



THESIS / THÈSE

MASTER EN INGÉNIEUR DE GESTION À FINALITÉ SPÉCIALISÉE EN DATA SCIENCE

Construction et optimisation de portefeuilles

comparaison des techniques conventionnelles avec les techniques de Machine Learning

GRIFNÉE, Martin

Award date:
2022

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Construction et optimisation de portefeuilles :
comparaison des techniques conventionnelles avec
les techniques de Machine Learning**



Martin Grifnée

Directeur : Prof. S. Béreau

Mémoire présenté en vue de l'obtention du titre de
Master 120 - Ingénieur de Gestion
Finalité Spécialisée Data Science

ANNÉE ACADÉMIQUE : 2021-2022

Remerciements

"Si vous n'essayez jamais, vous ne réussirez jamais, mais si vous essayez, vous risquez de vous étonner vous-même."

Charles-Augustin Sainte-Beuve

Pendant mes années d'études et durant la rédaction de ce mémoire, différentes personnes ont joué un rôle important, je tiens à les remercier.

Tout d'abord, je remercie ma promotrice, la Professeure Sophie Béreau, d'avoir accepté de m'encadrer pour ce mémoire.

Ensuite, je tiens à remercier mes amis proches avec qui j'ai pu avoir des discussions intéressantes liées à ce travail. Je suis reconnaissant pour leur soutien et leur bienveillance.

Enfin, je remercie mes parents, qui m'ont toujours soutenu et encouragé. Leur aide a été particulièrement précieuse lors de la relecture de ce mémoire.

Résumé/Summary

Ce mémoire analyse et compare les techniques traditionnelles de construction et d'optimisation de portefeuilles avec les techniques de *Machine Learning*. Une première rubrique décrit la littérature existante en termes de sélection d'actifs et d'optimisation de portefeuille, selon l'approche classique et moderne avec ses différents cas d'applications. La partie empirique de ce mémoire se concentre sur la construction et l'optimisation d'un portefeuille comprenant dix actions allemandes. Sur base des différentes estimations de rendements attendus et de variances fournies par l'approche classique et les algorithmes de Machine Learning, le but est d'analyser la performance des différentes compositions du portefeuille via le ratio de Sharpe. L'horizon temporel utilisé pour fournir les estimations et les résultats de performance est de cinq ans, allant du 1^{er} janvier 2016 au 31 décembre 2020. Le modèle le plus performant a obtenu un ratio de Sharpe de 1,09, battant ainsi le second modèle traditionnel (0,804) et l'indice de référence DAX (0,576).

This paper analyzes and compares traditional portfolio construction and optimization techniques with machine learning techniques. A first section describes the existing literature in terms of asset selection and portfolio optimization, according to the classical and modern approach with its different application cases. The empirical part of this paper focuses on the construction and optimization of a portfolio comprising ten German stocks. Based on the different estimates of expected returns and variances provided by the classical approach and the Machine Learning algorithms, the goal is to analyze the performance of the different portfolio compositions via the Sharpe ratio. The time horizon used to provide the estimates and performance results is five years, from January 1, 2016 to December 31, 2020. The best performing model achieved a Sharpe ratio of 1.09, beating the second traditional model (0.804) and the benchmark DAX (0.576).

Table des Matières

1	Introduction	6
2	Revue de littérature	9
2.1	La théorie moderne de gestion de portefeuille développée par Harry Markowitz	9
2.2	Faiblesses et solutions à la théorie de Markowitz	11
2.3	Le portefeuille équipondéré	14
2.4	Le portefeuille à variance minimale	15
2.5	La sélection d'actifs financiers	15
2.6	Le Machine Learning en finance	19
3	Méthodologie	22
3.1	Période d'échantillon et sélection des actions	22
4	Analyse des données et résultats	23
4.1	Base de données	23
4.2	Analyse des données	24
4.3	Optimisation	25
4.4	Machine Learning	27
4.5	Résultats	28
5	Limites et discussion	31
6	Conclusion	32
	Bibliographie	34
	Annexes	36

Table des Figures

2.1	Markowitz - Modèle de sélection de portefeuille	9
2.2	Black-Litterman - Modèle de construction de portefeuille	13
2.3	IA & ML - Nombres de travaux de recherche publiés au fil du temps (Bartram et al., 2020)	19
2.4	Techniques de ML - Nombres de travaux de recherche publiés au fil du temps (Bartram et al., 2020)	20
4.1	Rendements annualisés des actions sélectionnées et de l'indice DAX	24
4.2	Écart-types annualisés des actions sélectionnées et de l'indice DAX	24
4.3	Corrélations entre les actions et l'indice de référence	24
4.4	Frontière efficiente et combinaisons différentes du portefeuille . . .	25
4.5	Schéma des différentes étapes de la partie Machine Learning	27
4.6	Frontière efficiente et combinaisons différentes du portefeuille . . .	28
.1	Composition du DAX	37
.2	Matrice de corrélation des actifs sur la période 2016-2018	38
.3	Matrice de corrélation des actifs sur la période 2019-2020	38

Liste des Tableaux

4.1	Tableau d'allocation des actifs en pourcentage	26
4.2	Tableau des caractéristiques des portefeuilles et de l'indice	26
4.3	Tableau des allocations d'actions (%) issus des quatre modèles de ML	28
4.4	Tableau récapitulatif des résultats de performance	29

Construction et optimisation de portefeuilles : comparaison des techniques conventionnelles avec les techniques de Machine Learning

Martin Grifnée

Août 2022

1 Introduction

Au niveau mondial, l'industrie de la gestion d'actifs est un secteur économique majeur et sa valorisation a dépassé les 100 000 milliards d'euros fin 2020 (Heredia et al., 2021). La sélection d'actifs et la prédiction des futurs mouvements de ceux-ci et du marché représentent l'un des plus grands défis quotidiens pour les gestionnaires de fonds. Ils ont aujourd'hui accès à une énorme quantité de données représentant une grande variété d'information. La prolifération de données disponibles impacte l'industrie financière qui poursuit sa révolution technologique (Mike et al., 2021). Pour tirer profit de ce contexte et afin de prendre des décisions efficaces, différents outils et solutions sont développés pour mieux capturer des tendances dans ces données. Pour satisfaire l'objectif principal des gestionnaires de fonds, l'exploitation des données entend apporter de meilleures prédictions pour maximiser les rendements et donc les profits pour chaque niveau de risque. Comparé aux approches plus classiques d'optimisation de portefeuille, les techniques d'intelligence artificielle peuvent souvent fournir de meilleures estimations pour les rendements, variances et covariances. Ces estimations peuvent alors être utilisées dans les *frameworks* traditionnels d'optimisation de portefeuille. Parallèlement, l'intelligence artificielle peut être utilisée plus en amont directement pour la sélection des actifs et l'allocation des poids de ceux-ci constituant un portefeuille atteignant de meilleures performances cibles que les méthodes traditionnelles (Bartram et al., 2020).

Pour construire et optimiser un portefeuille, les gestionnaires de fonds passent généralement par deux grandes procédures. Dans un premier temps, ils sélectionnent des actifs financiers. Cette sélection se fait selon des critères précis et définis que le gestionnaire doit prendre en compte. En général, le gestionnaire s'efforce, en termes de performance, de dépasser un indice de référence souvent lié à sa sélection d'actifs. La méthode classique pour un gestionnaire de fonds est la sélection d'actifs sur base de la qualité financière et des perspectives de performance future de ces actifs. Il existe trois grandes méthodes couramment utilisées par les gestionnaires de fonds : l'analyse fondamentale, l'analyse technique et l'analyse de sentiment. L'analyse fondamentale d'un titre vise à évaluer la valeur intrinsèque d'une entreprise, puis à la comparer au cours actuel du marché. Le gestionnaire décide alors si cette différence est suffisante pour investir ou non dans le titre. Avec l'analyse technique, différents indicateurs et modèles sont construits cherchant à informer le gestionnaire de certaines tendances répétitives dans les données historiques de prix et de volume du titre pour permettre de prédire les rendements futurs des actifs financiers analysés. L'analyse de sentiment ne prend quant à elle pas en compte les données comptables et financières ou les données historiques d'une action, mais plutôt des données qualitatives comprises dans la presse financière ou autre publication. Le but de cette analyse est de sortir un sentiment global soit positif soit négatif permettant de dégager une prédiction sur le futur boursier de l'actif analysé (Kalyani et al., 2016). Dans un second temps, une fois la sélection d'actifs réalisée, le gestionnaire les intègre dans un portefeuille selon le profil du fond et possiblement différentes contraintes. Par exemple, des limitations sur le risque global du portefeuille ou encore le fait de ne pas utiliser de produits dérivés en prenant des positions à découvert. En considérant que le risque est quantifié par la variance des rendements, la théorie classique de la gestion de portefeuille propose un cadre d'allocation des capitaux de telle sorte que la variance d'un portefeuille soit minimisée pour chaque niveau de rendement considéré, en tenant compte de contraintes précédemment spécifiées (Markowitz, 1952).

Cependant, ces méthodes classiques de sélection et d'allocation sont remises en question par la théorie d'efficacité des marchés (Fama, 1970). Cette théorie stipule que le prix actuel des actifs reflète déjà les événements qui ont eu lieu

et les évènements qui sont attendus à se produire. Un marché est donc efficient si les prix reflètent toujours, à tout moment, parfaitement toute l'information disponible. Prédire les prix futurs des actifs serait donc inefficace, balayant l'analyse fondamentale, technique et de sentiment puisque toute sur-évaluation ou sous-évaluation d'un titre aurait déjà été prise en compte. La théorie d'efficience des marchés ne justifie donc pas le travail des gestionnaires de portefeuilles. En effet, elle souligne le fait que personne, pour le même niveau de risque, n'est en mesure de construire un portefeuille d'actifs qui surpasserait de manière constante un portefeuille composé d'actions choisies au hasard. En lien avec la théorie d'efficience des marchés, on retrouve la théorie de *Random Walk* qui soutient une fluctuation du cours boursier totalement aléatoire, qu'aucune tendance ne régit son évolution et qu'il est donc impossible de faire des prédictions sur le cours futur des actifs (Sitte, 2002). Différentes formes d'efficience de marché sont avancées par Fama : l'efficience faible (cours historique du titre reflété dans le prix), semi-forte (cours historique et information publique reflétés dans le prix) et forte (cours historique, information publique et privée reflétés dans le prix). Des études remettent en question la théorie d'efficience des marchés et montrent que les marchés peuvent devenir inefficients (Duan et al., 2009) et que l'efficience est plutôt une caractéristique qui varie dans le temps (Lo, 2005).

C'est dans ce contexte de construction et d'optimisation de portefeuilles qu'il est intéressant de parcourir les modèles de *Machine Learning* appliqués au monde de la finance, et de les comparer avec ceux plus traditionnels. Dans ce mémoire, nous allons dans un premier temps passer en revue la littérature scientifique sur la gestion de portefeuille, c'est-à-dire les théories et méthodes de construction et d'optimisation de portefeuilles utilisées afin de sur-performer un indice de référence. Dans un second temps, sur un cas d'étude empirique, le but est de sélectionner un nombre d'actions défini pour construire des portefeuilles à optimiser via différents modèles et ensuite les comparer en termes de performance avec le Ratio de Sharpe. La question de recherche est la suivante :

"Comment construire un portefeuille d'actions allemandes et l'optimiser en utilisant les techniques classiques et celles de *Machine Learning* ? Quelle est sa performance par rapport à l'indice de référence DAX ?"

2 Revue de littérature

La revue de littérature de gestion de portefeuille s'articule en deux grandes parties. La première décrit les théories et méthodes de l'approche classique et la seconde se concentre sur les modèles et cas d'applications du *Machine Learning*. Chacune des parties aborde les aspects de construction et d'optimisation, c'est-à-dire la sélection d'actifs financiers individuels et la combinaison de ceux-ci en allouant des poids à chacun par l'optimisation de portefeuille.

2.1 La théorie moderne de gestion de portefeuille développée par Harry Markowitz

Harry Markowitz (1952), connu comme le père de la théorie moderne de portefeuille, a avec son article *Portfolio Selection* posé les bases sur nos décisions d'investissement en termes d'optimisation et de diversification de portefeuilles. Il a développé des principes fondamentaux sous différentes hypothèses permettant aux investisseurs une répartition optimale de leurs fonds disponibles en étant exhaustif sur les choix possibles d'investissements. Markowitz a supposé qu'un investisseur rationnel préfère un portefeuille moins risqué pour un niveau de rendement donné et un portefeuille avec un meilleur rendement pour un niveau de risque donné (où le risque est mesuré comme la variance des rendements). Le rendement attendu est donc pour l'investisseur rationnel quelque chose de convoitable, à l'inverse de la variance.

Pour ce faire, Markowitz a quantifié le rendement et le risque des titres par des mesures statistiques, respectivement le rendement attendu (μ) et la variance (σ^2). Il prône le fait que ces deux mesures doivent être considérées ensemble pour répartir au mieux les fonds à investir afin de trouver parmi toutes les combinaisons d'investissements possibles, le meilleur compromis rendement/risque. Le modèle considère aussi qu'il existe un actif sans risque.

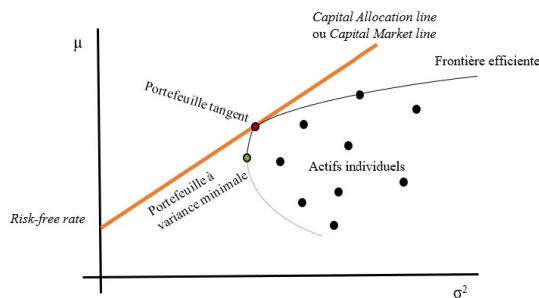


Figure 2.1 : Markowitz - Modèle de sélection de portefeuille

Il définit la fonction d'utilité quadratique de l'investisseur comme suit :

$$U(\mathbf{w}) = \mathbf{w}^T \mu - \frac{A}{2} \mathbf{w}^T \sigma^2 \quad (2.1)$$

où \mathbf{w} est le vecteur des poids assignés à chaque actif du portefeuille, μ est le vecteur des rendements attendus, A est un paramètre d'aversion aux risques propres au profil de chaque investisseur et σ^2 reprend les variances estimées. L'objectif est donc de choisir les poids qui maximisent cette fonction, c'est-à-dire l'utilité de l'investisseur. Dans la réalité, μ et σ^2 ne sont pas connus et sont estimés en termes d'espérances avec les rendements historiques (r_t) sur horizon temporel défini (T).

$$\begin{aligned} \hat{\mu} &= \frac{1}{T} \sum_{t=1}^T r_t \\ \hat{\sigma}^2 &= \frac{1}{T-1} \sum_{t=1}^T (r_t - \hat{\mu})^2 \end{aligned} \quad (2.2)$$

En plus du compromis rendement/risque, l'autre principe novateur apporté par Markowitz est la diversification de portefeuille. Avant la publication de ses travaux, la norme était d'investir séparément dans des actifs individuels avec des hautes perspectives de valeur et de rendements futurs par rapport à leur prix courant. Le principe de diversification est avantageux pour l'investisseur car le risque global d'un portefeuille n'est pas la moyenne pondérée des valeurs de risque individuelles des actifs compris dans le portefeuille. En effet, Markowitz a montré que le risque d'un portefeuille dépend des corrélations entre les actifs le constituant.

Intégrer des actifs individuels non liés entre eux dans un portefeuille, permet donc à l'investisseur de réduire le risque global tout en atteignant des niveaux de rendements plus élevés. Les distributions de rendements des actifs individuels sont différentes et ce sont leurs valeurs de corrélation reprises dans la matrice de covariances qui offrent l'avantage de diversification. Sur le schéma à la page précédente, la frontière efficiente matérialise l'avantage de la diversification (Schulmerich, 2013).

2.2 Faiblesses et solutions à la théorie de Markowitz

Premièrement, les hypothèses générales nécessaires pour l'optimisation de Markowitz sont fortes sur les paramètres d'*input* du modèle. D'un point de vue d'optimisation pure, le portefeuille efficient à choisir est celui offrant la meilleure relation rendement/risque. Ce portefeuille est celui maximisant le ratio de Sharpe ($\frac{R_p - r_f}{\sigma_p}$), c'est-à-dire la différence entre le rendement du portefeuille (R_p) et le taux sans risque (r_f), divisé par l'écart-type (σ_p). On le retrouve graphiquement sur la figure 1, c'est le portefeuille tangent en rouge sur la frontière efficiente, aussi appelé portefeuille de marché.

Pour que le modèle puisse être élaboré, de nombreuses hypothèses doivent être faites sur les rendements des actifs, les investisseurs et les marchés. On a par exemple pour les actifs, un accord sur une distribution normale des rendements d'actifs et l'existence d'un actif non risqué donnant le taux sans risque (r_f). Les investisseurs sont rationnels et souhaitent donc maximiser leur utilité, ils perçoivent le risque en terme de variance (relativement en écart-type) et sont averses aux risques en voulant minimiser le risque pour chaque niveau de rendement considéré. Les investisseurs ne peuvent pas influencer les prix, ils sont preneurs de prix. En ce qui concerne les marchés, ils sont efficaces et absorbent rapidement et parfaitement l'information. L'accès aux rendements et à la volatilité des actifs est correct et gratuit. De plus, on fait les hypothèses qu'il n'y a pas de frais de transaction et de taxes sur les marchés, qu'il est possible de prendre des positions à découvert autant que possible et que chaque actif est négociable à tout moment. Par conséquent, beaucoup de ces hypothèses sont questionnables (Füss, 2008).

Deuxièmement, il est difficile de produire de bonnes estimations pour les rendements attendus et leur variance. Le cadre d'optimisation de Markowitz est très sensible aux erreurs d'estimations en entrée impactant par la suite fortement la composition du portefeuille (Best and Grauer, 1991; Chopra and Ziemba, 1993). Cette sensibilité est directement liée au fait que la matrice de covariance doit être inversée dans le processus d'optimisation. Il s'avère aussi que le modèle est optimal *in-sample* offrant de bons résultats de performance, mais pas pour le cas *out-of-sample*. Les solutions sont souvent concentrées et instables (Jobson and Korbie, 1980; Michaud, 1989). En effet, il y a trois sources potentielles pour estimer les rendements attendus et leur covariances :

1. Les données historiques des cours afin de calculer les rendements attendus futurs et leurs (co)variances.
2. Les estimations prévisionnelles d'experts en analyse fondamentale.
3. Une combinaison des deux points ci-dessus.

Avec l'option 1, le choix de la période historique pour faire les estimations est crucial car l'impact est significatif pour le calcul d'allocation des poids des différents actifs et le calcul de la frontière efficiente. Et puis il y a aussi le fait que les performances passées d'un actif ne garantissent pas les résultats futurs de celui-ci (Black and Litterman, 1992). Les marchés boursiers étant volatiles, les différences de rendements entre différentes fenêtres d'échantillon peuvent être importantes, rendant l'exercice de sélection difficile avec des prédictions fortement sujettes à des erreurs d'estimations. De plus, pour le processus d'optimisation, il est important de s'accorder sur une fréquence de données boursières définie, soit journalière, hebdomadaire ou mensuelle, pour le calcul des rendements attendus.

Avec l'option 2, utiliser des estimations d'analystes fondamentaux a également ses faiblesses. En effet, les estimations peuvent être biaisées par un sentiment global d'optimisme ou de pessimisme. Par exemple, en tant de crise, le temps nécessaire aux analystes financiers fondamentaux pour ajuster leurs estimations peut aussi être long. Ce fut le cas lors de la crise des *subprimes* en 2008 concernant les prévisions des titres de nombreuses institutions financières (Schulmerich, 2013).

D'autres travaux ont également montré que les fonctions objectifs du modèle de Markowitz maximisent les erreurs d'estimations (Michaud, 1989; Black and Litterman, 1992). Nous avons vu précédemment qu'il n'existe aucune estimation exacte pour les rendements attendus, les variances et les covariances, expliquant de ce fait l'exposition aux erreurs d'estimation. Le cadre d'optimisation de Markowitz a tendance à sur-pondérer les actifs à rendements espérés élevés ayant des corrélations négatives et sous-pondérer les actifs avec des rendements espérés faibles corrélés positivement. Sans contrainte au niveau de la vente à découvert, les portefeuilles optimaux obtenus sont souvent composés d'actifs avec des poids soit fortement positifs, soit fortement négatifs, ce qui n'est pas optimal au regard de la performance *out-of-sample*.

Différentes extensions ont été proposées afin d'améliorer l'approche initiale de Markowitz. Elles sont reprises de manière non exhaustive et résumées ci-dessous :

- Le modèle de Black-Litterman (Black and Litterman, 1992) : le modèle est une approche de sélection de portefeuille axée sur l'estimation significative des rendements des actifs en intégrant des pronostics subjectifs.

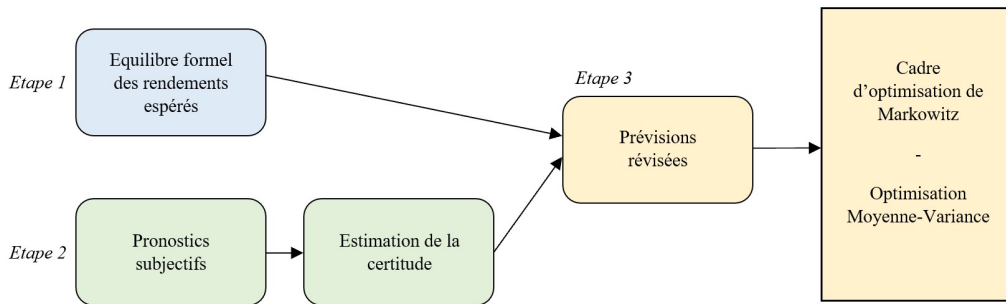


Figure 2.2 : Black-Litterman - Modèle de construction de portefeuille

- Litterman a également développé une extension permettant d'introduire les coûts de transaction dans le problème d'optimisation (Litterman, 2003).
- Plusieurs nouvelles contraintes peuvent également être rajoutées dans le problème d'optimisation, comme le fait d'investir la totalité de ses fonds en actifs risqués ou encore une contrainte sur les positions de vente à découvert (Jagannathan and Ma, 2003).
- Différentes techniques pour mieux estimer ou remplacer la matrice de covariance. Elles modélisent et quantifient l'impact des erreurs d'estimations des rendements attendus et des variances via les moments statistiques. Par exemple, le *shrinkage estimator* de la matrice de covariance (Ledoit and Wolf, 2004).

Enfin, mesurer le risque par la variance a comme défaut d'envisager que les investisseurs sont indifférents devant l'évolution de la valeur du portefeuille dans le sens de la hausse ou dans le sens de la baisse. En réalité, il s'avère que les investisseurs sont particulièrement plus sensibles aux pertes subies qu'aux gains effectués. Le comportement étant dépendant de la fonction d'utilité.

Sachant les défaillances d'utilisation de la variance comme mesure de risque, différents modèles ont vu le jour dans le but de tenir compte du comportement des investisseurs par rapport au risque et au fait que les distributions de rendement des titres peuvent être anormales. Parmi ces mesures de risque plus récentes, on retrouve les mesures dites de *DownSide Risk* qui comme son nom anglophone l'indique ne prend pas entièrement en compte la distribution des rendements, l'intérêt se porte sur la partie gauche de la distribution.

Plus récemment, une autre mesure de risque est de plus en plus utilisée, notamment dans les institutions financières, il s'agit de la *Value-at-Risk* (souvent notée VaR). La *Value-at-Risk* représente la perte potentielle maximale d'un investisseur sur la valeur d'un actif ou d'un portefeuille qui ne devrait être atteinte qu'avec une probabilité donnée sur un horizon donné. C'est donc la pire perte attendue sur un horizon de temps donné pour un certain niveau de confiance. D'autres modèles d'optimisation de portefeuilles existent sur base de la VaR, par exemple les modèles Moyenne-VaR ou Moyenne-CVaR.

Cependant, ce mémoire restera sur l'approche classique du risque en se concentrant sur le modèle Moyenne-Variance de Markowitz pour l'étude empirique.

2.3 Le portefeuille équipondéré

Outre le portefeuille maximisant le ratio de Sharpe, une autre solution plus simpliste est celle du portefeuille équipondéré, c'est-à-dire l'allocation du même poids pour tous les actifs sélectionnés dans le portefeuille. Le même montant est donc investi dans chaque titre. Les difficultés en termes de concentration et d'instabilité des résultats issus de l'optimisation Moyenne-Variance font qu'en pratique, un portefeuille équipondéré produit bien souvent de meilleures performances *out-of-sample* (DeMiguel et al., 2009). Si N actifs financiers sont sélectionnés pour composer un portefeuille équipondéré, on a les poids suivants :

$$\mathbf{w}^* = \frac{1}{N} \mathbf{1}_N \quad (2.3)$$

Les indices de référence les plus connus (S&P500, NASDAQ, Dow Jones, DAX, etc.) reprennent les titres des plus grosses entreprises présents sur la place boursière étudiée et une pondération par capitalisation de chaque titre par rapport au marché est réalisée.

Le concept de portefeuille à pondération égale a gagné en intérêt en raison de la performance historique des actions de petite capitalisation et de l'émergence de nombreux fonds négociés en bourse (ETFs) (Ernst et al., 2017).

2.4 Le portefeuille à variance minimale

Sur la figure 1, ce portefeuille est le point vert et se trouve sur le point le plus bas de la frontière efficiente. Il représente le portefeuille le moins risqué globalement, c'est-à-dire celui avec la plus petite variance parmi tous. En comparaison avec le portefeuille tangent (point rouge), des études empiriques ont montré que le portefeuille à variance minimale globale peut s'avérer plus performant, obtenant de meilleurs rendements hors échantillon (Kempf and Memmel, 2005). Le portefeuille à variance minimale globale est la solution du problème de minimisation suivant :

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^T \boldsymbol{\sigma}^2 \\ \text{s.c.} \quad & \mathbf{w}^T \underline{\mathbf{1}} = 1 \end{aligned} \tag{2.4}$$

où $\mathbf{w} = (w_1, \dots, w_N)$ est le vecteur des poids assignés à chaque actif du portefeuille et $\underline{\mathbf{1}}$ est un vecteur colonne unitaire de dimension $(1 \times N)$.

2.5 La sélection d'actifs financiers

Jusqu'à présent, les différentes méthodes et techniques classiques d'optimisation de portefeuille ont été abordées. Mais avant de pouvoir optimiser un portefeuille, il faut établir sa composition et sélectionner des actifs financiers parmi un univers d'investissement très large (actions, obligations, immobilier, etc.).

Dans ce mémoire, l'étude empirique se concentre sur une classe d'actifs financiers en particulier, les actions. La méthode de sélection d'actions, ou *stock picking* en anglais, consiste à sélectionner un ensemble de titres cotés en bourse se rapportant à des entreprises. Cette sélection se fait sur base de critères comptables et financiers (analyse fondamentale) ou de critères relatifs au marché (analyse technique) ou encore des critères plus qualitatifs (analyse de sentiment).

En introduction, l'hypothèse d'efficience des marchés (Fama, 1970) fut décrite, remettant donc en doute l'utilité et les compétences des analystes experts en

sélection d'actions. Cependant, des études montrent que les marchés peuvent devenir inefficients (Duan et al., 2009) et que l'efficacité est plutôt une caractéristique qui varie dans le temps (Lo, 2005). Il existe donc des périodes où des arbitrages peuvent avoir lieu et les *stocks pickers* peuvent en tirer profit.

En gestion de portefeuille, on identifie en général deux types différents. Le premier est la situation où le gestionnaire adopte un comportement actif et arbitre parmi les actifs du marché, c'est la gestion active. Il sélectionne des titres de manière à pouvoir sur-performer le portefeuille de marché (*benchmark*, souvent l'indice de référence), en maximisant la rapport rendement/risque du portefeuille construit. C'est sur ce premier type de gestion que se fera donc l'étude empirique.

A l'inverse, le second type de gestion est passive et le gestionnaire tente uniquement de reproduire un certain portefeuille de marché ayant pour cible le même rendement que le marché considéré. La gestion passive est généralement moins coûteuse car elle n'implique pas les coûts que la gestion active nécessite. Elle est aujourd'hui très facile à appliquer via les fonds négociés en bourse (ETFs).

Le paradigme de la finance comportementale met également en lumière le fait que les investisseurs peuvent adopter des comportements irrationnels lors de la sélection, la conservation et la ventes de titres. Les êtres humains souffrent de différents biais cognitifs et comportementaux impactant de manière néfaste leur processus décisionnel. On retrouve entre autres les biais suivants (Gill, 2018) :

- La sur-confiance : les gestionnaires surestiment leurs compétences de prévision.
- La familiarité : affection pour les titres, secteurs et/ou régions que les individus connaissent le mieux.
- La représentativité : raccourci mental d'interprétation et d'extrapolation à partir d'échantillon trop petits.
- Le conservatisme : l'individu ne modifie pas son comportement malgré la présence de nouvelles informations.

- La disponibilité : raisonnement se basant principalement sur des informations immédiatement accessibles, par conséquent l'investisseur peut surestimer ou sous-estimer la probabilité d'occurrence d'un évènement.
- La comptabilité mentale : l'investisseur considère des décisions séparément alors qu'elles devraient être prises en compte ensemble.

Après avoir considéré les deux types de gestion de portefeuille et les biais appliqués à la finance, revenons sur les approches possibles pour sélectionner des actions et composer un portefeuille géré activement. Pour rappel, on distingue trois processus d'analyse différents :

1. **L'analyse fondamentale** : elle étudie différentes variables comptables et financières de l'entreprise (chiffres et ratios propres à la croissance, à la rentabilité, aux investissements de l'entreprise) afin d'en estimer sa valeur intrinsèque. Deux autres aspects sont également pris en compte dans l'analyse, la situation macro-économique et l'industrie dans laquelle évolue la société pour identifier la conjoncture actuelle et pouvoir comparer l'entreprise à son secteur (Baresa et al., 2013).

Toujours dans la sélection d'actions, les investisseurs adoptent deux styles de gestion en général. Une approche de valeur ou une approche de croissance, voire une combinaison des deux (*Value vs Growth* en anglais). La première approche concerne des entreprises matures à croissance généralement stable et faible, les actions de ces entreprises sont souvent peu populaires auprès des investisseurs et peuvent donc être sous-évaluées par le marché (ratio P/E faible). L'investisseur optant pour cette approche espère donc une revalorisation future à la hausse de l'entreprise tout en jouissant souvent de versement de dividendes. A l'inverse, la seconde approche entend investir dans des entreprises à forte croissance qui sont souvent assez populaires car les potentiels de croissance future sont conséquents (P/E faible élevé.)

2. **L'analyse technique** : elle se base uniquement sur des caractéristiques du marché boursier. Le cours de bourse d'une entreprise est central dans l'analyse afin de prédire le meilleur moment d'achat ou de vente du titre. Pour cela, différents indicateurs et modèles sont construits cherchant à informer le gestionnaire de certaines tendances répétitives

dans les données historiques de prix et de volume du titre pour inférer l'intérêt ou non du titre.

3. **L'analyse de sentiment** : elle prend en compte des données qualitatives comprises dans la presse financière ou autre publication. Le but de cette analyse est de sortir un sentiment global soit positif soit négatif permettant de dégager une prédiction sur le futur boursier de l'actif analysé. Cette analyse est souvent associée à des algorithmes *Text Mining*, domaine du *Natural Language Processing* (NLP) (Das, 2014; Fisher et al., 2016).

En général, c'est l'analyse fondamentale qui est la plus utilisée, elle est considérée comme la pierre angulaire de la gestion de portefeuille.

2.6 Le Machine Learning en finance

Depuis plus d'une dizaine d'années, les algorithmes d'intelligence artificielle occupent une place toujours plus importante dans de nombreux secteurs et industries. Le *Machine Learning* a en effet, connu un regain d'intérêt récemment car la puissance de calcul informatique et la production et disponibilité des données ont considérablement augmenté. Parallèlement, les coûts des composants et de stockage ont diminué. Cela permet donc aux algorithmes d'être plus robustes et de les rendre applicables et utiles dans de nombreux cas pratiques. Une étude menée fin 2019 a d'ailleurs témoigné de cette évolution, la valorisation des fonds gérés par des algorithmes est désormais plus importante par rapport à la valorisation des fonds traditionnels, où la stratégie d'investissement est déterminée par l'homme (Thomas, 2019). C'est donc à l'heure actuelle l'équivalent de plus de 50 000 milliards d'euros d'actifs qui sont gérés par des algorithmes.

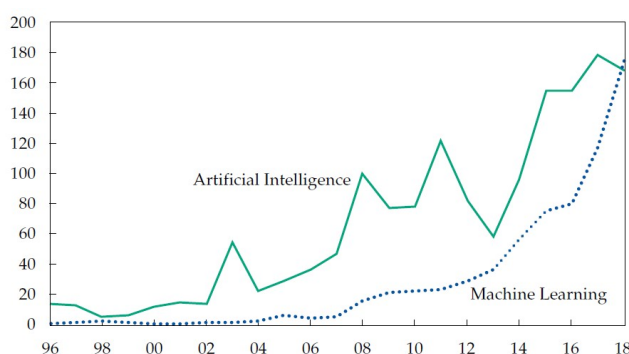


Figure 2.3 : IA & ML - Nombres de travaux de recherche publiés au fil du temps (Bartram et al., 2020)

Le but ici n'est pas de présenter et décrire théoriquement les différents algorithmes de *Machine Learning* mais plutôt de présenter ses cas d'applications et la littérature se rapportant à la gestion de portefeuille. Aujourd'hui, de nombreuses institutions financières utilisent différents modèles de *Machine Learning* pour leur département de gestion d'actif et de portefeuille. Il est également crucial de prendre en compte que ce qu'on appelle IA (Intelligence Artificielle) ou ML (*Machine Learning*) en finance n'est pas nouveau mais existe et a été créé à partir des années 50. L'essence des algorithmes est en réalité sous la forme de techniques de modélisation statistique ou économétrique.

En ce qui concerne la construction de portefeuille, l'intelligence artificielle peut être utile dans ce processus. En effet, en utilisant une série de variables économiques ou concernant l'entreprise, les modèles de Machine Learning sont capables d'identifier les actions susceptibles de sur-performer ou sous-performer un indice de référence par exemple. Les algorithmes utilisent aussi les données historiques pour fournir les données d'entrée (c'est-à-dire les prédictions de rendements attendus, de variances et covariances) du problème d'optimisation de répartition des actifs risqués pour atteindre les objectifs des gestionnaires de portefeuilles. De plus, tout en utilisant les données historiques, des techniques sont également capables d'utiliser des recommandations d'achat ou de vente d'actions émis par de véritables experts fondamentaux (Bew et al., 2019).

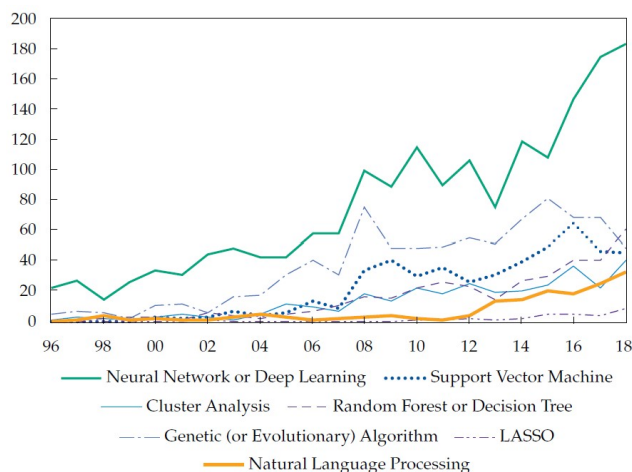


Figure 2.4 : Techniques de ML - Nombres de travaux de recherche publiés au fil du temps (Bartram et al., 2020)

Parmi les différentes techniques disponibles pour la prédiction de rendement, les réseaux de neurones artificiels (ANNs) et les *Support Vector Machines* (SVMs) apparaissent comme les plus performants comparés aux régressions LASSO et par moindres carrés ordinaires, aux forêts aléatoires et arbres de décision et aux algorithmes génétiques (Ballings et al., 2015). A en juger de la performance *out-of-sample*, les résultats des ANNs et SVMs sont en moyenne 30% plus précis en comparaison avec les autres méthodes. Le succès des ANNs et des SVMs réside dans leur capacité à capturer les relations non-linéaires complexes. En général, il est important de considérer que ces résultats sont spécifiques à la tâche et aux données.

Une pratique courante pour profiter des avantages des différentes techniques d'intelligence artificielle, consiste à prendre la moyenne des prédictions des différents résultats obtenus. Dans certains cas, cette approche ensembliste permet de produire de meilleures prédictions par rapport aux résultats individuels (Rasekhschaffe et al., 2019). Des algorithmes sont aussi capables d'offrir une approche alternative de construction de portefeuille sans passer par les traditionnelles techniques linéaires d'optimisation.

En ce qui concerne l'optimisation de portefeuille, nous avons vu précédemment que, dans les faiblesses de l'optimisation selon Markowitz, les poids optimaux sont très sensibles aux données en entrée, c'est-à-dire les estimations des rendements attendus. L'intelligence artificielle tackle donc ce problème en fournissant des estimations de rendements et de risque plus précises par rapport aux techniques conventionnelles, et ces estimations peuvent ensuite être utilisées dans le cadre classique d'optimisation de Markowitz.

La seconde source de problèmes est l'estimation de la matrice de variance-covariance. Pour l'estimer, cela nécessite de grandes séries de données et l'optimisation de Markowitz fait l'hypothèse que les rendements d'actifs ont des corrélations stables. Une technique de remplacement de la structure de la matrice a été développée, avec une structure arborescente en utilisant le clustering hiérarchique (de Prado, 2016). Cette approche demande moins d'estimations en entrée et permet d'obtenir une allocation plus stable et robuste des poids.

Dans la gestion de portefeuille, il est difficile à l'heure actuelle d'affirmer que les techniques de *Machine Learning* sont supérieures aux implémentations conventionnelles. Le sujet est comme on vient de le voir très en vogue et en continuel développement, les gestionnaires viennent seulement récemment de commencer à les utiliser. Mais au vu des avantages, il est indéniable qu'il est fort probable que dans un futur proche, les stratégies d'investissement algorithmiques s'améliorant, l'intelligence artificielle démontre une performance supérieure et absolue par rapport aux techniques traditionnelles.

3 Méthodologie

Cette partie présente la méthodologie utilisée et les choix réalisés pour réaliser l'étude empirique. Dans un contexte de gestion active, le but de l'étude est de construire et optimiser un portefeuille d'actions allemandes via les techniques classiques et modernes, pour enfin les comparer en termes de performances *out-of-sample*. L'étude est implémentée en R sur RStudio. Un intérêt est également porté quant à la capacité du portefeuille construit à battre l'indice de référence DAX.

3.1 Période d'échantillon et sélection des actions

Le DAX est l'indice de marché allemand et il est depuis le troisième trimestre de 2021, composé des actions des 40 plus importantes entreprises cotées à la Bourse de Francfort. Avant cette date, il était composé des 30 plus grandes entreprises cotées à la Bourse de Francfort. Pour que l'étude empirique reste cohérente, la période d'échantillon va du 1^{er} janvier 2016 au 31 décembre 2020. Elle ne va pas au-delà pour que les comparaisons entre le portefeuille et l'indice DAX restent pertinentes.

Notre échantillon comprend donc cinq années boursières complètes. Les trois premières années (2016, 2017 et 2018) sont la période d'entraînement et les deux suivantes (2019 et 2020) sont la période de test. C'est sur la première période que sont extraits les différents estimateurs de rendements attendus et de variance suite à l'implémentation de modèles classique et de *Machine Learning*. La deuxième période permet ensuite d'évaluer la performance des différentes compositions de portefeuilles faites par les différentes techniques.

Pour rappel, l'indicateur principal de performance utilisé dans ce mémoire est le ratio de Sharpe ($\frac{R_p - r_f}{\sigma_p}$), c'est-à-dire la différence entre le rendement du portefeuille (R_p) et le taux sans risque (r_f), divisé par l'écart-type du portefeuille (σ_p). Parallèlement, le rendement du portefeuille construit sera aussi comparé au rendement de l'indice de référence allemand. Trois types de portefeuilles ont été décrits dans la revue de littérature et seront également repris dans ce projet, le portefeuille à variance minimale, le portefeuille équilibré et le portefeuille maximisant le ratio de Sharpe considéré (en théorie) optimal.

Concernant la sélection d'actions composant le portefeuille étudié, un profil fictif d'investisseur a été imaginé. L'investisseur est peu averse aux risques et souhaite investir dans un nombre limité d'entreprises allemandes qui font partie du DAX, respectant ses exigences. Il veut sélectionner des actions d'entreprises à grande capitalisation pour s'assurer une bonne liquidité de marché et il souhaiterait obtenir un rendement moyen supérieur à l'indice de référence DAX. Des contraintes d'optimisation pour obtenir la répartition optimale des poids d'actions ont également été définies. Il entend investir tout son capital dans le portefeuille, c'est-à-dire que la somme des poids alloués aux actions est de 100%. De plus, l'investisseur veut éviter les positions de vente à découvert, il préfère se concentrer sur des positions longues.

La sélection de l'investisseur fictif comprend dix actions d'entreprises allemandes. Il adopte une approche par secteur qui sont ceux de l'assurance (Allianz), du logiciel (SAP), de l'industrie (Siemens), du transport aérien (Lufthansa), des télécommunications (Deutsche Telekom), de la distribution et logistique (Deutsche Post), des services financiers (Vonovia), l'automobile (BMW), le pharmaceutique et la santé (Merck) et le chimique (Symrise).

4 Analyse des données et résultats

4.1 Base de données

Les données historiques journalières de l'indice DAX et des 10 actions sélectionnées ci-dessus sont récoltées depuis l'API de Yahoo Finance via une fonction R. Ce sont les cours boursiers journaliers à la fermeture qui sont récupérés, c'est-à-dire les prix des actions en fin de journée ajustés à une éventuelle distribution de dividendes, un fractionnement d'actions ou d'autres facteurs qui se produisent après la fermeture du marché. Les données sont donc en série temporelle.

Les données du taux sans risque (r_f) sont importées depuis un fichier CSV extrait depuis *wsj.com*. L'actif considéré comme sans risque sont les obligations allemandes à 10 ans. Les taux d'intérêt de ces obligations nous permettent donc d'obtenir le taux sans risque moyen après avoir annualisé les rendements journaliers. Sur la période de 2016 à 2018, le r_f moyen est de 0,32%.

4.2 Analyse des données

Une fois les données d'entraînement de 2016 à 2018 chargés, les rendements journaliers et hebdomadaires des dix actions, de l'indice DAX et du taux sans risque, sont calculés sur la période. On utilise la formule de log-rendement pour avoir les taux de rendements journaliers et hebdomadaires des différents actifs. La formule est la suivante :

$$\text{Rendement d'actif} = \log\left(\frac{\text{Prix}_t}{\text{Prix}_{t-1}}\right) = \log(P_t) - \log(P_{t-1}) \quad (4.2)$$

Les rendements annualisés sont ensuite calculés avec une formule d'annualisation en R. On obtient pour les trois années entre 2016 et 2018, les rendements annualisés suivants :

	ALV.DE	SAP.DE	SIE.DE	LHA.DE	DTE.DE	DPW.DE	VNA.DE	BMW.DE	MRK.DE	SY1.DE
Annualized Return	0.06538642	0.0681951	0.05720794	0.06946252	0.007982604	-0.007515454	0.1348708	-0.06955253	0.005988259	0.01608327
DAX										
Annualized Return	-0.004168102									

Figure 4.1 : Rendements annualisés des actions sélectionnées et de l'indice DAX

Sur les dix actions sélectionnées, huit ont un rendement annualisé supérieur à celui de l'indice sur la période d'entraînement. Les écarts-types annualisés sont aussi calculés, et c'est celui de l'indice qui est le plus faible avec 16,09%.

	ALV.DE	SAP.DE	SIE.DE	LHA.DE	DTE.DE	DPW.DE	VNA.DE	BMW.DE	MRK.DE	SY1.DE
Annualized Standard Deviation	0.1999802	0.1908204	0.2043569	0.3195031	0.1787139	0.2062362	0.1948175	0.226502	0.2078617	0.2017378
DAX										
Annualized Standard Deviation	0.1608982									

Figure 4.2 : Écarts-types annualisés des actions sélectionnées et de l'indice DAX

Une matrice reprenant les corrélations entre chacune des actions est construite. Un vecteur avec les corrélations entre les actions et l'indice de référence est également calculé. Ces données seront par la suite utiles dans le cadre d'optimisation afin d'observer les avantages de la diversification.

	ALV.DE	SAP.DE	SIE.DE	LHA.DE	DTE.DE	DPW.DE	VNA.DE	BMW.DE	MRK.DE	SY1.DE
DAX	0.8460879	0.7620015	0.8150397	0.5219555	0.6853085	0.7472027	0.4382663	0.7845537	0.5590585	0.5427154

Figure 4.3 : Corrélations entre les actions et l'indice de référence

4.3 Optimisation

L'étape d'optimisation commence par définir les spécifications du portefeuille. En suivant l'approche classique de Markowitz, on ajoute ensuite les objectifs de rendement/risque dans les spécifications. Des contraintes sont également définies, c'est-à-dire d'investir la totalité du capital dans la sélection d'actions (la somme des poids est donc 1) et ne pas autoriser les positions de vente à découvert.

Afin d'afficher graphiquement les différentes combinaisons du portefeuille, on calcule dans un premier temps la frontière efficiente en ne considérant que les portefeuilles à variance minimale pour chaque niveau de rendement. Une série de combinaisons du portefeuille est également générée aléatoirement pour se rendre compte des possibilités dans l'espace possible.

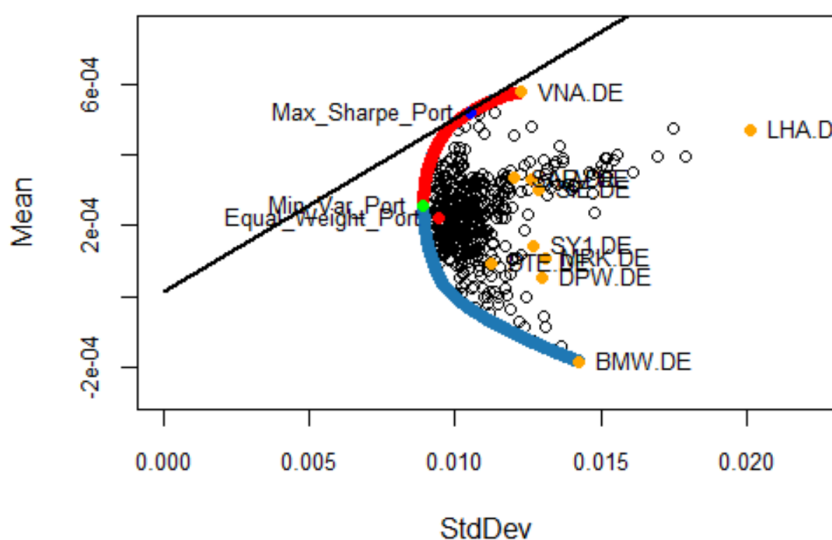


Figure 4.4 : Frontière efficiente et combinaisons différentes du portefeuille

Le graphe ci-dessus permet de visualiser les différents portefeuilles possibles en investissant dans le set des dix actions sélectionnées sur la période 2016-2018. Les points en orange représentent les actifs individuels, la courbe en rouge représente la frontière efficiente et la courbe en bleu représente la frontière Moyenne-Variance qui n'est pas optimale pour un investisseur rationnel.

Trois combinaisons particulières de portefeuille sont aussi représentées. Le portefeuille à variance minimale (point vert), le portefeuille équipondéré (point rouge) et enfin le portefeuille qu'on va considérer comme optimal, le portefeuille maximisant le ratio de Sharpe (point bleu).

Allocation des actifs (%)			
Type de portefeuille	Variance mini- male	Équipondéré	Maximisant le ratio de Sharpe
Allianz	6,31	10	12,84
SAP	8,11	10	3,93
Siemens	9,27	10	1,66
Lufthansa	0,63	10	9,18
Deutsche Telekom	19,05	10	0
Deutsche Post	4,56	10	0
Vonovia	22,38	10	72,39
BMW	2,42	10	0
Merck	14,2	10	0
Symrise	13,07	10	0

Table 4.1 : Tableau d'allocation des actifs en pourcentage

On remarque que le portefeuille maximisant le ratio de Sharpe ne sélectionne que cinq actions sur les dix avec une forte pondération sur Vonovia. Voyons maintenant les performances de chacun des portefeuilles sur la période 2016-2018.

Le portefeuille maximisant le ratio de Sharpe est clairement celui qui sur-performe les autres sur cette période et c'est également le plus optimal en considérant sa valeur de 0,671 pour son ratio. Dans un second temps, la performance de l'allocation de ce portefeuille sera étudiée *out-of-sample*.

Type de portefeuille	Variance mini- male	Équipondéré	Maximisant le ratio de Sharpe	DAX
Rendement annuelisé (%)	6,14	5,8	14,05	-0,42
Écart-type annuelisé (%)	14,15	14,96	16,74	16,09
Beta (β)	0,76	0,87	0,65	1
Ratio de Sharpe	0,335	0,229	0,671	-0,046

Table 4.2 : Tableau des caractéristiques des portefeuilles et de l'indice

4.4 Machine Learning

Dans cette partie, les modèles utilisés pour produire des estimations des rendements attendus sont détaillés. Les variables indépendantes utilisées pour les différents modèles sont la volatilité entre l'ouverture et la fermeture du cours journalier, la volatilité maximum (différence entre le point le plus haut et plus bas sur la journée) et enfin le rendement ajusté réalisé. La variable dépendante va soit être le rendement attendu directement pour les algorithmes de régression et prévision (LASSO et ANNs), soit une valeur binaire de hausse ou baisse du rendement attendu pour les algorithmes de classification (arbres de décision et SVMs).

La période d'entraînement comprend 60% des données de l'échantillon (2016 à 2018) et la période de test 40% du reste des données (allant de 2019 à 2020). Le but va donc être d'entraîner les modèles sur la première période afin de fournir des prévisions en termes de rendement et de variance pour la seconde période. C'est sur base de ces estimations produites par les différents modèles que l'allocation des poids d'actifs va être réalisée sur la deuxième période. Les performances de chacune des combinaisons de portefeuille pourront par la suite être comparées aux techniques traditionnelles.

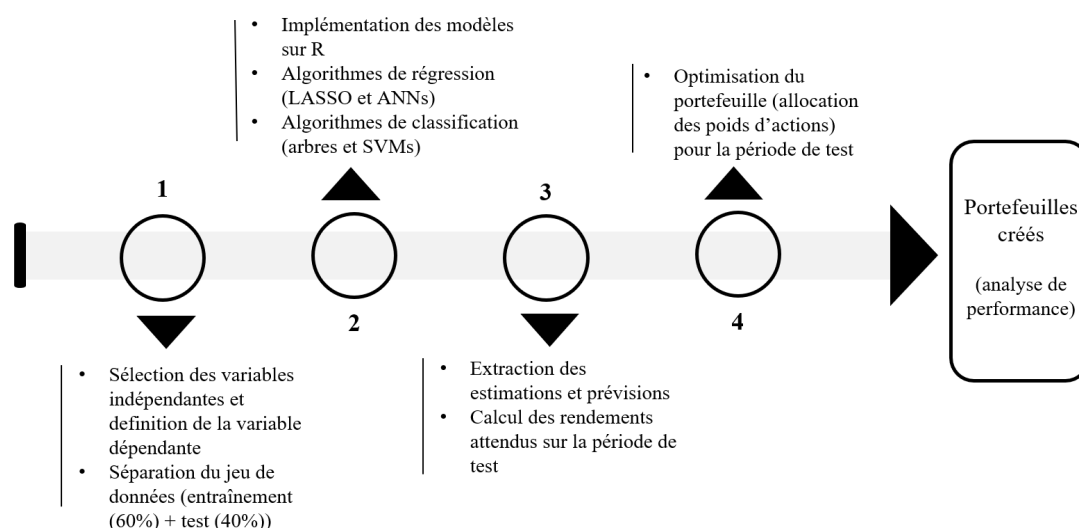


Figure 4.5 : Schéma des différentes étapes de la partie Machine Learning

4.5 Résultats

Allocation des actifs (%)					
Type de modèle ML/classique utilisé	ANN	LASSO	SVM	Arbre	Optimal (Sharpe)
Allianz	19,8	41,4	44,2	10	12,84
SAP	4,2	0	9	10	3,93
Siemens	5,6	0	1,6	10	1,66
Lufthansa	1,4	0	0	10	9,18
Deutsche Telekom	1	0,8	0,6	10	0
Deutsche Post	1,8	1	6	10	0
Vonovia	0	39,4	5,6	10	72,39
BMW	0,2	15,4	20	10	0
Merck	13,8	2	0	10	0
Symrise	52,2	0	13	10	0

Table 4.3 : Tableau des allocations d'actions (%) issus des quatre modèles de ML

On remarque que la répartition de portefeuille produite par le modèle d'arbre de décision est en réalité le portefeuille équipondéré pour la période de test. Une fois cette répartition faite sur base des rendements attendus estimés, il est intéressant de prendre les vrais rendements pour étudier comment performe chacune des répartitions.

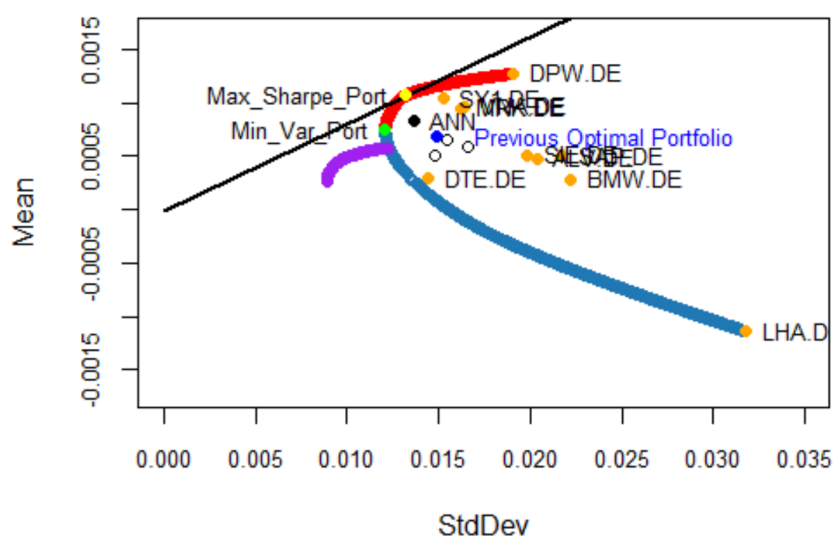


Figure 4.6 : Frontière efficiente et combinaisons différentes du portefeuille

Graphiquement à la page précédente, sur la période de test allant de 2019 à 2020, les combinaisons du portefeuille issues des techniques traditionnelles et de *Machine Learning* sont représentées. Tout d’abord, on peut observer la frontière optimale de la précédente période d’entraînement (en violet) et celle de la période de test (en rouge). Ensuite, on retrouve les actifs individuels en orange, le portefeuille maximisant le ratio de Sharpe en jaune et le portefeuille à variance minimale en vert. Enfin, les dernières combinaisons représentées sont le précédent portefeuille optimal maximisant le ratio de Sharpe sur la période d’entraînement (en bleu), le portefeuille ayant utilisé un modèle de réseaux de neurones artificiels pour l’allocation (en noir) et les trois dernières issues des modèles LASSO, SVM et d’arbres de décision.

Voyons maintenant les performances de chacun des portefeuilles sur la période 2019-2020. Le taux sans risque (r_f) annualisé pour la période est négatif, avec une valeur de -0,34% encourageant encore davantage à investir dans des actifs plus risqués pour ne pas perdre de l’argent pour un investisseur rationnel.

Type de modèle ML/classique utilisé	ANN	LASSO	SVM	Arbre	Optimal (Sharpe)	DAX
Rendement annualisé (%)	23,24	17,93	15,99	13,82	18,7	14,36
Écart-type annualisé (%)	21,64	24,48	26,29	23,47	23,67	25,54
Ratio de Sharpe	1,09	0,746	0,621	0,603	0,804	0,576

Table 4.4 : Tableau récapitulatif des résultats de performance

Le tableau d’allocation des poids des actifs à la page précédente montre une répartition assez hétérogène entre les différents modèles. Un point important à considérer est que les rendements et écarts-types réalisés sur la période 2019-2020 sont considérablement plus importants par rapport à la période d’entraînement 2016-2018. Le portefeuille construit à l’aide du réseau de neurones sur-performe les autres au regard du ratio de Sharpe, avec une valeur de 1,09. La deuxième répartition de portefeuille la plus performante est celle reprise de la méthode traditionnelle qui maximisait le ratio de Sharpe sur la

période 2016-2018, avec 0,804. Le modèle LASSO offre une performance proche avec un ratio de 0,746. Par rapport à l'indice de référence DAX, l'objectif de surpasser son rendement est accompli pour quatre des cinq modèles, seul le portefeuille construit avec les arbres de décision (qui est aussi le portefeuille équipondéré) offre un rendement légèrement inférieur.

Il semblerait que dans ce cas d'étude, les algorithmes sélectionnés se comportent correctement et sont capables de bien estimer et prédire les cours boursiers quand ces derniers sont en période haussière. Une étude similaire rejoint cette analyse, elle observe également que les techniques de *Machine Learning* sur-performent les autres stratégies quand les cours boursiers sont à la hausse (Macchiarulo, 2018).

5 Limites et discussion

Tout au long du cas d'étude empirique, des directions différentes auraient pu être explorées mais des choix devaient être faits afin d'éviter de rester trop généraliste. Cela a potentiellement eu pour conséquence d'impacter les résultats de l'étude. Il y a par exemple le choix de la période d'échantillon et la division du jeu de données de 60% dédiée à l'entraînement et 40% pour tester la performance des différentes méthodes. La sélection de variables indépendantes et la définition de la variable dépendante sont aussi déterminantes. Pour être plus précis et plus exhaustif sur la question, il pourrait être intéressant de considérer un plus long horizon temporel pour avoir encore plus de données financières et être moins sujet aux cycles de certains cours boursiers.

Le choix des paramètres des modèles est aussi important et peut par la suite avoir un impact sur leur performance. Pour cette étude, il y avait peu de paramètres disponibles. De plus, ces derniers étaient assez inflexibles à certaines modifications pouvant améliorer la précision des estimations des rendements attendus. D'autres techniques auraient également pu être utilisées pour rendre les estimations produites encore plus robustes. Par exemple, pour essayer d'éviter encore davantage l'*Overfitting*, la technique de *cross-validation* aurait pu être appliquée.

L'étude empirique a pour but de comparer la performance des techniques de construction et d'optimisation de portefeuilles, techniques à la fois conventionnelles en suivant le cadre développé par Markowitz avec le modèle d'optimisation Moyenne-Variance, et de *Machine Learning*. L'étude n'a considéré que quatre modèles différents, ceux qui paraissaient les plus pertinents au regard du cas d'étude analysé. Or, la liste des algorithmes est longue et il existe une vaste gamme de modèles. Dans la revue de littérature, nous avons vu que combiner certains de ces modèles pouvait encore rendre les estimations plus précises et robustes hors échantillon. Les techniques traditionnelles d'analyse auraient également été intéressantes à intégrer à la comparaison, avec des données en entrées issues de l'analyse technique ou encore l'analyse de sentiment.

6 Conclusion

Ce mémoire avait pour but de comparer la performance des techniques conventionnelles et de *Machine Learning* dans la gestion active de portefeuille, plus précisément les méthodes aidant à la construction et l'optimisation de portefeuilles d'actions allemandes avec un objectif sous-jacent de surpasser le rendement de l'indice de référence, le DAX. Nous avons la question de recherche suivante :

"Comment construire un portefeuille d'actions allemandes et l'optimiser en utilisant les techniques classiques et celles de *Machine Learning* ? Quelle est sa performance par rapport à l'indice de référence DAX ?"

La revue de littérature a dans un premier temps permis d'explorer les techniques classiques utilisées selon deux aspects. Le premier concerne la sélection d'actifs pour construire un portefeuille performant, optimal et diversifié. Le second aborde amplement le modèle d'optimisation de Markowitz et ses limites. Dans un second temps, le rôle et l'avantage du *Machine Learning* en finance et plus particulièrement dans la gestion de portefeuille est abordé. Les algorithmes qui nous ont intéressés pour l'étude sont ceux utilisant les données historiques pour fournir les données d'entrée (c'est-à-dire les prédictions de rendements attendus, de variances et covariances) du problème de répartition des actifs risqués pour atteindre les objectifs des gestionnaires de portefeuilles. Les quatre algorithmes de *Machine Learning* sélectionnés pour cette tâche sont à la fois des algorithmes de régression (ANNs et LASSO) et de classification (arbres de décision et SVMs).

Pour mener cette analyse sur un cas d'étude empirique, un marché boursier a été choisi, un horizon temporel défini et une sélection de dix actions a été faite selon des critères précis. En effet, ce travail se concentre sur ces dix actions du marché allemand et son indice de référence le DAX, récupérant ainsi des données financières sur une période d'échantillon allant du 1^{er} janvier 2016 au 31 décembre 2020. La période d'entraînement démarre au 1^{er} janvier 2016 et s'arrête au 31 décembre 2018, représentant donc 60% des observations. A l'inverse de la période de test qui comprend les derniers 40% des données financières, démarrant au 1^{er} janvier 2019 et s'arrêtant au 31 décembre 2020.

Nous obtenons des résultats intéressants sur la période de test. Le portefeuille construit à l'aide du réseau de neurones sur-performe les autres au regard du ratio de Sharpe, avec une valeur de 1,09. La deuxième répartition de portefeuille la plus performante est celle reprise de la méthode traditionnelle qui maximisait le ratio de Sharpe sur la période 2016-2018, avec 0,804. Le modèle LASSO offre une performance proche avec un ratio de 0,746. Par rapport à l'indice de référence DAX, l'objectif de surpasser son rendement est accompli pour quatre des cinq modèles, seul le portefeuille construit avec les arbres de décision (qui est aussi le portefeuille équipondéré) offre un rendement légèrement inférieur.

Bibliographie

- Ballings, M., VandenPoel, D., Hespeels, N., and Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications, 42(20):46–56.
- Baresa, S., Bogdan, S., and Ivanovic, Z. (2013). Strategy of stock valuation by fundamental analysis. UTMS Journal of Economics, 4:45–51.
- Bartram, S., Branke, J., and Motahari, M. (2020). Artificial intelligence in asset management. CFA Institute Research Foundation, pages 2–13.
- Best, M. and Grauer, R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. The Review of Financial Studies, 4(2):315–342.
- Bew, D., Harvey, C. R., Ledford, A., Radnor, S., and Sinclair, A. (2019). Modeling analysts' recommendations via bayesian machine learning. Journal of Financial Data Science, 1(1):75–98.
- Black, F. and Litterman, R. (1992). Global portfolio optimization. Journal of Financial Analysts, 48(5):28–43.
- Chopra, V. and Ziemba, W. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. Journal of Portfolio Management, 19(2):6–11.
- Das, S. R. (2014). Text and context: language analytics in finance. Foundations and Trends in Finance, 8(3):145–261.
- de Prado, M. (2016). Building diversified portfolios that outperform out-of-sample. Journal of Portfolio Management, 42(4):59–69.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naïve diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies, 22(5):3–12.
- Duan, Y., Hu, G., and Mclean, R. D. (2009). When is stock picking likely to be successful ? evidence from mutual funds. Financial Analysts Journal, 65(2):55–66.

- Ernst, P. A., Thompson, J. R., and Miao, Y. (2017). Portfolio selection: the power of equal weight. arXiv, page arXiv.1602.00782v3.
- Fama, E. (1970). Efficient capital markets: a review of theory and empirical work. Journal of Finance, 25(2):383–417.
- Fisher, I., Garnsey, M., and Hughes, M. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. Intelligent Systems in Accounting, Finance & Management, 23(3):157–214.
- Füss, R. (2008). Ansätze zur optimierung eines portfolios - why not markowitz. Financial Planning Praxis F.
- Gill, K. R. (2018). Study on behavioral finance, behavioral biases, and investment decisions. Journal of Accounting and Financial Management Research, 8(3):1–14.
- Heredia, L., Barletta, S., Carrubba, J., and Frankle, D. (2021). The \$100 trillion machine. BCG Global Asset Management Report 2021, pages 4–28.
- Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. Journal of Finance, 58(4):1651–1683.
- Jobson, J. and Korbie, B. (1980). Estimation for markowitz efficient portfolios. Journal of American Statistical Association, 371:544–554.
- Kalyani, J., Bharathi, P., and Jyothi, P. (2016). Stock trend prediction using news sentiment analysis. arXiv, page arXiv.1607.01958.
- Kempf, A. and Memmel, C. (2005). On the estimation of the global minimum variance portfolio. Working paper No. 05-02 of Centre of Financial Research in Cologne, pages 1–6.
- Ledoit, O. and Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. Journal of Portfolio Management, 30(4):110–119.
- Litterman, R. (2003). Modern investment management. John Wiley Sons Inc., Hoboken.
- Lo, A. W. (2005). Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. Journal of investment consulting, 7(2):21–44.

- Macchiarulo, A. (2018). Predicting and beating the stock market with machine learning and technical analysis. Journal of Internet Banking and Commerce, 23(1):1–22.
- Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7(1):77–95.
- Michaud, R. (1989). The markowitz optimization enigma: Is "optimized" optimal? Journal of Financial Analysts, 45(1):31–42.
- Mike, K., Gerry, H., William, M., Caroline, C., and Ross, T. (2021). Future of asset management. Accenture tech. Report 2021, pages 3–9.
- Rasekhschaffe, R., Christian, K., and Jones, R. (2019). Machine learning for stock selection. Journal of Financial Analysts, 75(3):70–88.
- Schulmerich, M. (2013). The efficient frontier in modern portfolio theory: weaknesses and how to overcome them. Journal of Investment Management Consultants Association, pages 27–32.
- Sitte, R. (2002). Neural networks approach to the random walk dilemma of financial time series. Applied Intelligence, 16(3):164–168.
- Thomas, P.-H. (2019). Votre épargne bientôt gérée par les robots? Trends-Tendance, 48:42–44.

Annexes

Société	Secteur d'activité	Poids dans l'indice (%)
Adidas	Industrie textile	1.22
Allianz	Assurance	7.34
BASF	Industrie chimique	5.62
Bayer AG	Industrie pharmaceutique	7.06
Beiersdorf	Cosmétique	0.75
BMW	Constructeur automobile	1.63
Commerzbank	Banque	0.96
Daimler	Constructeur automobile	5.37
Deutsche Bank	Banque	3.53
Deutsche Börse	Service boursier	2.44
Deutsche Lufthansa	Compagnie aérienne	1.13
Deutsche Post	Service postier	2.22
Deutsche Postbank	Banque	0.23
Deutsche Telekom	Télécommunication	7.39
E.ON	Électricité et Eau	1.37
Fresenius Medical Care	Industrie pharmaceutique	1.43
Henkel	Produits de grande consommation	0.93
Infineon	Semi-conducteur	0.11
K+S	Industrie chimique	1.14
Linde	Ingénierie	2.26
MAN	Constructeur de poids lourds	0.82
Merck	Industrie pharmaceutique	0.90
Metro	Distribution	0.72
Munich Re	Ré-assurance	5.08
RWE	Électricité et eau	5.73
Salzgitter	Métallurgie	0.42
SAP	Informatique	5.08
Siemens	Conglomérat d'ingénierie	9.83
ThyssenKrupp	Industries diversifiées	1.38
Volkswagen Group	Constructeur automobile	7.28

Figure .1 : Composition du DAX

	ALV.DE	SAP.DE	SIE.DE	LHA.DE	DTE.DE	DPW.DE	VNA.DE	BMW.DE	MRK.DE	SY1.DE
ALV.DE	1.0000000	0.5975009	0.6656528	0.4720385	0.5874252	0.6333608	0.2934005	0.6398876	0.4202625	0.3794074
SAP.DE	0.5975009	1.0000000	0.5820922	0.3455187	0.5119131	0.5252025	0.4094042	0.5072663	0.4740111	0.4991832
SIE.DE	0.6656528	0.5820922	1.0000000	0.4131807	0.4886084	0.5836616	0.2924542	0.5829465	0.4225265	0.3877022
LHA.DE	0.4720385	0.3455187	0.4131807	1.0000000	0.3356526	0.4660391	0.2362248	0.4373103	0.2178967	0.2490889
DTE.DE	0.5874252	0.5119131	0.4886084	0.3356526	1.0000000	0.5161477	0.4687945	0.4944028	0.4368113	0.4472770
DPW.DE	0.6333608	0.5252025	0.5836616	0.4660391	0.5161477	1.0000000	0.3687698	0.5750923	0.4284951	0.3825733
VNA.DE	0.2934005	0.4094042	0.2924542	0.2362248	0.4687945	0.3687698	1.0000000	0.2581046	0.3354530	0.4673664
BMW.DE	0.6398876	0.5072663	0.5829465	0.4373103	0.4944028	0.5750923	0.2581046	1.0000000	0.3939379	0.3768378
MRK.DE	0.4202625	0.4740111	0.4225265	0.2178967	0.4368113	0.4284951	0.3354530	0.3939379	1.0000000	0.3798786
SY1.DE	0.3794074	0.4991832	0.3877022	0.2490889	0.4472770	0.3825733	0.4673664	0.3768378	0.3798786	1.0000000

Figure .2 : Matrice de corrélation des actifs sur la période 2016-2018

	ALV.DE	SAP.DE	SIE.DE	LHA.DE	DTE.DE	DPW.DE	VNA.DE	BMW.DE	MRK.DE	SY1.DE
ALV.DE	1.0000000	0.5512779	0.8099514	0.6330677	0.6883038	0.6742282	0.3921324	0.7965911	0.5229336	0.4015992
SAP.DE	0.5512779	1.0000000	0.5878484	0.3717181	0.4878527	0.5560060	0.3599450	0.5068899	0.4529428	0.4149262
SIE.DE	0.8099514	0.5878484	1.0000000	0.5762471	0.6442010	0.7125484	0.3616389	0.7619002	0.5589392	0.4440566
LHA.DE	0.6330677	0.3717181	0.5762471	1.0000000	0.4727144	0.4942007	0.1959336	0.6178006	0.3586130	0.1800344
DTE.DE	0.6883038	0.4878527	0.6442010	0.4727144	1.0000000	0.5903873	0.5110561	0.5960049	0.5110001	0.4623624
DPW.DE	0.6742282	0.5560060	0.7125484	0.4942007	0.5903873	1.0000000	0.4427923	0.6416874	0.5513304	0.5226160
VNA.DE	0.3921324	0.3599450	0.3616389	0.1959336	0.5110561	0.4427923	1.0000000	0.2644718	0.4489419	0.4644083
BMW.DE	0.7965911	0.5068899	0.7619002	0.6178006	0.5960049	0.6416874	0.2644718	1.0000000	0.4867451	0.2923550
MRK.DE	0.5229336	0.4529428	0.5589392	0.3586130	0.5110001	0.5513304	0.4489419	0.4867451	1.0000000	0.4927026
SY1.DE	0.4015992	0.4149262	0.4440566	0.1800344	0.4623624	0.5226160	0.4644083	0.2923550	0.4927026	1.0000000

Figure .3 : Matrice de corrélation des actifs sur la période 2019-2020

Remarque : Toutes les données, les méthodes et le code implémenté sont disponibles sur mon Github (<https://github.com/MartinGrf/Memoire-Code>)