malgré ses défauts, ce modèle connaît un tel succès car il présente de nombreux avantages : sa simplicité d'application et de formulation, son utilisation répandue par les opérateurs du marché, mais surtout parce qu'il permet de calculer un paramètre important en finance : la volatilité.

La volatilité est l'épine dorsale de la finance dans le sens où elle fournit non seulement un signal d'information aux investisseurs, mais elle constitue également un élément d'entrée dans divers modèles financiers. Elle reflète l'incertitude et la variabilité des mouvements des prix sur les marchés financiers. Dans le domaine de la finance, de nombreux modèles ont été développés afin d'estimer la volatilité future, offrant ainsi aux investisseurs la possibilité de prendre des décisions éclairées.

Dans ce mémoire, nous nous concentrons sur l'utilisation de l'apprentissage automatique, plus précisément de la régression linéaire, pour estimer la volatilité future des prix d'action. La littérature existante regorge de nombreuses contributions portant sur la prévision de la volatilité. Cependant, la plupart des approches avancées proposées par différents chercheurs pour modéliser et anticiper la volatilité des rendements s'appuient en grande partie sur des modèles paramétriques tels que le modèle de moyenne mobile pondérée exponentielle (EWMA) [2], le modèle d'autorégression conditionnelle hétéroscédastique (ARCH) et le modèle d'autorégression conditionnelle hétéroscédastique généralisée (GARCH) [3].

Bien que ces modèles de prévision de volatilité soient couramment utilisés et utiles, ils présentent également quelques inconvénients, comme la sensibilité aux données récentes. En effet, ces modèles sont souvent sensibles aux données les plus récentes, ce qui signifie qu'ils peuvent réagir de manière excessive aux fluctuations à court terme de la volatilité. Cela peut entraîner des prévisions instables ou bruitées.

Pour toutes ces raisons, la pertinence de la prévision de la volatilité s'ensuit comme une conséquence naturelle. Bien que la littérature sur ce sujet soit abondante et que

de nombreux modèles aient été proposés, nous avons remarqué ces dernières années à une augmentation des applications des techniques de Machine Learning (ML) dans de nombreux secteurs différents.

Le Machine Learning (ou apprentissage automatique), branche de l'intelligence artificielle, connaît un vif regain d'intérêt ces dernières années grâce au Big Data [4]. Appliquée dans des domaines variés comme le marketing digital, la biologie, l'économie ou l'industrie automobile, l'analyse prédictive représente un enjeu stratégique considérable.

En finance, les applications sont multiples : tarification, détection de fraudes, gestion de portefeuilles. Elles concernent à la fois les banques de détail et les marchés financiers. Dans ce domaine, les données ont toujours été utilisées, mais la tendance était jusqu'à présent orientée vers la conception de modèles probabilistes complexes (processus stochastiques). Le Machine Learning ouvre un champ de recherche complémentaire, qui se recentre sur l'exploitation des données [5]. Depuis l'hypothèse d'efficience des marchés [6], la question de la prédictibilité des cours a été régulièrement étudiée, en particulier au travers d'études d'événements, à la fois avec des méthodes traditionnelles (régressions, analyse en composante principale, analyse factorielle) et plus récemment grâce aux nouveaux algorithmes de Machine Learning.

Le principal défi auquel nous sommes confrontés est de trouver un moyen précis et fiable pour estimer la volatilité future des prix d'action. La volatilité joue un rôle clé dans la prise de décision en matière d'investissement, car elle permet de mesurer le risque associé à un actif financier. En utilisant la régression linéaire et les techniques d'apprentissage automatique, nous espérons développer un modèle capable de prédire avec précision la volatilité future, offrant ainsi aux investisseurs un outil puissant pour évaluer les risques et prendre des décisions éclairées.

La régression linéaire est un outil puissant en apprentissage automatique pour modéliser la relation entre une variable dépendante et des variables indépendantes en utilisant une fonction linéaire. Elle est largement utilisée dans de nombreux domaines, tels que l'économie, la finance, la biostatistique et bien d'autres. Ce mémoire est organisée en trois chapitres principaux. Le premier chapitre aborde l'importance de la volatilité dans le domaine financier et présente les différents modèles utilisés pour estimer cette dernière. Le deuxième chapitre se concentre sur l'apprentissage automatique, en mettant particulièrement l'accent sur les modèles récents, renommés et puissants, Le troisième chapitre met en évidence notre contribution, qui consiste en le développement d'une application Python utilisant le modèle de régression linéaire pour prédire la volatilité future des prix d'action.

# Chapter 1

## Estimation de la volatilité

### 1.1 Introduction

Ce chapitre vise à fournir un aperçu de quelques différentes méthodes d'estimation, Nous discuterons des concepts clés, des hypothèses et des limites associées à chaque méthode.

### 1.2 Revue de la littérature

#### 1.2.1 Recherche académique sur l'estimation de la volatilité

Les chercheurs se sont penchés sur différentes méthodes et approches pour estimer la volatilité des actifs financiers. Ces études visent à améliorer notre compréhension de la volatilité et à développer des modèles plus précis pour prédire les mouvements des marchés.

Les chercheurs ont exploré diverses techniques d'estimation de la volatilité, telles que les modèles ARCH (Hétéroscédasticité conditionnelle autorégressive) et GARCH (Hétéroscédasticité conditionnelle autorégressive généralisée). Ces modèles prennent en compte la volatilité conditionnelle, c'est-à-dire la volatilité qui varie en fonction des conditions du marché.

La recherche académique a également examiné l'utilisation de données historiques pour estimer la volatilité. Les chercheurs ont étudié l'efficacité des différentes périodes de données historiques, ainsi que l'impact des événements économiques et politiques sur la volatilité.

En outre, Les chercheurs visent à évaluer les performances des différentes méthodes d'estimation de la volatilité. ils ont comparé les résultats obtenus par différentes approches

et ont cherché à identifier les modèles les plus précis et les plus fiables.

Les chercheurs continuent d'explorer de nouvelles méthodes et approches pour estimer la volatilité, dans le but d'améliorer la précision des prévisions et de mieux gérer les risques financiers.

### 1.2.2 Pratiques de l'industrie dans l'estimation de la volatilité

Les pratiques de l'industrie dans l'estimation de la volatilité peuvent varier en fonction du secteur et des instruments financiers concernés. Cependant, certaines approches couramment utilisées incluent :

- Modèles ARCH et GARCH : Les modèles ARCH (Autoregressive Conditional Heteroskedasticity) et GARCH (Generalized Autoregressive Conditional Heteroskedasticity) sont largement utilisés pour estimer la volatilité. Ces modèles prennent en compte la volatilité conditionnelle, c'est-à-dire la volatilité qui varie en fonction des conditions du marché.
- Données historiques : Les données historiques sont souvent utilisées pour estimer la volatilité. Les praticiens de l'industrie peuvent utiliser différentes périodes de données historiques en fonction de la fréquence des données disponibles et de la stabilité des marchés.
- Méthodes statistiques : Différentes méthodes statistiques peuvent être utilisées pour estimer la volatilité, telles que l'estimation par maximum de vraisemblance (MLE) ou l'estimation par moindres carrés.
- Modèles de volatilité stochastique : Certains praticiens de l'industrie utilisent des modèles de volatilité stochastique, tels que le modèle de volatilité stochastique à sauts (SVJ), pour capturer les caractéristiques spécifiques de la volatilité des actifs financiers.

Il est important de noter que les pratiques de l'industrie peuvent évoluer avec le temps et peuvent varier d'une institution financière à l'autre. Les praticiens de l'industrie doivent prendre en compte les spécificités de leur domaine d'activité et les caractéristiques des actifs financiers qu'ils estiment, les hypothèses sous-jacentes et les objectifs de leur analyse pour choisir la meilleure approche pour leur étude.

## 1.3 Quelques notions financières

Nous allons commencer par présenter quelques bases des modèles financiers qui sont fondés sur la revue de la littérature issue de [7] et [8]

### 1.3.1 Marché financier

Le marché financier est un lieu (parfois virtuel) où l'on achète et vend des titres financiers (ou actifs financiers).

Les principales caractéristiques d'un marché financier sont :

- La rencontre de deux contreparties (vendeurs et acheteurs).
- La cotation continue des produits financiers.
- L'élaboration de bonnes conditions pour les transactions, en prenant en compte les objectifs opposés des acteurs du marché.

### 1.3.2 Option financière

Une option financière [9], est un produit dérivé qui confère le droit (et non l'obligation) d'acheter (option d'achat, ou call) ou de vendre (option de vente, ou put) une quantité donnée d'un actif financier (action, obligation, indice boursier, devise, matière première, autre produit financier dérivé, etc.) appelé actif sous-jacent, à un prix préalablement fixé (prix d'exercice), à une date d'échéance donnée (option européenne) ou avant une date donnée (option américaine).

#### Problème des options

Un **call** est une option d'achat, et un **put** une option de vente. Le payoff d'une option (ou valeur intrinsèque) est le maximum entre 0 et le flux engendré par un exercice immédiat de l'option. Pour un call, le payoff est  $\max(0, S - K) = (S - K)_+$ .

La valeur temps d'une option vient du fait qu'il est possible que le prix du sous-jacent varie encore avant l'échéance.

Une option est dans la monnaie engendrait un flux positif si elle était exercée immédiatement. Un call est dans la monnaie si et seulement si  $S > K$ .

Une option est à la monnaie engendrerait un flux nul si elle était exercée immédiatement. Un call est à la monnaie si et seulement si  $S = K$ .

Une option est en dehors de la monnaie engendrerait un flux négatif si elle était exercée immédiatement. Un call est dans la monnaie si et seulement si  $S < K$ .

### 1.3.3 Risque financier

Le risque financier fait référence à la possibilité de subir une perte financière résultant d'une opération financière ou d'une activité économique.

### 1.3.4 Actifs financiers

Les actifs financiers sont des contrats où les parties s'échangent des flux d'argent.

### 1.3.5 Portefeuille

Un portefeuille (en finance) désigne une collection d'actifs financiers détenus par un établissement ou un individu. Cela peut aussi désigner des valeurs mobilières détenues à titre d'investissements, de dépôt, de provision ou de garantie.

### 1.3.6 Actions financières

- Une action est un titre de propriété d'une entreprise qui n'est pas remboursable.
- Le prix d'une action est défini par sa cotation en bourse. Une action peut être vendue ou achetée à n'importe quel moment (pendant les heures d'ouverture de la bourse).
- Le détenteur d'une action devient un associé, proportionnellement au nombre de titres qu'il détient.
- Les émetteurs des actions sont des entreprises, l'émission d'actions permet de recouvrir son investissement initial et ses bénéfices.
- Une action est un produit très volatile, lié à la fois aux performances de l'entreprise et à la situation du marché. Sa cotation est constamment réévaluée en fonction de l'offre et de la demande sur les marchés financiers.

## 1.4 La volatilité financière

### 1.4.1 Définition

La volatilité fait référence au degré de variation ou de fluctuation du prix d'un actif financier sur une période donnée. Elle est utilisée pour évaluer les risques et les rendements potentiels d'un investissement ou d'un portefeuille. La volatilité est souvent considérée comme une mesure de la stabilité ou de l'incertitude d'un marché financier.

### 1.4.2 L'impact de la volatilité en finance

Une volatilité élevée indique une plus grande incertitude et des mouvements de prix plus importants, ce qui peut être associé à des risques plus élevés. À l'inverse, une volatilité faible indique une plus grande stabilité et des mouvements de prix moins importants, ce qui peut être considéré comme moins risqué.

La volatilité joue un rôle crucial dans différentes applications financières, selon [10], Elle est utilisée pour la gestion des risques, permettant aux investisseurs de mesurer et de gérer les risques associés à leurs investissements. Elle est également utilisée dans la valorisation des actifs, en prenant en compte la volatilité pour estimer la valeur d'un actif financier. De plus, la volatilité influence les décisions d'investissement, car les investisseurs doivent tenir compte de la volatilité attendue pour évaluer les rendements potentiels et prendre des décisions éclairées.

## 1.5 Aperçu des méthodes traditionnelles d'estimation

Les méthodes traditionnelles d'estimation de la volatilité comprennent la volatilité historique et la volatilité implicite, comme a dit Nabil Saïm [11].

**La volatilité historique** est calculée en utilisant les données historiques des prix d'un actif financier. Elle mesure la variation des prix dans le passé et est souvent utilisée pour estimer la volatilité future. Cette méthode est basée sur l'idée que les mouvements passés des prix peuvent donner des indications sur les mouvements futurs.

La formule de volatilité historique est donnée par :

$$\sigma = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - \mu)^2} \quad (1.1)$$

- $r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$  est le retour au jour  $t$ .
- $P_t$  : est le prix au temps  $t$ .
- $\mu$  : est la moyenne de  $r_t$ .

**La volatilité implicite**, quant à elle, est calculée à partir des prix des options sur un actif financier. Elle représente la volatilité attendue par le marché pour l'actif sous-jacent. Les options sont des contrats financiers qui donnent le droit d'acheter ou de vendre un actif à un prix prédéterminé à une date future. En utilisant les prix des options, il est possible de rétrocalculer la volatilité implicite qui correspondrait à ces prix.

La formule de volatilité implicite est donnée par :

$$\text{Volatilité implicite} = \text{Fonction inverse de Black-Scholes (P, S, K, T, r)}, [12] \quad (1.2)$$

### Formule de Black-Scholes

La formule de Black-Scholes est utilisée pour calculer la valeur théorique d'une option européenne. La formule est donnée comme suit :

$$C = S * N(d1) - X * e^{(-r * T)} * N(d2)$$

Où :

- C représente le prix de l'option
- S est le prix actuel de l'action sous-jacente
- X est le prix d'exercice de l'option
- r est le taux d'intérêt sans risque
- T est la durée jusqu'à l'expiration de l'option

Les formules pour calculer N(d1) et N(d2) sont les suivantes :

$$N(d1) = \frac{\ln(S/X) + (r + \frac{\sigma^2}{2}) \cdot T}{\sigma \cdot \sqrt{T}}$$

$$N(d2) = N(d1) - \sigma \cdot \sqrt{T}$$

Où :  $\sigma$  est la volatilité de l'action sous-jacente.

La fonction inverse de Black-Scholes, également connue sous le nom de volatilité implicite, est utilisée pour calculer la volatilité d'un actif financier à partir du prix de l'option et des autres paramètres du modèle Black-Scholes. Malheureusement, il n'existe pas de formule analytique simple pour calculer directement la volatilité implicite. Cependant, il existe des méthodes numériques itératives, telles que la méthode de Newton-Raphson ou la méthode de dichotomie, qui permettent de trouver une estimation de la volatilité implicite en utilisant le prix de l'option et les autres paramètres du modèle Black-Scholes.

Ces deux méthodes sont largement utilisées dans la finance pour estimer la volatilité des actifs financiers. Elles ont leurs avantages et leurs limites, et il est souvent recommandé de les utiliser conjointement pour obtenir une estimation plus robuste de la volatilité. Dans notre cas, nous utiliserons la volatilité historique dans le chapitre de contribution car elle est plus simple.



## 1.6 Fondements mathématiques et concepts associés

L'estimation de la volatilité s'appuie sur divers concepts tels que :

- **Rendements** : La volatilité est généralement mesurée à partir des rendements d'un actif financier. Les rendements représentent les variations de la valeur d'un actif sur une période donnée. Ils sont calculés en comparant les prix d'un actif à différents moments, le rendement se calcule comme suit :

$$\text{Rendement} = \frac{\text{Prix final} - \text{Prix initial}}{\text{Prix initial}}$$

- **Variance** : La variance est une mesure de la dispersion des rendements. Elle représente la moyenne des carrés des écarts entre chaque rendement et la moyenne des rendements. La variance est souvent utilisée comme mesure de la volatilité, on peut la calculé comme suit :

$$\text{Variance} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

- **Écart-type** : L'écart-type est la racine carrée de la variance. Il représente la mesure de la volatilité d'un actif financier. Plus l'écart-type est élevé, plus la volatilité est importante, il est mesuré comme suit :

$$\text{Écart type} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

où

- $\mathbf{x}$  représente chaque valeur individuelle dans l'ensemble de données.
- $\bar{x}$  représente la moyenne des données.
- $\mathbf{N}$  représente le nombre total de valeurs dans l'ensemble de données.

## 1.7 Conclusion

L'estimation de la volatilité est un processus continu qui nécessite une surveillance et des ajustements constants à mesure que les conditions du marché évoluent.

Tout au long de ce chapitre, nous avons exploré diverses méthodes et techniques utilisées pour estimer la volatilité, allant des simples calculs de volatilité historique aux modèles économétriques plus avancés, équipant les lecteurs des connaissances et des outils nécessaires pour naviguer dans le monde complexe de la volatilité financière.

# Chapter 2

## Machine Learning

### 2.1 Introduction

L'apprentissage automatique, également connu sous le nom de machine learning, est une branche de l'intelligence artificielle qui se concentre sur le développement de techniques permettant aux ordinateurs d'apprendre et de prendre des décisions sans être explicitement programmés. Au lieu de suivre des instructions spécifiques, les systèmes d'apprentissage automatique sont conçus pour analyser les données, identifier des modèles et tirer des conclusions afin de prendre des décisions ou de faire des prédictions, donc c'est quoi un modèle ?

### 2.2 Signification 'Modèle' en apprentissage automatique

Un modèle de machine learning selon un cours du [Microsofts](#), est un fichier ou une structure qui a été entraîné pour reconnaître certains types de modèles ou de schémas dans les données. Il peut être considéré comme une représentation mathématique ou statistique des relations entre les variables d'un problème donné. Le modèle est créé à partir d'un ensemble de données d'entraînement et est utilisé pour faire des prédictions ou prendre des décisions sur de nouvelles données.

Le modèle de machine learning peut prendre différentes formes en fonction de l'algorithme utilisé et du type de problème à résoudre. Par exemple, il peut s'agir d'un modèle de régression linéaire pour prédire une valeur numérique, d'un modèle de classification pour classer des données dans des catégories prédéfinies, ou d'un modèle de réseau de neurones pour reconnaître des motifs complexes.

L'objectif principal d'un modèle de machine learning est de généraliser à partir des

données d'entraînement afin de pouvoir faire des prédictions précises sur de nouvelles données. Pour cela, le modèle est ajusté en utilisant des techniques d'apprentissage automatique telles que la rétropropagation du gradient, l'optimisation des paramètres ou l'ajustement des hyperparamètres.

Dans le domaine de l'apprentissage automatique, il existe différents types d'algorithmes utilisés pour entraîner les modèles. Certains des algorithmes couramment utilisés sont les classifieurs, les réseaux de neurones, les processus gaussiens, les arbres de décision, les méthodes de régression, etc. Ces algorithmes permettent de modéliser les relations complexes entre les variables et de faire des prédictions ou des classifications basées sur les données d'entraînement.

## 2.3 Revue de la littérature sur les techniques d'apprentissage automatique en finance

Les techniques d'apprentissage automatique sont de plus en plus utilisées dans le domaine de la finance pour analyser et prédire les marchés financiers, détecter les fraudes, évaluer les risques, optimiser les portefeuilles, etc.

Voici quelques exemples de techniques et modèles :

- Réseaux de neurones : Les réseaux de neurones sont des modèles d'apprentissage automatique qui imitent le fonctionnement du cerveau humain. Ils sont utilisés pour prédire les prix des actions, détecter les anomalies dans les transactions financières, etc.
- Arbres de décision : Les arbres de décision sont des modèles qui utilisent des règles de décision basées sur des caractéristiques des données pour prédire des résultats. Ils sont utilisés pour la classification des clients, la détection de fraude, etc.
- Machines à vecteurs de support (SVM) : Les SVM sont des modèles d'apprentissage automatique utilisés pour la classification et la régression. Ils sont utilisés pour prédire les tendances du marché, évaluer les risques, etc.
- Méthodes d'ensemble : Les méthodes d'ensemble combinent plusieurs modèles d'apprentissage automatique pour améliorer les performances de prédiction. Des exemples de méthodes d'ensemble comprennent les forêts aléatoires, les boosters de gradient, etc.

## 2.4 Application de l'apprentissage automatique en finance

Les applications de l'apprentissage automatique (ML) dans le domaine de la finance sont nombreuses et variées. Voici quelques exemples d'applications courantes :

- **Trading algorithmique** : L'apprentissage par renforcement est utilisé pour développer des algorithmes de trading automatisés qui prennent des décisions d'achat et de vente en fonction de modèles et de signaux du marché.
- **Prévision des prix** : Les modèles d'apprentissage automatique peuvent être utilisés pour prévoir les mouvements des prix des actifs financiers tels que les actions, les devises et les matières premières.
- **Gestion des risques** : L'apprentissage automatique peut être utilisé pour évaluer et gérer les risques financiers en analysant de grandes quantités de données historiques et en identifiant les tendances et les modèles de risque.
- **Détection de fraude** : Les algorithmes d'apprentissage automatique peuvent être utilisés pour détecter les transactions frauduleuses et les comportements suspects dans les opérations financières.
- **Prévision des défaillances** : L'apprentissage automatique peut être utilisé pour prévoir les risques de défaillance des entreprises en analysant les données financières et les indicateurs de performance.
- **Analyse de sentiment** : L'analyse de sentiment selon Nedioui Med Abdelhamid dans son thèse [13] utilise des techniques d'apprentissage automatique pour analyser les opinions et les émotions des utilisateurs à partir de données textuelles. Elle est utilisée pour évaluer la réputation des entreprises, prédire les mouvements du marché, etc.
- **Gestion de portefeuille** : L'apprentissage automatique peut être utilisé pour optimiser la gestion de portefeuille en recommandant des allocations d'actifs et des stratégies de trading basées sur l'analyse de données historiques et en temps réel.

Ces exemples ne sont qu'une petite partie des techniques d'apprentissage automatique utilisées en finance. Il existe de nombreuses autres techniques et approches spécifiques à différents problèmes financiers.

## 2.5 Conclusion

La pertinence de l'apprentissage automatique (ML) en finance réside dans sa capacité à analyser de grandes quantités de données financières et à extraire des informations précieuses pour la prise de décision.

# Chapitre 3

## Contribution

### 3.1 Introduction

Dans ce chapitre de contribution, nous allons utiliser un modèle appelé "Régression linéaire" en Python afin d'estimer la volatilité future des actions de l'entreprise Apple en utilisant la volatilité historique. Nous avons choisi ce modèle parmi d'autres options en raison de sa simplicité et de sa facilité de mise en œuvre, ce qui est particulièrement important étant donné notre niveau de compétence en programmation. De plus, nous avons déjà étudié la régression linéaire dans d'autres modules, ce qui nous confère une compréhension préalable de ce modèle.

L'utilisation de la régression linéaire dans notre projet nous permettra de mieux appréhender les concepts fondamentaux de l'apprentissage automatique tout en nous offrant un point de départ solide pour explorer des modèles plus avancés à l'avenir. La volatilité est un aspect essentiel de la prévision des marchés financiers, et la régression linéaire peut nous aider à estimer cette volatilité en utilisant des données historiques.

En utilisant le modèle de régression linéaire, nous serons en mesure d'analyser les tendances passées de la volatilité et de les utiliser pour anticiper les fluctuations futures. Cela nous fournira des informations précieuses pour prendre des décisions éclairées dans le domaine financier.

Dans les sections suivantes de ce chapitre, nous décrirons en détail notre approche pour implémenter la régression linéaire afin de prévoir la volatilité. Nous fournirons des explications claires ainsi que des exemples de code Python pour illustrer les

étapes clés de l'implémentation.

En fin de compte, l'objectif de ce chapitre de contribution est de poser une base solide pour notre projet en utilisant la régression linéaire pour les prévisions de volatilité, tout en renforçant nos compétences en programmation Python.

Dans les sections suivantes, nous expliquerons plus en détail les étapes que nous suivrons pour mettre en œuvre ce modèle. Mais tout d'abord, qu'est-ce que ce modèle ?

## 3.2 Régression Linéaire

Le modèle de régression linéaire, comme illustré dans la Figure 3.1 est une méthode statistique utilisée pour modéliser et analyser la relation entre une variable dépendante continue et une ou plusieurs variables indépendantes. Il suppose une relation linéaire entre la variable dépendante et les variables indépendantes.

Le modèle de régression linéaire simple décrit la relation entre une seule variable indépendante ( $X$ ) et une variable dépendante continue ( $Y$ ). L'équation du modèle de régression linéaire simple peut être exprimée comme suit :

$$Y(t) = \beta_0 + \beta_1 X(t) + \varepsilon$$

où

- $Y$  représente la variable dépendante,
- $X$  représente la variable indépendante,
- $\beta_0$  est l'intercept (la valeur de  $Y$  lorsque  $X$  est égal à zéro),
- $\beta_1$  est la pente de la droite de régression (le changement de  $Y$  pour un changement unitaire de  $X$ ),
- $\varepsilon$  est le terme d'erreur (représentant l'écart entre la valeur réelle de  $Y$  et la valeur prédite par le modèle).

Le modèle de régression linéaire multiple permet d'inclure plusieurs variables indépendantes ( $X_1, X_2, \dots, X_n$ ) pour prédire la variable dépendante ( $Y$ ). L'équation du modèle de régression linéaire multiple peut être formulée comme suit :

$$Y(t) = \beta_0 + \beta_1 X_1(t) + \beta_2 X_2(t) + \dots + \beta_n X_n(t) + \varepsilon$$

Les coefficients  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  représentent les effets des variables indépendantes sur la variable dépendante. Le terme d'erreur  $\varepsilon$  représente la variation non expliquée par le modèle.

Pour estimer les coefficients du modèle de régression linéaire, des méthodes d'estimation telles que les moindres carrés ordinaires (MCO) sont utilisées. L'objectif est de minimiser la somme des écarts quadratiques entre les valeurs réelles de la variable dépendante et les valeurs prédites par le modèle.

Une fois que le modèle de régression linéaire a été ajusté aux données, il peut être utilisé pour prédire les valeurs de la variable dépendante en fonction des valeurs des variables indépendantes. De plus, une analyse des résidus peut être effectuée pour évaluer la qualité du modèle et vérifier si les hypothèses de la régression linéaire sont satisfaites.

En résumé, le modèle de régression linéaire est une approche statistique qui permet de modéliser et d'analyser la relation entre une variable dépendante continue et une ou plusieurs variables indépendantes, en supposant une relation linéaire entre elles. Il est largement utilisé dans de nombreux domaines de recherche et appliqués pour comprendre et prédire des comportements et des tendances.

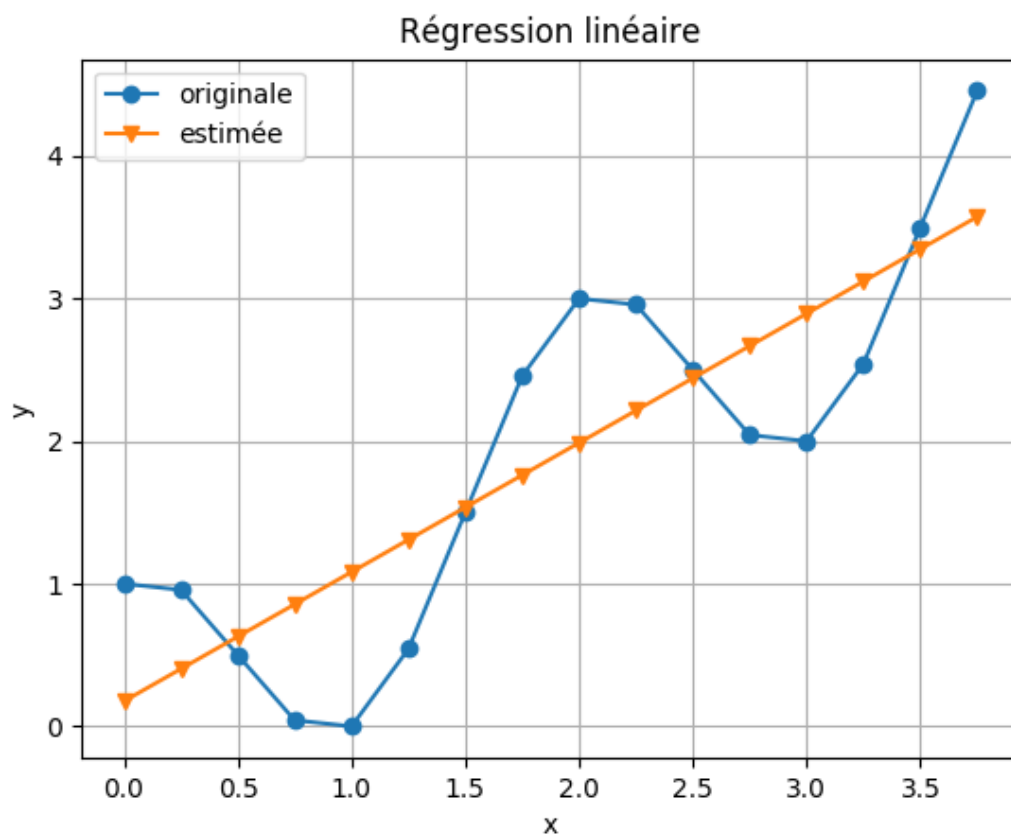
La figure 3.1 est une illustration d'un exemple de régression linéaire pour estimer la fonction suivante :

$$f(x) = x * \cos x * \pi$$

où

- $x \in \mathbb{R}$
- $\pi \approx 3.14$ .





**Figure 3.1:** Illustration d'un exemple de régression linéaire .

Dans cette section, nous prédirons la volatilité du lendemain (écart-type) pour Apple en utilisant quelques caractéristiques indépendantes. L'écart-type est la variable dépendante dans ce cas. Pour illustrer le concept, le modèle sera simplifié.

## 3.3 Outils utilisés

### 3.3.1 Python

Python est un langage polyvalent, convivial et largement utilisé dans le domaine de la programmation. Il offre de nombreuses bibliothèques et frameworks, qui facilitent le traitement des données, la création de modèles statistiques et l'analyse des résultats. Dans notre cas, nous avons utilisé ceux-ci :

### 3.3.1.1 Scikit-learn

Scikit-learn (également connu sous le nom de sklearn) est une bibliothèque open-source en Python qui fournit une large gamme d'outils pour l'apprentissage automatique (machine learning). Dans notre cas, nous l'utilisons pour construire un modèle de régression linéaire, qui est une technique d'apprentissage supervisé permettant de prédire une valeur continue en fonction d'un ensemble de variables indépendantes.

### 3.3.1.2 Pandas

Pandas est une bibliothèque Python qui offre des structures de données et des outils d'analyse de données performants. Elle est souvent utilisée pour manipuler et analyser des données tabulaires, notamment des tableaux et des séries chronologiques. Dans notre cas, nous l'utilisons pour charger, nettoyer et préparer les données avant de les utiliser dans notre modèle de régression linéaire.

### 3.3.1.3 Numpy

Numpy est une bibliothèque fondamentale en Python pour le calcul numérique. Elle fournit des structures de données efficaces pour travailler avec des tableaux multidimensionnels, ainsi que des fonctions mathématiques pour effectuer des opérations sur ces tableaux. Dans notre cas, nous utilisons Numpy pour effectuer des calculs matriciels et des opérations mathématiques nécessaires à la régression linéaire.

### 3.3.1.4 Yfinance

Yfinance est une bibliothèque Python qui permet de récupérer des données financières à partir de Yahoo Finance. Elle permet d'accéder à des informations telles que les cours des actions, les données historiques, les statistiques financières, etc. Dans notre cas, nous utilisons yfinance pour obtenir les données financières nécessaires à notre modèle de régression linéaire, par exemple les prix des actions ou les données économiques pertinentes.

```
[2] import datetime as dt
import pandas as pd
import numpy as np
import yfinance as yf
from pandas_datareader import data as pdr
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error
```

**Figure 3.2:** Importer les bibliothèques en Python.

En résumé, ces bibliothèques 3.2, jouent un rôle essentiel dans notre cas d'utilisation en utilisant Scikit-learn pour la régression linéaire, Pandas pour la manipulation des données, Numpy pour les calculs numériques et yfinance pour l'accès aux données financières.

Python est également apprécié pour sa syntaxe claire et sa grande communauté de développeurs qui partagent des ressources et des solutions.

### 3.3.2 Google Colab

Nous avons utilisé Google Colab 3.3, car c'est une plateforme en ligne gratuite qui permet d'exécuter du code Python et de collaborer facilement sur des projets. Google Colab offre également une grande puissance de calcul grâce à l'utilisation de ressources cloud, ce qui est particulièrement utile pour des tâches de traitement de données ou d'apprentissage automatique qui nécessitent des capacités de calcul importantes.

The screenshot shows a Google Colab notebook titled 'Master.ipynb'. The code cell contains the following commands:

```
[1] !pip install pandas
!pip install yfinance
!pip install pandas-datareader
```

The output shows that all requirements are already satisfied for the installed packages:

```
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (1.5.3)
Requirement already satisfied: python-dateutil<=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz<=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2023.3)
Requirement already satisfied: numpy<=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas) (1.23.5)
Requirement already satisfied: six<=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil<=2.8.1->pandas) (1.16.0)
Requirement already satisfied: yfinance in /usr/local/lib/python3.10/dist-packages (0.2.28)
Requirement already satisfied: pandas<=1.3.0 in /usr/local/lib/python3.10/dist-packages (from yfinance) (1.5.3)
Requirement already satisfied: numpy<=1.16.5 in /usr/local/lib/python3.10/dist-packages (from yfinance) (1.23.5)
Requirement already satisfied: requests<=2.31 in /usr/local/lib/python3.10/dist-packages (from yfinance) (2.31.0)
Requirement already satisfied: multitasking<=0.0.7 in /usr/local/lib/python3.10/dist-packages (from yfinance) (0.0.11)
Requirement already satisfied: lxml<=4.9.1 in /usr/local/lib/python3.10/dist-packages (from yfinance) (4.9.3)
Requirement already satisfied: appdirs<=1.4.4 in /usr/local/lib/python3.10/dist-packages (from yfinance) (1.4.4)
Requirement already satisfied: pytz<=2022.5 in /usr/local/lib/python3.10/dist-packages (from yfinance) (2023.3)
Requirement already satisfied: frozendict<=2.3.4 in /usr/local/lib/python3.10/dist-packages (from yfinance) (2.3.8)
Requirement already satisfied: beautifulsoup4<=4.11.1 in /usr/local/lib/python3.10/dist-packages (from yfinance) (4.11.2)
Requirement already satisfied: html5lib<=1.1 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4<=4.11.1->yfinance) (1.1)
Requirement already satisfied: soupsieve<=1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4<=4.11.1->yfinance) (2.4.1)
Requirement already satisfied: six<=1.9 in /usr/local/lib/python3.10/dist-packages (from html5lib<=1.1->yfinance) (1.16.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from html5lib<=1.1->yfinance) (0.5.1)
```

The execution completed at 8:02AM in 2 seconds.

Figure 3.3: Google Colab.

## 3.4 Étapes clés du processus de développement

### 3.4.1 Collecte de données et prétraitement

### 3.4.1.1 Collecte de données

Dans cette section, nous expliquerons ce qu'est un ensemble de données et comment nous l'avons obtenu à partir de Yahoo Finance.

Un **ensemble de données** est un ensemble structuré de données qui représente des informations spécifiques. Il peut être utilisé pour entraîner et tester des modèles d'apprentissage automatique ou pour effectuer des analyses statistiques. Dans notre cas, nous avons collecté un ensemble de données contenant des informations sur les prix des actions.

Nous avons obtenu notre ensemble de données à partir de **Yahoo Finance**(<https://finance.yahoo.com>), une plateforme en ligne qui fournit des informations financières et des données sur les marchés boursiers. Yahoo Finance propose des données historiques sur les prix des actions de nombreuses entreprises cotées en bourse.

#### Choix de l'ensemble de données (Prix des actions d'Apple)

Parmi les nombreuses options disponibles, nous avons choisi de travailler avec les données des prix des actions d'Apple. Apple est une entreprise technologique bien connue et ses actions sont largement échangées sur les marchés financiers.

En utilisant l'API de Yahoo Finance ou en téléchargeant directement les données historiques des prix des actions d'Apple, nous avons obtenu un ensemble de données contenant les prix d'ouverture, de clôture, les plus hauts et les plus bas, ainsi que le volume des transactions pour chaque jour de négociation.

### 3.4.1.2 Prétraitement des données

Avant d'utiliser notre ensemble de données pour l'analyse ou l'apprentissage automatique, nous avons effectué certaines étapes de prétraitement pour garantir la qualité et la cohérence des données. Voici les étapes courantes de prétraitement des données :

- **Nettoyage des données** : Nous avons vérifié et supprimé toute donnée manquante dans notre ensemble de données.
- **Transformation des données** : Nous avons converti les formats de données si nécessaire, par exemple, convertir les colonnes de dates en formats de date

appropriés.

Le prétraitement des données est une étape essentielle pour garantir la qualité et la fiabilité des résultats. En effectuant ces étapes, nous avons préparé notre ensemble de données des prix des actions d'Apple pour une utilisation ultérieure dans nos analyses ou modèles.

## ▼ Collecter les données & prétraitement

```

✓ [4] end = dt.datetime.now()
0s start = dt.datetime(2019,1,1)
# 'AAPL' = les prix d'action Apple
df = yf.download(['AAPL'], start, end).tail(1826).dropna()
#Afficher le tableau des prix d'action, il y a 1826 jours dans 4 ans.
df

[*****100%*****] 1 of 1 completed

```

Figure 3.4: Collecte de données et prétraitement.

### 3.4.2 Calcul de prédicteurs

Pour utiliser le modèle de régression linéaire dont nous avons parlé précédemment, nous devons d'abord calculer ses prédicteurs d'entrée, Ces métriques serviront comme **variables indépendantes** dans notre modèle de régression linéaire pour estimer la volatilité future d'un actif financier. La figure 3.5 nous explique le calcul de prédicteurs en Python.

```

▶ #Calculer les prédicteurs
df['Volatilite historique'] = df['Close'].rolling(window=5).std()
#Calculer la moyenne
df['Moyenne'] = df['Close'].rolling(window=5).mean()
#Calculer la Variable Cible
df['Variable Cible'] = df['Volatilite historique'].shift(-1)
last_row = df.iloc[[-1]]
#Enlever les valeurs NaN.
df.dropna(inplace=True)
predictors = ['Open', 'High', 'Low', 'Close', 'Volume', 'Volatilite historique', 'Moyenne']
df

```

Figure 3.5: Le calcul de prédicteurs en Python.

La prochaine étape consiste à déclarer nos prédicteurs et notre variable cible. Notre variable cible sera l'écart type de demain, que nous obtenons en décalant l'écart

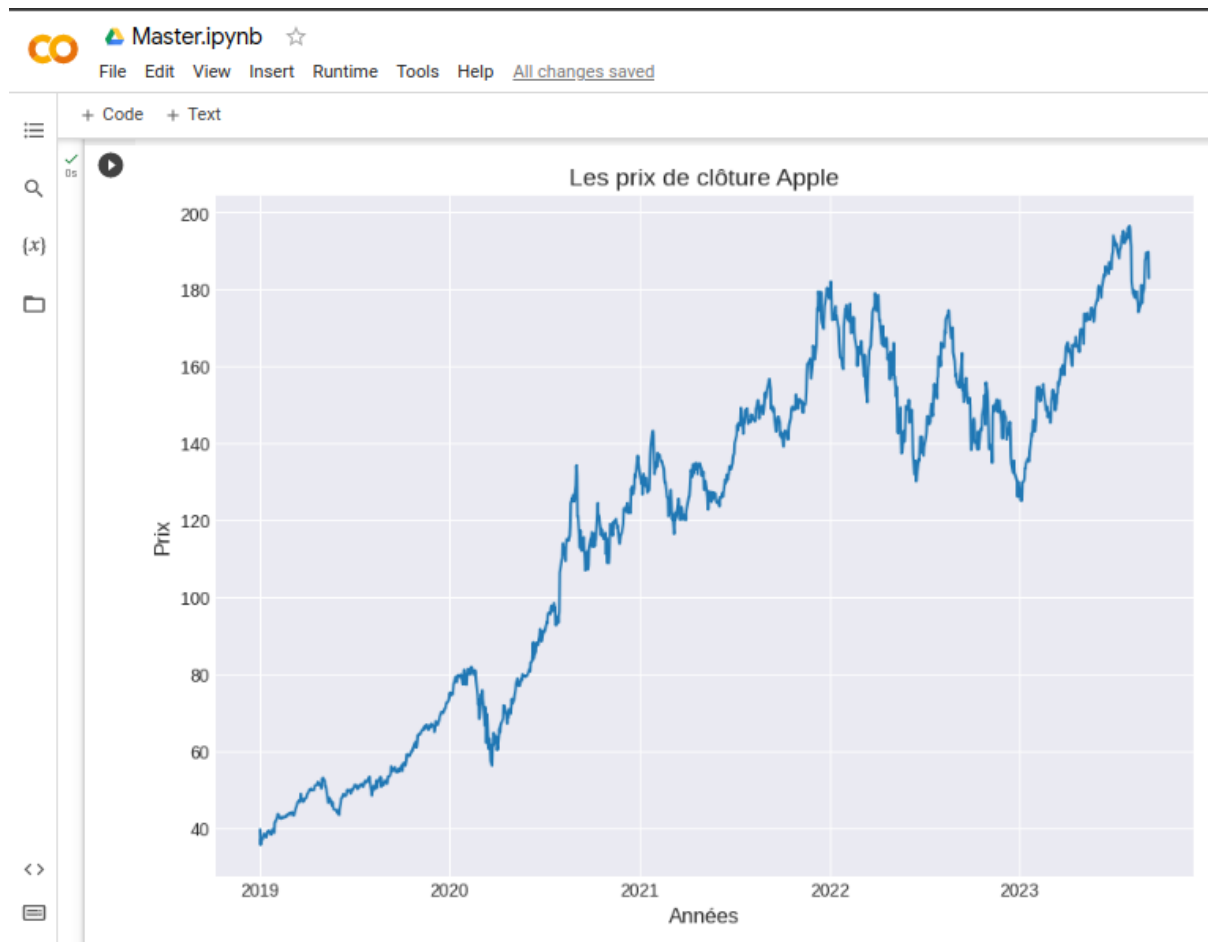
type roulant d'un jour.

Il est important de noter que ces calculs sont basés sur les prix de clôture de l'actif financier, car ils représentent le prix final de l'actif à la fin de chaque période de trading.

Le code Python pour extraire et afficher ces prix est exprimé dans Les figures 3.6 et 3.7

```
▶ #Pour la visualisation
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use("seaborn-v0_8-darkgrid")
#Plotter les prix "close price"
plt.figure(figsize=(10,7))
df['Close'].plot()
#titre et les labels des axes
plt.title('Les prix de clôture Apple', fontsize=14)
plt.xlabel('Années',fontsize =12)
plt.ylabel('Prix',fontsize =12)
plt.plot(Close)
plt.show()
```

Figure 3.6: Code Python pour afficher un graphique des prix de clôture.



**Figure 3.7:** Les prix de clôture de l'actif financier AAPL.

### 3.4.2.1 Volatilité historique

La volatilité historique est une mesure statistique qui quantifie l'écart-type des rendements d'un actif financier sur une période donnée. Elle permet d'évaluer la variabilité des prix d'un actif et d'estimer le niveau de risque associé à cet actif. La volatilité historique est calculée en utilisant les rendements logarithmiques des prix de clôture, comme expliqué précédemment dans le premier chapitre.

Les rendements logarithmiques sont une mesure des variations relatives des prix d'un actif financier. Ils sont calculés en prenant le logarithme naturel du rapport entre le prix de clôture d'un jour et le prix de clôture du jour précédent. Les rendements logarithmiques sont utilisés car ils permettent de prendre en compte les variations relatives des prix, ce qui est important pour l'analyse des séries temporelles financières. Le tableau suivant est extrait de notre code montrant quelques rendements logarithmiques calculés :

Date	Les rendements logarithmiques
2019-01-03	-0.104924
2019-01-04	0.041803
2019-01-07	-0.002228
2019-01-08	0.018884
2019-01-09	0.016839
...	
2023-09-01	0.008428
2023-09-05	0.001266
2023-09-06	-0.036450
2023-09-07	-0.029686
2023-09-08	0.004103

**La volatilité historique** peut être exprimée de la manière suivante :

Écart-type des rendements logarithmiques \*  $\sqrt{\text{nombre de périodes dans l'année}}$

Le code Python dans les figures 3.8 et 3.9 est pour calculer et afficher le graphe de la volatilité historique par rapport au temps.

```

}
✓ [13] TRADING_DAYS = 30
0s returns = np.log(df['Close']/df['Close'].shift(1))
] returns.fillna(0, inplace=True)
volatility = returns.rolling(window=TRADING_DAYS).std()*np.sqrt(TRADING_DAYS)
vh=volatility.dropna()
vh

```

**Figure 3.8:** Code Python pour calculer la volatilité historique.

### 3.4.2.2 La moyenne

La moyenne par rapport au temps est également calculée en utilisant les rendements logarithmiques. Elle représente la tendance générale des prix de l'actif sur la période considérée.

**La moyenne** est mesurée comme suit :

Somme des rendements logarithmiques / nombre de périodes.

Après avoir effectué le calcul de la moyenne, voici le tableau obtenu :



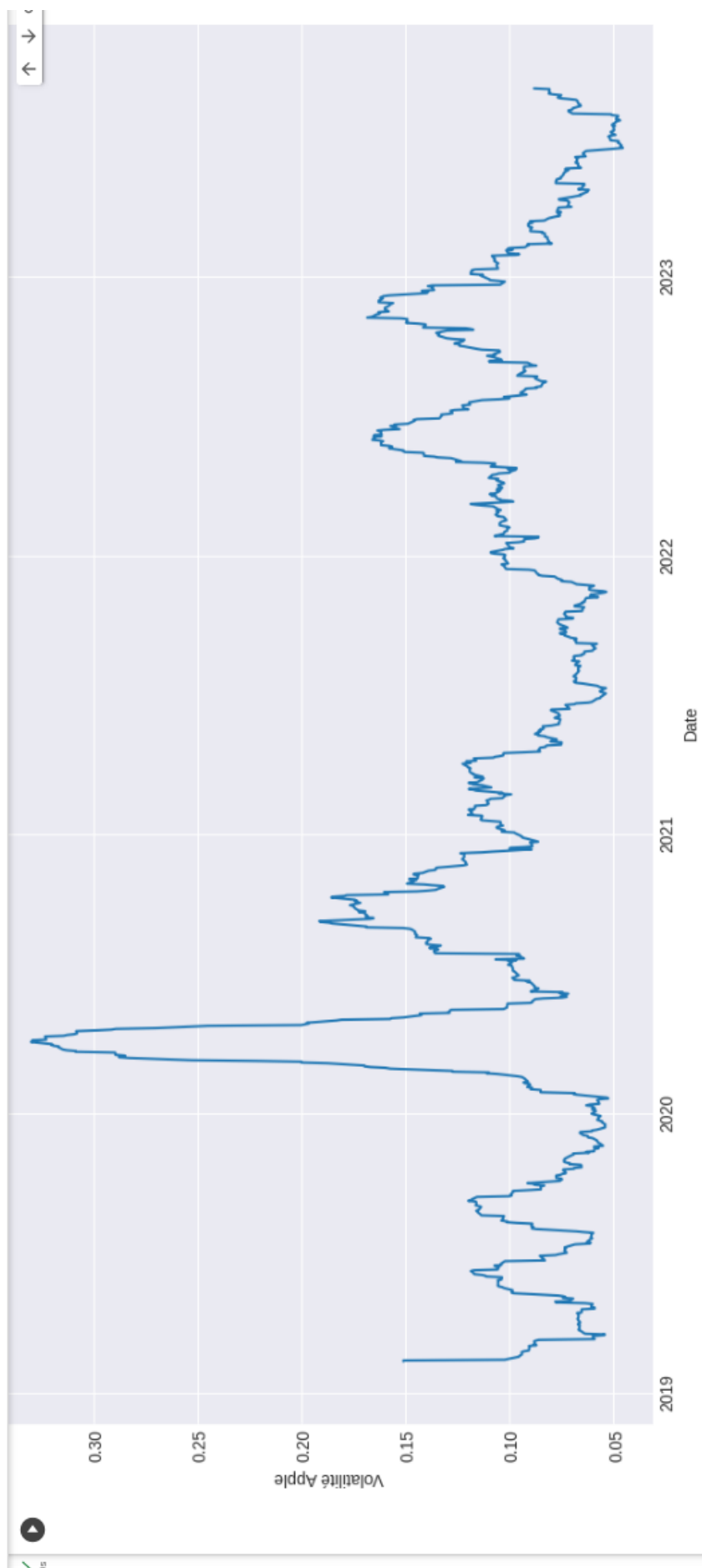
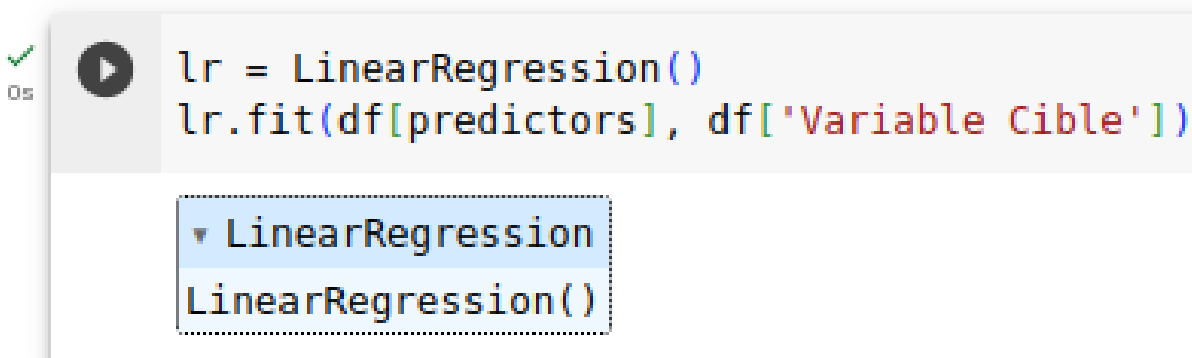


Figure 3.9: Le graphe de la volatilité historique.

Date	Moyenne
2019-01-18	38.534501
2019-01-22	38.699501
2019-01-23	38.742001
2019-01-24	38.630000
2019-01-25	38.725000
...	
2023-08-29	180.084000
2023-08-30	181.389999
2023-08-31	183.687997
2023-09-01	185.857999
2023-09-05	187.759998

### 3.4.3 Entraînement du modèle

Après avoir calculé les prédicteurs, nous devons entraîner le modèle de régression linéaire sur l'ensemble de données qu'on a prétraité. Pour ce faire, il suffit d'appeler le module "LinearRegression()" en Python 3.10, En ajoutant les prédicteurs comme valeurs d'entrée et la variable cible comme sortie.



```
lr = LinearRegression()
lr.fit(df[predictors], df['Variable Cible'])
```

LinearRegression  
LinearRegression()

Figure 3.10: Entraînement du modèle.

Effectuer une division de l'ensemble de données en ensembles d'entraînement et de test est en effet une pratique courante lors de la construction de modèles d'apprentissage automatique, y compris la régression linéaire. Cependant, dans certains scénarios, il peut y avoir des arguments valables pour ne pas effectuer cette division et former le modèle jusqu'au temps  $t$  pour prédire la valeur au temps  $t+1$ .

L'un des arguments en faveur de cette approche est l'information temporelle. En formant le modèle sur toutes les données disponibles jusqu'au temps  $t$  et en prédisant la valeur au temps  $t+1$ , le modèle peut capturer les informations temporelles les plus récentes et les motifs dans les données. Cette approche peut être utile lorsque

l'objectif est de réaliser des prédictions précises basées sur les données les plus récentes.

Par conséquent, nous ajustons le modèle en utilisant l'ensemble de données complet, en utilisant les prédicteurs que nous avons définis précédemment.

### 3.4.4 Tester le modèle

Une fois que le modèle est entraîné, nous pouvons l'utiliser pour faire une prédiction. Dans ce cas, nous sommes intéressés par la prédiction de l'écart type de demain. Nous rassemblons les caractéristiques pour demain, qui représentent les prédicteurs, à partir du dernier point de données disponible. Nous avons sauvegardé la dernière ligne avant de la supprimer en raison des valeurs NaN. Nous transmettons les valeurs/caractéristiques sauvegardées de la ligne au modèle et obtenons la prédiction de l'écart type pour l'écart type du lendemain.

$$VFE = 4.672259967148503$$

tel que VFE représente la volatilité future estimée.

La prédiction signifie que le modèle s'attend à ce que le prix du lendemain fluctue de +/- 4,67 USD par rapport au prix du jour précédent.

### 3.4.5 Evaluation du modèle

Les métriques couramment utilisées en python pour évaluer un modèle sont l'erreur moyenne absolue (MAE), l'erreur quadratique moyenne (MSE) ou l'erreur quadratique moyenne racine (RMSE).

L'erreur quadratique moyenne (MSE) mesure la différence quadratique moyenne entre les valeurs prédites et les valeurs réelles. Elle est calculée en prenant la moyenne des différences au carré entre chaque valeur prédite et sa valeur réelle correspondante. Le MSE est couramment utilisé comme métrique d'évaluation pour les modèles de régression.

$$MSE = 0.4712069540787772$$

D'autre part, l'erreur moyenne absolue (MAE) calcule la différence absolue moyenne entre les valeurs prédites et les valeurs réelles. Elle est calculée en prenant la moyenne des différences absolues entre chaque valeur prédite et sa valeur réelle correspondante. Le MAE est également une métrique couramment utilisée pour évaluer les modèles de régression.

$$MAE = 0.4684843261974196$$

Le RMSE, quant à lui, est simplement la racine carrée du MSE. La différence réside dans la manière dont elles mesurent l'erreur dans un modèle de prédiction.

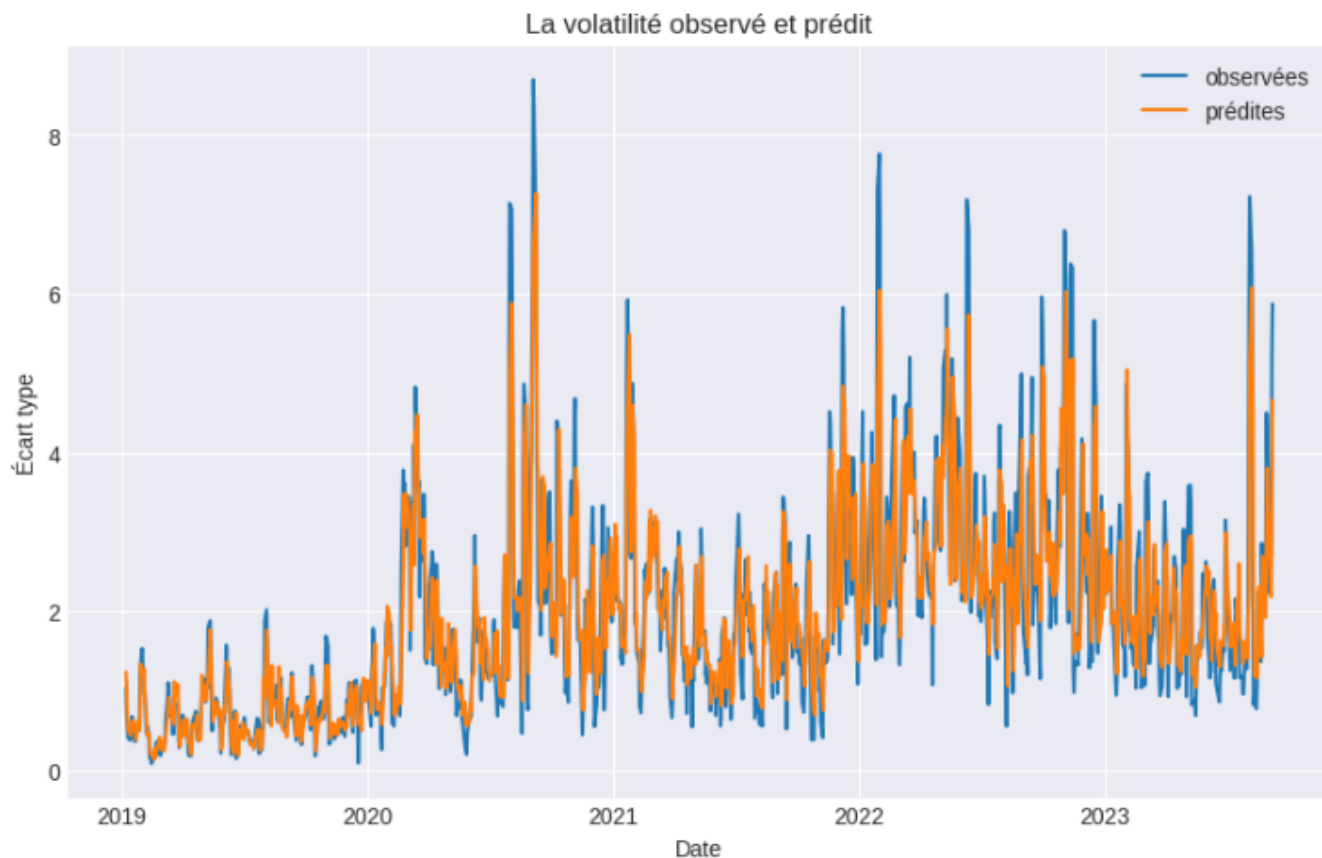
$$RMSE = 0.6864451573714955$$

### 3.4.6 Analyse et Interpretation des resultats

Des valeurs plus basses du MSE et du MAE indiquent de meilleures performances du modèle, car elles indiquent de plus petites erreurs entre les valeurs prédites et les valeurs réelles.

Cependant, il est important de noter que même si le MSE et le MAE indiquent de petites erreurs, il peut y avoir des valeurs aberrantes notables dans les prédictions qui méritent une attention particulière. Ces valeurs aberrantes peuvent avoir un impact significatif sur les performances globales du modèle et doivent être prises en compte.

De plus, pour avoir une compréhension visuelle des performances du modèle, On a généré un graphique dans la figure 3.11 illustrant la variance entre les valeurs observées et les valeurs prédites. Cela peut fournir des informations sur la façon dont le modèle s'ajuste aux données d'entraînement et aider à identifier les motifs ou les écarts.



**Figure 3.11:** La volatilité observé vs prédit.

Les résultats démontrent que l'utilisation du machine learning peut fournir des estimations plus précises de la volatilité des prix d'action. Cette avancée est d'une grande importance pour les professionnels de la finance, car elle permet d'améliorer leur prise de décision en matière d'investissement. La capacité du machine learning à analyser de grandes quantités de données et à identifier des modèles complexes permet d'obtenir des estimations plus précises de la volatilité, ce qui est essentiel pour évaluer les risques et les rendements potentiels.

Cependant, il convient de souligner que l'approche du machine learning présente également des limites :

- L'utilisation de modèles de machine learning peut nécessiter des données volumineuses et de haute qualité.
- Les performances de ces modèles peuvent être compromise si les données disponibles sont insuffisantes ou de mauvaise qualité.
- Les modèles de machine learning sont souvent considérés comme des "boîtes noires", ce qui signifie qu'il peut être difficile de comprendre comment ils prennent leurs décisions. Cela peut poser des problèmes de transparence et de confiance pour les professionnels de la finance qui souhaitent comprendre le

raisonnement derrière les estimations de volatilité.

Malgré ces limites, il existe plusieurs domaines de recherche futurs qui pourraient améliorer l'estimation de la volatilité en finance en utilisant le machine learning. Tout d'abord, il serait intéressant d'explorer l'utilisation de techniques de machine learning plus avancées, telles que l'apprentissage en profondeur, pour améliorer encore davantage les modèles de volatilité. Ces techniques peuvent être capables de capturer des modèles plus complexes et d'obtenir des estimations encore plus précises.

Deuxièmement, il serait pertinent d'étudier l'application de l'apprentissage par renforcement dans l'estimation de la volatilité. Cette approche pourrait permettre aux modèles de s'adapter et d'apprendre en temps réel à partir des données du marché, ce qui pourrait améliorer leur capacité à prédire la volatilité future.

Enfin, il serait intéressant d'explorer l'utilisation de sources de données supplémentaires, telles que les données des médias sociaux ou les données macroéconomiques, pour améliorer l'estimation de la volatilité. Ces sources de données peuvent fournir des informations supplémentaires qui pourraient être utiles pour prédire les mouvements futurs des prix d'action.

En conclusion, l'approche du machine learning offre des avantages significatifs pour l'estimation de la volatilité en finance, en fournissant des estimations plus précises qui peuvent améliorer la prise de décision des investisseurs. Cependant, il est important de prendre en compte les limites de cette approche, telles que les exigences en matière de données et les problèmes de transparence. Les domaines de recherche futurs, tels que l'utilisation de techniques plus avancées et l'exploration de nouvelles sources de données, pourraient contribuer à améliorer davantage l'estimation de la volatilité en finance.

## 3.5 Conclusion

Dans ce chapitre, nous avons démontré comment utiliser la régression linéaire pour prédire l'écart type de l'action AAPL. En analysant les données historiques et en exploitant des prédicteurs pertinents tels que l'écart type et la moyenne par rapport au temps, nous avons présenté un cadre pour construire un modèle capable de fournir des informations sur la volatilité future du marché. Il s'agit d'un modèle simplifié que nous pouvons améliorer à l'avenir. Nous avons bénéficié d'une aide précieuse pour programmer notre modèle grâce à Piotr Szymanski et son [tutoriel](#).

# Conclusion Générale

Ce mémoire vise à combiner les concepts de la finance et de l'apprentissage automatique pour développer un modèle simple mais efficace permettant d'estimer la volatilité future des prix d'action.

L'objectif est d'améliorer la prise de décision des investisseurs et d'ouvrir de nouvelles perspectives dans le domaine de l'analyse financière. En utilisant des techniques avancées de machine learning, nous avons pu explorer différentes approches de modélisation de la volatilité et évaluer leur performance.

Les résultats obtenus ont montré que l'utilisation d'un régresseur linéaire peut fournir des estimations précises de la volatilité, ce qui peut être d'une grande valeur pour les professionnels de la finance.

Cependant, il convient de noter que ce modèle n'est pas sans limites et que des recherches supplémentaires sont nécessaires pour améliorer sa robustesse et son applicabilité dans des contextes financiers spécifiques.

En conclusion, ce mémoire ouvre la voie à exploiter d'autres modèles et de nouvelles opportunités de recherche et offre des perspectives prometteuses pour l'utilisation de l'apprentissage automatique dans l'estimation de la volatilité en finance.

# Bibliographie

- [1] Fischer Black et Myron Scholes. The theory of relativity. *The Journal of Political Economy*, 81(3) :637–654, (May - Jun, 1973).
- [2] Ayad Assoil. La mesure et la gestion du risque de liquidité sur le marché boursier du cac 40. *HAL Open science*), 23 Apr 2021.
- [3] Jonas Kibala Kuma. “ modelisation heteroscedastique : Les modeles arch-garch ”. *HAL Open science*), (18 Apr 2018).
- [4] Amir Gandomi and Murtaza Haider. Beyond the hype : Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2) :137–144, 2015.
- [5] Michaela et Saltelli Andrea et Tarantola Stefano et Hoffman Anders et Giovannini Enrico Nardo, Michela et Saisana. *Handbook on Constructing Composite Indicators and User Guide*, volume 2005. 09 2008.
- [6] Eugene F. Fama. Efficient capital markets : A review of theory and empirical work, papers and proceedings of the twenty-eighth annual meeting of the american finance association new york, n.y.december, 28-30, 1969 (may, 1970), pp. 383-417 (35 pages), published by : Wiley. *The Journal of Finance*), 25(2), (May - Jun., 1973).
- [7] Damien Lamberton et Bernard Lapeyre. Introduction au calcul stochastique appliqué à la finance. ellipse, février, 2012.
- [8] Thierry Roncalli. La gestion des risques financiers. le 23/10/2009 (2eme édition).
- [9] Fatah Cheurfa. "analyse de sensibilité et la quantification de l'incertitude épistémique dans les modèles de risques classiques".
- [10] Nicolas Huth. Some properties of the correlation between the high-frequency financial assets., 27 May 2013.
- [11] Nabil Saïm. Estimation de la volatilité et filtrage non linéaire, 2002.
- [12] Jacin Jerbi. Évaluation des options et la gestion des risques financiers par les réseaux de neurones et par les modèles de volatilité stochastique, mathéma-



tiques[math] de l'université panthéon-sorbonne-paris,2006. Franais. ffNNT : ff. fftel-00308623f.

- [13] Techniques d'apprentissage automatique pour l'analyse et la fouille des sentiments dans les réseaux sociaux ,[online] accessible. <http://thesis.univ-biskra.dz/5581/1/these%20doc.pdf/>. 2020/2021.

## Résumé

Ce mémoire se concentre sur l'application du Machine Learning pour estimer la volatilité des actifs financiers dans le contexte des marchés volatils et incertains.

Dans notre étude sur la volatilité, nous avons choisi de nous concentrer principalement sur la volatilité historique, et d'utiliser la régression linéaire comme méthode de "machine learning" qui a gagné en popularité ces dernières années dans ce contexte. Ces choix ont été faits en raison de leur simplicité, de leur interprétabilité et de leur utilisation répandue dans le domaine financier. La régression linéaire nous permet de modéliser la relation entre la volatilité et les données historiques de manière simple et compréhensible.

**Mots Clés :** Machine Learning, Volatilité historique, Régression linéaire.

## Abstract

This dissertation focuses on the application of Machine Learning to estimate the volatility of financial assets in the context of volatile and uncertain markets.

In our study on volatility, we chose to primarily focus on historical volatility and to use linear regression as a Machine Learning method that has gained popularity in recent years in this context. These choices were made due to their simplicity, interpretability, and widespread use in the financial domain. Linear regression allows us to model the relationship between volatility and historical data in a simple and understandable way.

**Keywords :** Machine Learning, Historical Volatility, Linear Regression.

