Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers

par

Rachid MIFDAL

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE
AVEC MÉMOIRE EN GÉNIE, CONCENTRATION PERSONNALISÉE
M. Sc. A.

MONTRÉAL, LE 12 NOVEMBRE 2019

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

PRÉSENTATION DU JURY

CE RAPPORT DE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Edmond Miresco, directeur de mémoire
Département de génie de la construction à l'École de technologie supérieure

M. Adel Francis, président du jury
Département de génie de la construction à l'École de technologie supérieure

M. Luc Bégnoche, membre du jury externe
DIVERSIFAI INC. - AIDVISORS.COM

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 11 OCTOBRE 2019

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Tout d’abord, je voudrais présenter ma profonde gratitude envers mon directeur de recherche Edmond Miresco, Professeur et directeur du programme d’ingénierie financière à l’école de technologie supérieure, qui m’a encadré tout au long de ce projet. Il était toujours disponible pour m’aider à avancer dans ma recherche et la rédaction de ce mémoire. Son support était essentiel pour que ce travail puisse voir le jour.

Je dois aussi souligner ma gratitude et mes remerciements à ma femme Kaouthar. Ce travail est le fruit de plusieurs sacrifices. Je suis très reconnaissant de ces années de compréhension, de privation et d’efforts communs. Sans oublier Adam, mon fils qui vient d’achever sa première année dans ce monde. Sa naissance était la plus belle chose qui s’est produite lors de ce projet. Même si ça n’a pas été toujours facile de se concentrer dans ma rédaction.

Finalement je tiens à remercier ma mère, mon frère Lachen, mes sœurs Imane et Malika et tous mes proches et amis, qui m'ont accompagné, aidé, soutenu et encouragé tout au long de la réalisation de ce mémoire.

À tous ces intervenants, je présente mes remerciements, mon respect et ma gratitude.

Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers

Rachid MIFDAL

RÉSUMÉ

Ce mémoire examine l'utilisation des algorithmes et des techniques d'apprentissage automatique pour la prédiction de la direction des cours des actifs financiers à très court terme (par exemple, de quelques minutes à quelques heures). L'objectif à la fin est de concevoir un système décisionnel autonome capable d'automatiser la négociation des titres financiers en générant une prédiction de la tendance future du prix, en se basant sur les flux de données des transactions antérieures.

Nous avons étudié comment la distribution des mouvements boursiers peut être asymétrique. Ce biais dans les données ajoute plus de complexité à la tâche de prédiction. On a proposé trois méthodes pour faire face à ce problème. Entre autres, le rééchantillonnage des données, la modification des seuils des probabilités qui servent à générer la classification et l'utilisation des coûts pour pénaliser les mauvaises classifications de la classe majoritaire.

On a aussi constaté l'existence de dépendances temporelles dans les séries des prix des titres boursiers utilisés. Cette composante devra être considérée lors de la génération de nos ensembles d'entrainement et d'évaluation. En effet, l'exactitude de l'inférence faite à partir des données d'entrainement sur de nouvelles données, dépendra de la similarité des propriétés statistiques des deux ensembles de données.

Nous avons aussi illustré que le choix des variables à utiliser, les indicateurs techniques pour ce projet, est une tâche qui nécessite beaucoup de rigueur. Comme nous l'avons démontré, il existe de fortes corrélations entre les indicateurs techniques qui sont dérivés des mêmes séries de données. C'est pourquoi, il faut être prudent dans le choix de ces indicateurs. Utiliser des indicateurs qui reflètent la même information, risque d'introduire beaucoup de bruit dans nos modèles et affecter leurs performances. Nous avons expliqué comment on peut faire face à ce problème et comment on peut réduire la dimensionnalité de nos données ainsi que l'interaction entre les variables explicatives.

L'existence des dérives conceptuelles dans les séries temporelles financières est un phénomène courant. En comparant des données sur des fenêtres temporelles différentes, on peut observer comment la distribution de la variable cible change. Une solution que nous proposons dans ce projet consiste à utiliser une technique d'évaluation dynamique, où chaque nouvelle instance sera prédite à partir d'un historique de données qui est dynamique. À chaque fois, la valeur collectée la plus récente est comparée à notre prédiction et intégrée dans nos données d'entrainement.

À la fin de ce projet nous avons illustré qu'il est possible de prédire les mouvements des titres boursiers en utilisant les techniques d'apprentissage automatique. La performance de nos prédictions était significativement plus élevée que si on suppose que le mouvement boursier suit une marche aléatoire. Nous avons aussi montré qu'il est possible d'implémenter des stratégies de négociation gagnantes basées sur l'apprentissage automatique et qui génèrent beaucoup plus de profits, que si on opte pour une gestion passive.

Mots-clés : Intelligence artificielle, apprentissage automatique, négociation algorithmique, séries temporelles financières, analyse prédictive, indicateurs boursiers techniques, analyse des sentiments des marchés financiers

Machine Learning techniques for predicting the trend of financial securities

Rachid MIFDAL

ABSTRACT

This thesis examines the use of machine learning techniques for predicting the trend of financial intraday time series (e.g.: minutes to hours) and explain how we could design an autonomous decision-making system to automate the trading of financial securities.

One of the challenges of financial time series is the asymmetric distribution of the price movements that introduce bias to the data used to train our models. Three methods are proposed in this thesis to deal with this problem. Among other things, resampling the data, modifying the default probability cutoffs used to generate the classifications and the use of cost sensitive learning to penalize the misclassifications of the majority class.

We have observed the existence of time dependencies in the price series of the securities used in this project. This component needs to be considered during the process of training and evaluating our models. Indeed, the accuracy of the inference made from the training data on new unseen data will depend on the similarity of the statistical properties of the two sets of data.

The existence of drifts in financial time series is a common phenomenon. By comparing data on different time windows, we can see how the distribution of the target variable changes over time. One solution we propose in this thesis is to use a dynamic rolling evaluation and forecasting technique, where each new instance is predicted from a dynamic set of data. For each new data point, the most recent collected value is compared to the model prediction and integrated into the training data.

We also illustrated that the selection of features to use in prediction (the technical indicators for this project), is a task that requires a lot of rigor. Strong multicollinearities can be found for technical indicators that are derived from the same series (e.g. price series). Using indicators that reflect the same information may introduce a lot of noise into our models and affect their performance. We have explained how we can deal with this problem and how we can reduce the dimensionality of our data as well as the interaction between the technical indicators.

At the end of this thesis, we demonstrated that it is possible to predict intraday financial time series trend by using machine-learning techniques. From a statistical perspective, the performance of our predictions was significantly better than a random guess, which contradicts the hypothesis of the random walk. We also back-tested our models to see if it is possible to implement winning trading strategies based on machine learning techniques. The results are significantly higher than the expected performance of a passive trading strategy.

Keywords: Artificial Intelligence, Machine Learning, Supervised Learning, Algorithmic Trading, Financial Time Series, Financial Predictive Analysis, Technical Indicators, Sentiment Analysis, Market Trend, Market efficiency hypothesis

TABLE DES MATIÈRES

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

Page

INTRODUCTION

Le domaine de l'intelligence artificielle connait de nos jours une très grande évolution et il risque de changer catégoriquement le monde à travers nous. La détection de la fraude, les systèmes de recommandation, la reconnaissance faciale et l'automatisation de la prise de décision, ne sont que quelques exemples parmi une variété où l'intelligence artificielle a déjà pris place. Le domaine financier n'a pas été épargné. D'ailleurs, on estime qu'environ plus de 80% des transactions boursières effectuées sur les marchés américains d'actions sont faites par des robots : négociation algorithmique(Glantz & Kissell, 2013).

Comme son nom l'indique l'intelligence artificielle est une intelligence (un ensemble de compétences) acquise par des machines dans le but de fonctionner et réagir comme des humains. L'apprentissage automatique (appelé aussi apprentissage machine) est une branche de l'intelligence artificielle qui concerne la conception et le développement d'algorithmes permettant à un ordinateur (une machine au sens large) d'apprendre à exécuter des tâches très complexes sans avoir été explicitement programmé(Koza, Bennett, Andre, & Keane, 1996).

Pour automatiser la négociation des titres financiers, il est possible d'utiliser deux types d'approches :

Un système algorithmique basé sur des règles prédéfinies : ce genre de systèmes est une version d'intelligence artificielle simple, qui utilise une série d'instructions prédéfinies qui conduit une machine à générer une prédiction ou une recommandation d'achat ou de vente. Les règles de ce genre de systèmes sont généralement produites à partir des connaissances d'experts humains. Un négociateur de titres financiers qui utilise des indicateurs techniques pour entrer dans des positions, pourra automatiser cette tâche en transformant les observations et signaux graphiques en règles et formules mathématiques.

Quand ces règles sont rencontrées, un signal d'entrée ou de sortie est généré. Ensuite, l'ordre est envoyé sur le marché pour être exécuté. L'avantage de ce genre de systèmes est qu'il est facile à monter et expliquer. Toutefois, il est très peu envisageable pour que ce genre de systèmes puisse couvrir un grand nombre de possibilités et s'adapter aux changements dans les propriétés statistiques des données quand elles évoluent au cours du temps d'une manière imprévue (dérives conceptuelles[1]).

Un système algorithmique dynamique basé sur l'apprentissage automatique : Il s'agit d'une approche alternative qui peut aider à résoudre certains problèmes des systèmes basés sur des règles. En effet, au lieu d'essayer d'imiter pleinement le processus de décision d'un expert, les méthodes d'apprentissage automatique cherchent à trouver des associations entre les données et inférer la connaissance extraite sur des données nouvelles.
Dans ce mémoire on considère la problématique à étudier comme étant une *classification* faisant partie de la famille de *l'apprentissage supervisé*, où on cherche à prédire la classe d'une variable cible, la tendance dans notre cas, en utilisant plusieurs attributs tels que les données fondamentales, les indicateurs techniques, les nouvelles des marchés financiers et les sentiments des investisseurs.

En apprentissage automatique il existe les deux approches suivantes pour modéliser ce problème :
L'approche générative où on s'intéresse à modéliser mathématiquement les probabilités conditionnelles de la distribution des données.

[1] Une dérive conceptuelle est un phénomène relié aux flux de données qui se reflète par un changement dans les propriétés statistiques de la variable qu'un modèle prédictif cherche à prédire. Les prédictions générées deviennent moins précises au fur et à mesure que les données évoluent dans le temps.

Exemple : la probabilité que la tendance soit à la hausse pour la prochaine séance de négociation étant donné que nous avons un croisement de deux moyennes mobiles 20 et 50. Cette probabilité pourra être estimée en utilisant le théorème de Bayes[2].

L'approche discriminante où on cherche à maximiser l'exactitude de la classification sans chercher à modéliser la distribution des données. Autrement dit, on estime la probabilité conditionnelle $P(Y = yi\ |X1$:t) directement à partir des données.
Où :
Y représente la variable cible à prédire,
Et X représente l'ensemble des attributs utilisées pour prédire Y.

Pour ce projet de recherche on va se limiter aux modèles discriminants qui ne dépendent que des données à utiliser et nécessitent moins d'hypothèses sur les distributions de ces données. L'autre point est qu'en pratique, la performance des modèles discriminants surpasse celle des modèles génératifs pour des données volumineuses qui arrivent en flux (Raina, Shen, Mccallum, & Ng, 2004).

Dans la partie sur la revue des modèles de prévision de ce mémoire on va discuter aussi les modèles économétriques autorégressifs classiques largement discutés et utilisés en finance quantitative. Par la suite, et en utilisant des tests de validité rétroactifs, on calculera le rendement obtenu en utilisant chacune des méthodes proposées. L'objectif est de voir si l'implémentation d'une stratégie basée sur l'apprentissage automatique mènera à la réalisation de rendements positifs.

---

[2] Ce théorème provient de la théorie des probabilités conditionnelles. Il est utilisé pour faire de l'inférence à partir des observations et des lois de probabilité de ces observations. (Source : Wikipédia)

$$\mathrm{P(A|B)} = \frac{\mathrm{P(B|A).\,P(A)}}{\mathrm{P(B)}}$$

On abordera également les problèmes rencontrés reliés à l'apprentissage à partir des flux de données financières, à savoir la distribution non équilibrée des mouvements, les dérives conceptuelles, la sélection des attributs et l'effet des nouvelles ou des comportements anormaux sur la performance globale.

CHAPITRE 1

CONTEXTE ET REVUE DE LITTÉRATURE

Ce chapitre a pour objectif de définir les concepts qui seront discutés dans ce projet de recherche. Il peut être réparti en trois parties. Dans la première, nous allons présenter le contexte global de notre projet. La seconde est consacrée aux objectifs de ce projet et les hypothèses qui étaient nécessaires pour le réaliser, ainsi qu'une introduction aux défis rencontrés. La troisième est une revue de la littérature qui consiste à résumer l'ensemble de la littérature et situer le sujet par rapport aux recherches antérieures et l'information existante.

1.1 Mise en contexte

Les séries temporelles financières (exemple : le prix à chaque 5 minutes de l'action AAPL) ne sont pas stationnaires, et en outre, des stratégies qui sont rentables à un certain moment et sous certaines conditions peuvent cesser de l'être, dès que les conditions nécessaires à leurs succès, ne sont plus rencontrées ou quand elles sont découvertes par un grand nombre d'opérateurs boursiers. D'où la nécessité d'avoir un système dynamique qui est capable de détecter ces changements comportementaux et les intégrer dans l'équation de prévision.

La prédiction des séries financières est un défi de taille. Il s'agit aussi d'une tâche très complexe, en partie à cause de la nature des données des séries financières qui se caractérisent par un comportement erratique, l'existence de bruit dans les données, les valeurs aberrantes et les données manquantes. D'autre part, par l'impact des performances intrinsèques, l'inefficience informationnelle, les annonces micro et macroéconomiques, l'influence des facteurs psychologiques et la microstructure des marchés financiers. Malgré cela, le sujet attire une grande attention, qui peut être justifiée par les récompenses potentielles qui peuvent être générées si on est en mesure de prévoir correctement le marché boursier.

Dans ce projet on cherche à tester la capacité des algorithmes d'apprentissage automatique à prédire la tendance (direction) des prix de certains actifs boursiers.

1.2 Objectifs et hypothèses de recherche

1.2.1 Objectifs du projet

Ce mémoire a pour objectif d'affirmer ou infirmer les points suivants :

- les mouvements des prix des actifs financiers forment des tendances qui peuvent être prévisibles en ayant recours à la modélisation prédictive;
- les techniques d'apprentissage automatique peuvent être utilisées pour mieux prédire les mouvements des actifs financiers transigées en bourse;
- les algorithmes d'apprentissage automatique peuvent être utilisées pour trouver les caractéristiques et les variables des séries temporelles financières, les plus importantes pour prédire la tendance de ces séries;
- en utilisant les techniques d'apprentissage automatique on peut battre les stratégies de gestion passive des portefeuilles.

Pour ce faire nous avons répartie notre travail en trois grandes tâches :

Revue de la littérature :

À travers une étude approfondie de la littérature, couvrant des sujets pertinents reliés à la prévision du marché boursier, nous avons cherché à comprendre ce que les autres recherches ont conclues et quel est le point commun entre elles. Exemple : la question sur l'efficience des marchés financiers.

Analyse expérimentale :

Où on expose les différentes démarches adoptées pour répondre aux points définis en haut. L'objectif étant d'utiliser des données historiques réelles pour expérimenter les algorithmes d'apprentissage automatique et voir si on arrive à des résultats concluants.

Recommandation d'une méthode :

À la fin de ce mémoire nous devons montrer que les techniques d'apprentissage automatique sont utiles pour la prévision des mouvements boursiers, et en conséquence recommander une méthode qui pourra être déployée en pratique, ou démontrer leur faiblesse et leurs incapacités à battre le marché et donc rejeter l'hypothèse de la possibilité de prévoir les mouvements des marchés financiers en utilisant ces techniques.

1.2.2 Les hypothèses de recherche

Plusieurs hypothèses ont été nécessaires pour la réalisation de ce projet. En voici la liste :

1. **Prévisibilité des mouvements des marchés financiers à court terme :**

   C'est l'hypothèse principale. Nous consacrons une grande partie de ce chapitre à la discuter. Cette question divise un grand nombre de chercheurs et spécialistes des marchés financiers. Nous pensons qu'il existe des poches d'inefficience informationnelle qui peuvent être attribuées à plusieurs facteurs que nous discuterons plus tard.

   Si nous admettons que les mouvements des titres boursiers sont purement aléatoires il n'y a aucune raison d'essayer de se lancer dans la prévision de ces séries. Pour ce travail, on pose comme hypothèse que la trajectoire des titres financiers fournie assez d'information pour prédire la tendance de ces séries, jusqu'à un certain niveau, pour des données financières de courte période (minute, heure à jour) et pour le type de titres à utiliser. Nos résultats à la fin devront nous conclure si cette hypothèse tienne la route.

2. **Le coût des transactions :**

   La deuxième hypothèse que nous utiliserons pour des fins de simplification est celle sur les coûts des transactions. On assume dans un premier lieu qu'il n'y a pas de coûts associés aux transactions. Notre but au début, est de tester le pouvoir prédictif des modèles à utiliser et non pas implémenter des algorithmes permettant de tester leurs rentabilités.
   Une comparaison entre les stratégies à tester sera faite à la fin, où nous allons intégrer ces coûts pour pouvoir quantifier la rentabilité générée de ces algorithmes. Une estimation de ces coûts par transaction est utilisée pour cette fin.

3. **Les mécanismes d'échange et la microstructure des marchés financiers :**

   Seuls les actifs boursiers liquides sont couverts par ce travail et on assume qu'il est possible de les échanger en tout moment avec le prix historique de fermeture. Cette hypothèse ne tient pas compte de la microstructure des marchés, où il existe en général, une différence entre le prix attendu d'une transaction et le prix auquel la transaction est réellement exécutée ('slippage'). Ce phénomène est accentué durant les périodes de forte volatilité lorsque des ordres du marché sont utilisés, ou également quand il n'y a pas assez de demande pour maintenir le prix attendu.

   Pour faire face au biais de rentabilité qui sera généré dans nos résultats théoriques et s'adapter à la réalité du marché, nous appliquerons à la fin un facteur d'ajustement pour pénaliser la rentabilité calculée. Exemple : pour ouvrir des positions longues, un facteur de coût sera intégré en ajoutant aux prix utilisés, une moyenne de la différence entre le prix 'Bid' et le prix 'Ask' de l'actif utilisé. La même chose est faite pour fermer ces positions mais en déduisant ce facteur du prix de fermeture.

4. **Les ventes à découvert :**

   On va supposer ici qu'il est possible d'effectuer des ventes à découvert à tout moment. Autrement dit, on ne veut pas se limiter juste à prédire un mouvement haussier pour acheter le titre et le vendre par la suite.

Mais aussi, quand le prix tend à avoir un mouvement baissier, on peut commencer par l'emprunter pour le vendre et le racheter après, à un prix que nous pensons être inférieur à celui de la vente.

En pratique, la vente à découvert n'est pas permise pour certains titres ou pendant certaines périodes de crises.

5. **Seuil de profit :**
   La dernière hypothèse est reliée au niveau de rendement espéré pour chaque transaction. En d'autres termes, On ne veut pas entrer dans des positions si on ne s'attend pas à ce que le prix varie d'un certain pourcentage (seuil) pendant un intervalle de temps prédéterminé. Exemple : une variation d'un cent à la hausse pour une action qui vaut 100$ n'est pas du tout intéressante pour la considérer comme un mouvement haussier. Nous utiliserons cette logique dans la définition de la tendance que nous utiliserons comme variable cible pour entrainer nos modèles. On abordera plus en détail cette question dans la partie sur la méthodologie et l'analyse expérimentale.

1.2.3 Les problèmes de recherche

Les données utilisées sont des séries financières de contrats à termes représentants des transactions sur des indices boursiers et des matières premières qui ont une grande liquidité.
Certains défis ont été rencontrés lors de l'analyse et la prédiction de ces séries financières et méritent d'être discutés, à savoir :

1. **Le traitement des données** : nettoyage, agrégation, balancement des données, transformation, imputation des données manquantes, création et sélection d'attributs.

2. **Fenêtre temporelle des données financières** : c'est la périodicité d'agrégation des données. Elle pourra varier dépendamment de l'objectif de la modélisation prédictive, de quelques millisecondes ou minutes à quelques jours. Dans le but d'avoir plus de flexibilité en fonction du choix de ce paramètre, nous avons créé une fonction qui permet d'agréger

les données en fonction de ce paramètre. Exemple : si la fenêtre est égale à 5 minutes, les données utilisées dans l'entrainement de nos modèles et la prédiction des résultats seront d'une granularité de 5 minutes au lieu d'une minute.

3. **Le processus de modélisation et d'apprentissage** : plusieurs points sont à considérer au niveau de la construction des modèles prédictifs. Entre autres, le sur-apprentissage et le sous-apprentissage, les dérives conceptuelles, sélection des paramètres des modèles prédictifs à utiliser, le biais des modèles et l'interprétabilité des résultats.

4. **L'évaluation des modèles prédictifs** : dans le but de quantifier la validité et la précision des prédictions, plusieurs métriques et techniques d'évaluation peuvent être utilisés. Le choix du modèle à utiliser pourra être basé sur plusieurs points qu'on discutera plus tard au chapitre 4.

Tous ces points vont être abordés en détails dans la section sur l'analyse expérimentale.

1.3 Revue de littérature

Dans cette section, nous présentons un résumé de l'état d'art en mettant en perspective les idées, les théories et les concepts, qui étaient nécessaires à comprendre pour la réalisation de ce projet.

Nous commençons par aborder la question sur l'efficience des marchés financiers, après dans les sections suivantes nous présentons les principales méthodes de prévision utilisées, tel que l'analyse fondamentale et l'analyse technique. Ensuite, on finira ce chapitre par une brève description des modèles économétrique linéaires.

1.3.1 Efficience des marchés financiers

1.3.1.1 Définition de l'efficience du marché

La théorie financière traditionnelle a toujours considéré que les marchés financiers sont efficients. C'est-à-dire, que toute l'information disponible est totalement reflétée dans le prix des actifs financiers et que les possibilités de déséquilibre du marché, menant à la réalisation de profits sont impossibles.

La littérature financière accorde beaucoup d'intérêt à cette question. D'une part, il y' a ceux qui défendent l'idée que les prix des actifs financiers varient de façon aléatoire et supportent l'hypothèse que dans un marché efficient, aucune stratégie d'investissement ne pourra générer plus de profits que la gestion indicielle passive à long terme. D'autre part, plusieurs autres chercheurs défendent la prévisibilité des séries financières et cherchent à affirmer qu'il est possible «de battre le marché pour un niveau de risque proportionnel [3]».
Les travaux de Louis Bachelier (1900) stipule que la trajectoire des prix des actions n'est qu'une succession de pas aléatoires (Bachelier, 1900). Ainsi, est née " l'hypothèse de la marche aléatoire des mouvements des cours boursiers". Selon cette hypothèse, l'espérance mathématique d'un spéculateur est nulle.

Dans son analyse empirique sur l'efficience du marché, réalisée par Fama en 1965, il définit pour la première fois, c'est quoi un marché efficient. Il conclut que les prix des actifs boursiers suivent une marche aléatoire et qu'ils représentent le mieux la valeur intrinsèque réelle. Les prix s'adaptent instantanément à l'arrivée de nouvelles informations (Fama, 1995).

Samuelson a fourni le premier argument économique formel pour démontrer que "les prix des titres cotés en bourse fluctuent aléatoirement" (Samuelson, 2016). Son argument est basé sur le concept d'une martingale, plutôt que sur une marche aléatoire (comme dans Fama, 1965).

---

[3] C'est-à-dire qu'il est possible pour des investisseurs ayant la même aversion au risque, d'avoir des rendements différents. Ici, on veut défendre l'idée selon laquelle, un investisseur ayant accès à des outils quantitatifs avancés pourra avoir un meilleur rendement pour un niveau de risque comparable que d'autres investisseurs passifs.

Harry Roberts (1967) a mis en évidence la difficulté de tester l'efficience en une seule forme (globale) et propose de la diviser en trois formes en fonction du type d'information que le marché prend en considération pour refléter les prix actuels des titres.

Il fait la distinction entre les trois formes d'efficience suivantes :

1. **La forme faible** : Toutes les données sur les prix historiques sont prises en compte dans les prix actuels.

2. **La forme semi-forte** : Toutes les informations pertinentes disponibles publiquement sont reflétés dans les prix actuels, annonces des résultats, opérations fusions-acquisition.

3. **La forme forte** : Toutes les informations pertinentes, même initiés, sont considérées dans les prix actuels des titres.

Si les marchés sont efficients, ceci revient à dire qu'il est impossible de faire des profits anormaux et que les acteurs agissent de façon rationnelle. Une implication directe de cette hypothèse est la formation de l'équilibre, c'est-à-dire qu'en aucun moment le marché va sous-estimer ou surestimer la valeur du titre transigé. Toujours selon cette hypothèse, toute tentative d'analyser ou prédire le mouvement des titres financiers est une perte de temps. La seule loi qui affecte le rendement espéré est la proportion du risque pris par rapport au rendement espéré. Autrement dit, un investisseur qui génère plus de profits est quelqu'un qui prend nécessairement plus de risques.

1.3.1.2 Définition d'une marche aléatoire

Une marche aléatoire est un processus aléatoire où les changements d'un état à un moment donné, à un état subséquent sont indépendants. C'est à dire, qu'à chaque instant, la direction de la marche dépend seulement de son état actuel, mais pas de son passé. On dit qu'il s'agit d'un processus sans mémoire.

Par exemple, lancer une pièce de monnaie non biaisée ‘n’fois est une marche aléatoire de n pas. Supposons qu’à chaque lancer on assigne +1 si le résultat est pile et -1 si le résultat est face. Alors, on ne peut pas prédire le résultat au $(n+1)^{éme}$ lancée en se basant sur la séquence de résultats des n lancers précédents. Les résultats des lancers sont donc dits, indépendants.

Mathématiquement, soit :

- *n* est le nombre de lancers effectués;
- $[X1, X2, \dots Xn]$ est une suite de variables aléatoires indépendantes qui représentent le résultat obtenu à chaque essaie i (*Xi* prend les valeurs +1 ou -1);
- p est la probabilité que le résultat soit face $= 1/2$;
- q est la probabilité que le résultat soit pile $q = 1 - p = 1/2$;
- S est le score obtenu après avoir fait n lancers $S = \sum_{i=0}^{n} Xi$; $X0 = 0$.

On définit : $S_{t+1} = S_t + X_{t+1}$

L’espérance mathématique de cette expérience en faisant n essaies :

$$E[S] = E\left[\sum_{i=0}^{n} Xi\right] = \sum_{i=0}^{n} E[Xi] = \sum_{i=0}^{n} P[Xi].Xi = 0\,,\ \ 0 < i < \infty \qquad (1.1)$$

Ce qui est équivalent à dire que l’espérance mathématique est nulle en cas d’équiprobabilité et où le gain attendu est égal à la perte à laquelle on est exposé.

Le graphique en bas représente des exemples de marches aléatoires que nous avons générés en utilisant les définitions posées précédemment.

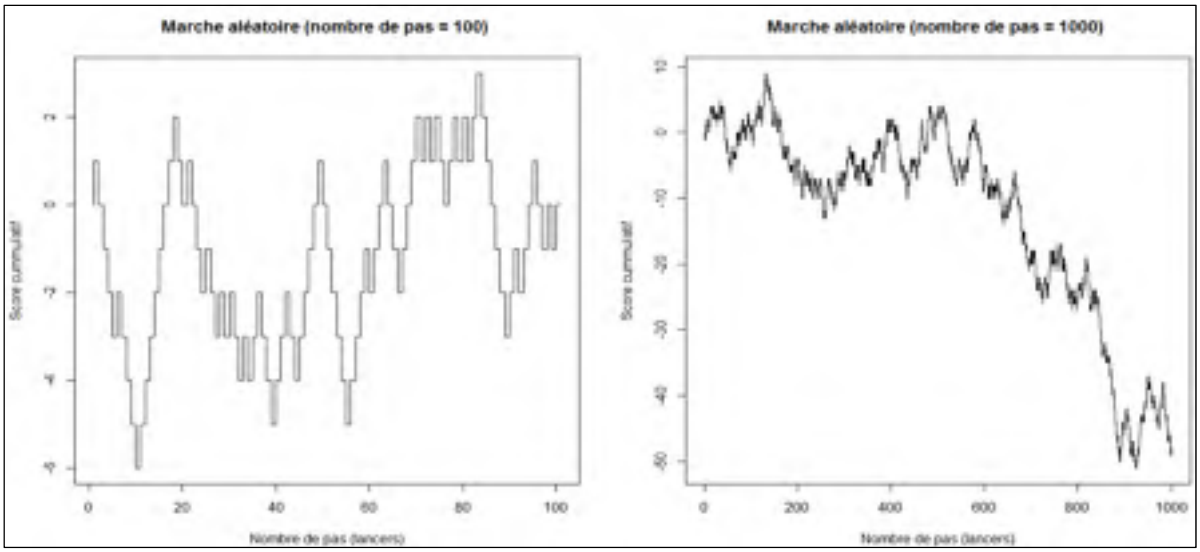


Figure 1.1 Exemple de deux marches aléatoires artificiellement générées de 100 et 1000 pas chacune

Remarque : Pour les deux exemples en haut nous avons posé que l'origine (le point de départ) de la marche est égal à zéro. Il est possible de prendre n'importe quelle autre valeur.

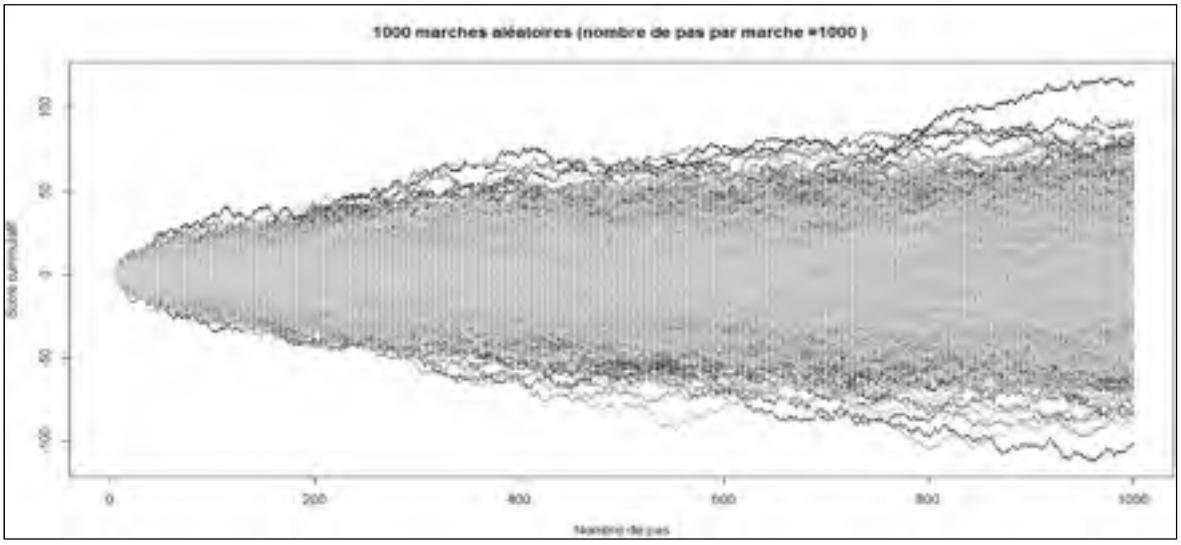


Figure 1.2 Exemple de 1000 marches aléatoires artificiellement générées de 1000 pas chacune

Nous avons utilisé un générateur de nombres aléatoires pour aboutir aux représentations dans la figure 1.2.

Il s'agit de 1000 marches aléatoires de 1000 pas chacune. En effet, à chaque fois qu'on refait l'expérience, on obtient une nouvelle marche aléatoire avec une allure différente. En faisant l'exercice 1000 fois, nous avons obtenus l'ensemble des marches aléatoires en considérant que toute les marches commencent au même point de départ et qui est nul.

On voit que la distribution des 1000 marches est centrée autour de zéro, qui est l'espérance mathématique de l'ensemble des marches générées. Ceci peut être prouvé en utilisant le théorème de la loi faible des grands nombres[4]. En d'autres termes, si le nombre de fois où on refait une expérience aléatoire est tellement grand, pour une marche aléatoire équilibrée et ayant comme origine la valeur zéro, le gain cumulatif qui sera obtenu tend vers zéro.

De façon générale, si on veut calculer X, le nombre de piles obtenus en faisant n lancers indépendants. Il s'agit d'une variable aléatoire qui suit une distribution binomiale de paramètres (n : # d'essai et p : la probabilité d'avoir un succès).

La probabilité d'avoir *K* succès en faisant *n* essaies :

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1.2}$$

Le nombre moyen espéré de succès en faisant n essaie est :

$$E[X] = n.p \tag{1.3}$$

Un partisan de l'hypothèse de l'efficience du marché, pourra utiliser le même exemple pour essayer de prouver l'impossibilité de prévoir les mouvements d'un titre en bourse.

---

[4] L'espérance mathématique d'une suite d'essaies (n) indépendants de même distribution est égale à la moyenne arithmétique quand n tend vers l'infini.

Si un marché est efficient, la variation du prix d'un titre est indépendante de ses variations historiques, parce que les changements des prix historiques sont déjà reflétés dans le prix actuel. Il peut argumenter que prédire le prochain mouvement d'un titre boursier est équivalent à lancer une pièce de monnaie. Autrement dit, si on admet que les cours boursiers suivent une marche aléatoire, il est imprévisible de savoir où va se situer le prix après n prochains pas (unités de temps : minutes, heures, jours, etc.).

L'hypothèse sur la marche aléatoire a précédé de 70 ans celle sur l'efficience du marché. Toutefois, Il s'agit d'un résultat de l'hypothèse de l'efficience du marché et non pas une condition nécessaire pour sa validité.

1.3.2 Critique de l'hypothèse de l'efficience du marché

Dans la section précédente nous avons essayé de relater les arguments derrière la pensée que les marchés financiers sont efficients et purement aléatoires. Dans la section suivante nous donnerons nos propres arguments à propos de cette hypothèse.

Les séries des prix des actifs financiers ne sont pas totalement régi par le hasard. Si les mouvements des actifs boursiers sont totalement aléatoires. Alors, pourquoi les gestionnaires de fonds ou les négociateurs des titres financiers à haute fréquence sont-ils parmi les professionnels les mieux payés au monde. Pourquoi même les banques d'investissement investissent des sommes colossales pour l'innovation et la création de stratégies gagnantes ? Plusieurs facteurs laissent planer le doute autour de la théorie sur l'efficience des marchés financiers. De plus, l'éventualité que l'information pertinente ne soit pas utilisée de façon adéquate, n'est pas prise en considération dans l'hypothèse de la marche aléatoire des mouvements des cours boursiers.

D'ailleurs, la performance exceptionnelle et récurrente réalisée par certains investisseurs ou certains spéculateurs est un contre-exemple parfait pour mettre en doute la validité de cette hypothèse. Ces anomalies sont attribuées à la stratégie adoptée.

Par exemple, le célèbre investisseur Warren Buffet attribue son succès à une analyse fondamentale approfondie des sociétés où il investit. Il achète les actions des compagnies sous-évaluées et qui ont un grand potentiel de croissance à long terme. Sur une période de 52 ans, Buffet a généré des profits 155 fois plus élevés que l'indice boursier S&P 500 avec un taux de rendement annuel moyen de 20.8%(Miles, 2004). Pendant 48 ans d'activité, il a réalisé des bénéfices 46 fois (contre 37 fois pour l'indice S&P 500) et il a réussi à battre le marché 39 fois. Mathématiquement, ceci n'est probable que dans 4 fois sur 1 milliard (10^{12})(Frazzini, Kabiller, & Pedersen, 2013).

La performance est tellement étonnante que quand on interpose le rendement annuel de l'action de la compagnie détenue par Buffet « Berkshire Hathaway » avec le rendement du S&P, le dernier a une allure d'une ligne flatte.

Figure 1.3 Comparaison entre le rendement du S&P 500 et Berkshire Hathway
Tirée de Yahoo finance

Un autre exemple est celui du mathématicien James Simon qui a fondé le fond d'investissement Renaissance et qui utilise des techniques quantitatives et computationnelles avancés pour faire de la négociation algorithmique. Le fond a généré 35% de rendement annuel en moyenne depuis sa création en 1988. Entre 1994 et 2014 le fond a généré 71% de rendement annuel moyen. Le fond emploi des chercheurs quantitatifs qui manipulent des quantités de données énormes en utilisant des technologiques hautement efficaces et évolutives pour le calcul et l'exécution d'algorithmes quantitatives très complexes.

1.3.2.1 L'existence d'anomalies

Une rentabilité est dite anormale si elle n'est pas expliquée par l'évolution générale du marché ou bien encore de la rentabilité attendue sur le titre. Peu importe la forme d'efficience considérée, des irrégularités ont été observées sur les marchés boursiers mondiaux de façon persistante, nous citons quelques-unes dans ce qui suit.

L'effet de janvier : les actions qui ont sous-performé au quatrième trimestre de l'année précédente ont tendance à surperformer en janvier. Ceci peut être justifié par la compensation de l'impôt sur les gains en capital que représente les pertes pour des actions sous performantes que les investisseurs cherchent à abandonner en fin d'année. La pression sur la vente est accrue de telle sorte que l'action devient sous-évaluée (Rogalski & Tinic, 1986).

L'effet du jour de la semaine : La recherche a montré que les titres financiers tendent à se déplacer davantage les vendredis que les lundis, et qu'il y a un biais vers des performances positives les vendredis. Ce n'est pas un écart énorme, mais c'est un écart persistant (Cengiz, Bilen, Büyüklü, & Damgacı, 2017).

Les fusions-acquisitions : Plusieurs études ont démontré que, sur le court terme, les titres pour lesquelles une fusion a eu lieu, se caractérisent par des performances supérieures comparées à des actions de sociétés similaires.

Ce rendement anormal est d'autant plus prononcé si la fusion a été réalisée par une offre publique financée en argent, et non en actions (Hong, Kaplan, & Mandelker, 1978).

Les actions des petites firmes performent mieux que celles des grandes compagnies : Ceci peut être expliqué par le cycle de vie de ces compagnies. Les possibilités de croissance d'une petite compagnie sont beaucoup plus importantes que celles d'une grande. Un partisan de la pensée financière classique pourra justifier cette anomalie par le risque que représente une petite compagnie par rapport à une grande. En fait il pourra dire qu'il est normal de s'attendre à avoir plus de rendement en investissant dans une petite compagnie, car le risque est plus élevé aussi.

1.3.2.2 Les tests d'autocorrélation des rendements

En même temps que la découverte d'anomalies, les progrès de la modélisation statistique, notamment les tests statistiques d'autocorrélation, ont permis de tester et mettre en question la validité de l'hypothèse de la marche aléatoire des mouvements boursiers. Ces tests ont démontré l'existence d'autocorrélations positives dans les mouvements des prix. C'est-à-dire que quand le prix d'un titre varie dans une direction, il y a une forte probabilité qu'il continue dans la même direction. Autrement dit, la direction ou la tendance du prix dépend de ses mouvements historiques, ce qui est contradictoire avec l'hypothèse de la marche aléatoire.

Le premier test qui a été développé est celui de BOX et Pierce (1970)(McLeod, 1978). Ce test consiste à examiner les résidus (ε) obtenus en comparant une série de données à une série de nombres aléatoires. S'il existe une autocorrélation entre ces résidus, alors on peut conclure que la série ne suit pas une marche aléatoire. Une autre approche qui est la plus couramment utilisée pour tester cette hypothèse, repose sur le test de Dickey-Fuller où on cherche à tester si une racine unitaire est présente dans les séries testées(Kwiatkowski, Phillips, Schmidt, & Shin, 1992).

Mathématiquement, ceci consiste à faire le test d'hypothèses suivant :

$$H0 : \rho = 1 \text{ Versus } H1 : \rho < 1 \quad (1.4)$$

Dans l'équation suivante :

$$y_t = \rho . y_{t-1} + \mu_t \quad (1.5)$$

L'existence d'une racine unitaire ($\rho = 1$) signifie que la série n'est pas stationnaire et que la trajectoire des prix change en permanence. Cela implique que les mouvements du prix sont dus à des chocs aléatoires qui ne peuvent pas être prédits sur la base d'information des prix historiques. Cependant, l'existence d'une racine signifie que la série est stationnaire.

Lo et MacKinlay (1988) ont démontré l'existence de fortes autocorrélations pour des données hebdomadaires et mensuels des indices boursiers. Ils ont conclu que le comportement stochastique des rendements de ces indices ne suit pas une marche aléatoire (Lo & MacKinlay, 2002).

L'analyse empirique de Jegadeesh (1990) a révélé l'existence d'une autocorrélation négative très significative des rendements boursiers mensuels des actions individuellement, mais une forte autocorrélation positive à douze mois (Jegadeesh, 1990).

Zhou (1996) a trouvé que les données sur les rendement des taux de change à haute fréquence ont une autocorrélation négative très élevée (Zhou, 1996). Ceci a été soutenu par (Dacorogna et al. 2001); (Gençay, Dacorogna, Muller, Pictet, & Olsen, 2001) qui sont arrivés à la même conclusion en utilisant des données d'une minute.

Ahn et al. (2002) ont examiné les autocorrélations quotidiennes des indices boursiers et ils ont constaté qu'ils sont positifs (Ahn, Jo, & Lee, 2002).

Toutes ces anomalies et études expérimentales arrivent à la même conclusion, que les cours boursiers sont auto-corrélés et donc, il est possible de prédire la direction du prix d'un titre en se basant sur ses données historiques. Dans la section suivante, nous allons voir les techniques qui peuvent être utilisée pour faire ces prédictions.

1.3.3 Méthodes de prévision des actifs financiers

Il existe trois approches pour prévoir la direction des cours boursiers. À savoir l'analyse fondamentale, l'analyse technique ou graphique et l'analyse purement quantitative. Le but de ces analyses est le même, toutefois, elles reposent sur des bases différentes et elles peuvent être combinées pour affiner la négociation des titres boursiers.

Les partisans de l'analyse fondamentale se tournent vers l'étude des facteurs micro et macro-économiques pour expliquer le changement dans les prix. L'idée est qu'en examinant les indicateurs fondamentaux macro-économiques tels que l'inflation ou le taux de chômage et micro-économiques comme les ventes, les niveaux d'endettement et les bénéfices on devrait être en mesure d'avoir une bonne perspective sur la tendance du prix future à moyen et à long terme.

De l'autre côté, ceux qui croient plus à l'analyse technique, pensent que les données historiques sur l'activité d'un titre boursier tels que les rendements historiques, les cours des actions et les volumes des transactions fournissent assez d'informations pour pouvoir prédire ses mouvements futurs. Ils argumentent que les séries chronologiques financières reflètent toute l'information sur les comportements et anticipations des investisseurs. Ces séries sont le fruit de la rencontre de l'offre et la demande et finissent par former des tendances haussières, neutres ou baissières.

Les méthodes quantitatives utilisent la modélisation mathématique, l'économétrie et des techniques computationnelles avancées pour l'analyse et la prévision des mouvements des séries financières et l'implémentation des stratégies.

La négociation quantitative englobe l'identification de la stratégie, les tests rétroactifs de profitabilité, l'optimisation de l'exécution des ordres et la gestion des risques.

À part les publications académiques, les publications sur les méthodes quantitatives ne sont pas très fréquentes. Ceci pourra s'expliquer par le fait que les détenteurs des stratégies ou techniques gagnantes n'ont pas d'incitatifs à les partager.

1.3.3.1 Analyse fondamentale

La plupart des investisseurs qui désirent investir en bourse à long terme commencent par l'analyse fondamentale pour l'évaluation d'une société, d'une action, d'un contrat à terme ou du marché dans son ensemble. Ce type d'analyses cherche à déterminer la valeur intrinsèque (théorique) d'un actif financier pour la comparer à sa valeur marchande. Cette comparaison permet de savoir s'il est avantageux d'investir dans l'actif car il est sous-évalué et s'attendre à le voir atteindre sa valeur intrinsèque dans le futur, ce qui représente une opportunité de gain. À l'opposé, si la valeur actuelle en bourse du titre est supérieure à sa valeur intrinsèque, alors le titre est surévalué et il faut s'attendre à une baisse de sa valeur.

Différents ratios financiers et boursiers sont utilisés en analyse fondamentale dans le but d'évaluer des investissements potentiels.

Les fondamentalistes divisent l'analyse fondamentale en quatre catégories :

- **l'analyse économique** : où il faut commencer par analyser l'environnement économique dans son ensemble, en utilisant les indicateurs économiques globaux tel que le PIB, le taux de chômage, les taux d'intérêt, la production industrielle, etc. L'objectif est de déterminer l'état de santé de l'économie dans son ensemble et identifier des secteurs ayant des perspectives d'évolution importantes.

- **l'analyse sectorielle** : consiste à étudier les spécificités économiques et concurrentielles du secteur visé par l'analyse, en étudiant ses caractéristiques telle que la distribution des parts de marché, l'évolution du secteur, les concurrents et les barrières à l'entré. Ce qui permet de connaitre le potentiel du secteur et son évolution

- **l'analyse financière** : l'objectif de cette étape est de valoriser l'entité étudiée et la positionner par rapport à ses concurrents en analysant ses documents comptables (bilan, compte de résultats, etc.) et en comparant ses ratios financiers avec ceux de la concurrence (exemple : rentabilité, endettement)

- **l'analyse boursière** : en utilisant les ratios boursiers tel que le ratio cours bénéfice, la valeur comptable par action, bénéfice par action, etc., déterminer si l'action est sous-évaluée ou surévaluée en comparant sa valeur marchande à sa valeur théorique et son secteur d'activité.

Les modèles d'évaluation comptables

1. **Le modèle d'actualisation des dividendes**

   Cette méthode a été élaborée par Gordon et Shapiro (Abarbanell & Bushee, 1997) et considère que le prix d'une action est égal à la valeur présente des dividendes espérés, qui sont supposés augmenter à perpétuité avec un taux 'g'.

$$P0 = \sum_{k=0}^{\infty} D0 \frac{(1+g)^t}{(1+r)^t} \qquad (1.6)$$

$$P0 = D0 \frac{(1+g)}{(r-g)} = \frac{D1}{(r-g)} \qquad (1.7)$$

   Où :

   P0 est la valeur présente de l'action,

   D0 est le dividende payé par la compagnie au moment de l'évaluation,

D1 est le dividende de l'année prochaine,
g est le taux de croissance à perpétuité des dividendes,
r est le coût du capital.

Ce modèle pose certains problèmes au niveau des hypothèses utilisées et ne pourra pas facilement refléter la valeur réelle de l'action. On peut citer entre autres, la sensibilité du prix calculé au taux de croissance utilisé, la croissance constante et infinie des dividendes et qui est supposée être inférieure à celle du coût du capital.

2. Le ratio cours / bénéfice

Le ratio cours / bénéfice ($\frac{P}{E}\ ratio$) résulte à la fois des variations du cours de l'action et du bénéfice par action. Le prix de l'action selon cette méthode est dérivé de façon simple à partir de l'équation suivante :

$$P0 = BPA_1\ \frac{P}{E}\ ratio \tag{1.8}$$

Où :
BPA_1 est le bénéfice par action espéré pour la prochaine période.

Chaque trimestre, lorsque les sociétés publient leurs résultats, une variation importante du bénéfice par action peut augmenter ou diminuer le ratio cours / bénéfice. Le facteur le plus important, cependant, est le changement dans les attentes des investisseurs, qui se reflète dans le cours de l'action.

Ce modèle repose sur deux hypothèses peu réalistes. La première suppose que les compagnies peuvent faire des profits de façon perpétuelle. Donc, dans le cas où la compagnie fait des pertes, le modèle n'est pas applicable. Et la deuxième est qu'il est possible de prédire la valeur future du bénéfice par action.

3. **Le modèle d'actualisation des flux de trésorerie (Discounted Cash-Flow (DCF))** Il s'agit d'un modèle qui utilise les flux de trésorerie libres qui ne sont autres que les flux monétaires restant après la déduction des dépenses d'exploitation et des dépenses en capital. Pour les actualiser, le modèle utilise un coût moyen pondéré des capitaux qui représente l'attente du rendement des prêteurs et des investisseurs.

   Dans son étude de 1992, Penman(Abarbanell & Bushee, 1997) a mentionné que malgré l'absence de son utilisation par de nombreux chercheurs et praticiens, il s'agit de l'un des modèles d'évaluation comptable les plus importants. Néanmoins, dans son étude, il a montré que son application dans la pratique n'est pas claire.

Résultats de l'analyse fondamentale

Plusieurs études ont été publiées visant à vérifier le pouvoir de l'analyse fondamentale à prévoir correctement le cours d'une action. Ces études ont cherché à tester la capacité relative des bénéfices et des flux de trésorerie pour déterminer la valeur intrinsèque future.

En effectuant une série d'analyses statistiques au cours de la période entre 1988 et 2000, plusieurs de ces études ont conclu que la prédiction de la valeur intrinsèque des actions est possible en utilisant des modèles qui reposent sur des ratios financiers et le coût de l'équité (Cheung, Chung, & Kim, 1997; Chung & Kim, 2001; Lo & MacKinlay, 2002).

L'impact des informations fondamentales sur la valeur des actions

Les informations utilisées pour l'analyse fondamentale proviennent des rapports financiers publiés par les compagnies chaque trimestre. Le graphe dans la figure 1.4, montre comment la valeur de l'action Microsoft a varié durant la période du mois de Juillet 2017 et jusqu'à la fin du mois de septembre 2018. Les flèches en rouge et en vert montrent le mouvement du cours de l'action le jour de la publication des résultats de la compagnie.

Figure 1.4 Impact de l'annonces des résultats sur le cours de l'action de Microsoft

On voit comment le prix de l'action fluctue de façon anormale entre le jour de l'annonce des résultats (en général après la fermeture ou avant l'ouverture des séances de négociation) et le jour suivant. On peut facilement noter l'existence d'écarts importants entre le prix de la fermeture de la séance du jour où les résultats sont publiés et le prix d'ouverture du jour suivant. Par exemple, le 26 Octobre 2017 Microsoft a publié un bénéfice par action de 0.84$ tandis que les estimations des analystes financiers étaient aux alentours de 0.72$. Soit une surprise de plus de 17%. Le même jour, le prix de fermeture de l'action (juste avant la publication des résultats) était de 78.76$. Le jour d'après, l'action a affiché un prix d'ouverture de 84.37$, soit une augmentation de plus de 7%.

En fait, Les investisseurs se précipitent pour acheter l'action dans le cas où les annonces sont positives ou dépassent les attentes prévues et cherchent à se débarrasser des titres dans le cas où les annonces sont négatives ou inférieures aux prévisions des analystes.

Les annonces macro-économiques impactent aussi directement ou indirectement le cours des titres boursiers. Ces annonces concernent l'inflation, les politiques monétaires et les taux d'intérêt. Ces annonces vont en général impacter le marché dans son ensemble. Plus un titre est corrélé avec le marché, plus l'impact sera important.

Toutefois, les rapports et les annonces macro-économiques sont publiés périodiquement. Donc, ils sont peu utiles lorsqu'il s'agit de déterminer les prix intra-journaliers. Cependant, il est possible que les marchés affichent une grande volatilité durant des séances de négociation où des annonces sont faites.

L'examen des principes fondamentaux n'est pas donc le seul facteur suffisant pour décider comment un actif financier fluctuera surtout quand il s'agit de prédire le prix à court terme

1.3.3.2 L'analyse technique

L'analyse technique (AT), appelé aussi l'analyse graphique est totalement différente de l'analyse fondamentale, Elle ne prend pas en considération la valeur intrinsèque ou les autres facteurs fondamentaux. L'analyse technique cherche à déterminer le moment opportun pour acheter ou vendre un titre. Seul l'historique des prix et des volumes et leurs variations antérieures importent.

Le but de l'analyse technique est la prévision de la direction future des cours des titres boursiers en analysant les graphiques des prix. L'idée de base derrière l'AT est que le prix de l'action reflète l'ensemble de l'information disponible, et que les prix varient en formant des tendances. Les prix futurs sont plus susceptibles de continuer dans la direction de la tendance que l'inverse. Ces tendances sont dues à un déséquilibre entre l'offre et la demande provoquant l'augmentation ou la baisse dans le cours des actions.

Au lieu d'interpréter, calculer et analyser les ratios fondamentaux d'un rapport financier, les spécialistes de l'analyse technique examinent les variations des prix et des volumes en utilisant des indicateurs techniques. Ces derniers sont des modèles mathématiques calculés à partir de l'historique des prix et des volumes et sont utilisés pour fournir des signaux pour prévoir les variations futures des prix. L'analyse technique s'applique à tout type de marché financier où nous avons accès aux flux de données sur les transactions effectuées : actions, contrats à terme, devises, taux d'intérêt, matières premières, etc.

L’analyse technique utilise des indicateurs techniques qui sont dérivées à partir des données des prix et des volumes des transactions boursières. Ces données sont granulaires et peuvent être agrégées en fonction de la fréquence désirée pour les analyser. Contrairement à l'analyse fondamentale qui s'appuie sur des informations de longues périodes et donc ne pourra être utilisée que pour des prévisions à long terme, l’analyse technique est utilisée davantage pour le court terme, à cause de la disponibilité et la granularité des données à court terme. Il est possible de combiner l’analyse technique avec l’analyse fondamentale pour évaluer les investissements à long terme.

Résultats de l’analyse technique

Plusieurs études ont été menées dans le but d’évaluer le pouvoir de l’AT à prédire les mouvements des actifs financiers.

Lo et al. (2000) ont analysé les résultats de l'analyse technique sur les marchés d’actions américains de 1962 à 1996 et ils ont constaté que plusieurs indicateurs techniques peuvent être utilisés pour l’analyse et la prédiction des séries temporelles financières (Lo, Mamaysky, & Wang, 2000).

Brock et al. (1992) ont testé plusieurs règles d’analyse technique sur un historique de 90 ans de cours boursiers quotidiens faisant partie du Dow Jones, et ils ont montré que ces règles étaient capables de générer plus de profit que le marché (Brock, Lakonishok, & LeBaron, 1992).

(Neely & Weller, 2001) ont fait recours à la programmation génétique pour générer des règles de négociation. Ces règles ont affiché une performance très élevée sur les marchés des taux de change.

(Fernandez-Rodrıguez, Gonzalez-Martel, & Sosvilla-Rivero, 2000) ont utilisé l'apprentissage automatique, en choisissant les réseaux de neurones comme modèles prédictifs et les indicateurs techniques comme données d'entrée, pour prévoir les prix sur le marché boursier de Madrid. Ils ont découvert que cette stratégie de négociation surpasse celle d'une stratégie passive pour des marchés baissiers et stables, mais non haussiers.

L'analyse techniques a démontré aussi une grande popularité auprès des négociateurs en bourse. Ceci a été démontré par plusieurs sondages qui ont été effectués. La majorité des répondants ont été positifs a son utilisation, surtout pour le court terme (Taylor & Allen, 1992); (Lui & Mole, 1998).

1.3.3.3 Modèles Économétriques

Les modèles économétriques sont des modèles mathématiques probabilistes qui essaient de décrire les relations aléatoires entre les variables incluses dans ces modèles. Ils ont été utilisés en finance de marché pour tenter d'expliquer les autocorrélations positives des séries temporelles des prix des actifs financiers.

L'un des modèles les plus populaire en économétrie et qui a été utilisé pour la prévision des séries financières temporelles est le modèle ARMA (Autoregressive Moving Average) de Box et Jenkins (Box, Jenkins, Reinsel, & Ljung, 2015)

$$y_t = \sum_{i=1}^{p} \alpha_i \, . y_{t-1} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-1} + \varepsilon_t + \delta \tag{1.9}$$

Où ε_t sont supposés être des variables aléatoires indépendantes identiquement distribuées qui suivent une distribution normale,

$$\varepsilon_t \sim N(0, \sigma^2) \tag{1.10}$$

Il est à noter qu'il existe plusieurs extensions du modèle ARMA. On trouve entre autres, le modèle ARIMA et SARIMA.

Ce dernier tient compte de la saisonnalité des séries à modéliser.

On peut citer aussi le modèle ARCH (Autoregressive Conditional Heteroskedasticity en anglais), qui modélise la variance à un pas de temps comme une fonction des résidus à partir d'un processus de moyenne (par exemple, la moyenne nulle)

$$\sigma^2{}_t = \sum_{i=1}^{n} \alpha_i \,.\, \varepsilon_{t-i}^2 + \delta \tag{1.11}$$

Là aussi, il faudra noter qu'il y a plusieurs extensions de ce modèle. On trouve le modèle (GARCH) introduit par Bollerslev en 1986 et qui a été largement utilisé pour la prévision de la volatilité (Engle, 1982).

$$\sigma^2{}_t = \sum_{i=1}^{n} \alpha_i \,.\, \varepsilon_{t-i}^2 + \sum_{i=1}^{n} \beta_i \,.\, \sigma_{t-i}^2 + \delta \tag{1.12}$$

Les travaux de (Hamilton, 1994) et (Lo et Mackinaly, 2002) décrivent l'ensemble des modèles et techniques économétriques utilisées pour l'analyse et la prédiction des séries chronologiques financières.

Dans une étude menée sur le marché des actions japonais, Bonnie Ray, Shaw Chen and Jeffrey Jarrett ont démontré, en utilisant le modèle ARIMA et son extension ARFIMA, l'existence de composantes temporelles pouvant être utilisées dans la prédiction des prix des actions japonaises (Ray, Chen, & Jarrett, 1997).

Jarrett, J., & Schilling, J. ont conclu d'après leur expérimentation sur le marché allemand que les modèles économétriques autorégressives peuvent prédire le changement dans les rendements des actions utilisées (Jarrett & Schilling, 2008).

Toutes ces recherches ont montré l'existence d'autocorrélations positives pour les séries financières chronologiques, ce qui veut dire que les modelés autorégressifs sont utiles pour la prédiction de ces séries.

CHAPITRE 2

LES SÉRIES CHRONOLOGIQUES FINANCIÈRES

Prédire le comportement futur d'une série chronologique financière est un défi de taille. Connaître à l'avance la direction que prendra un titre, même avec une fiabilité et une précision limitée, pourra être très efficace pour optimiser la gestion d'actifs financiers. Dans ce chapitre nous allons mettre l'accent sur les propriétés et le comportement des séries chronologiques financières, le traitement nécessaire pour pouvoir les utiliser comme données d'entrée dans un système de prévision dynamique et les défis reliés à leurs utilisations tels que les valeurs manquantes, les valeurs aberrantes et la stationnarité de ces séries.

2.1 Description d'une série chronologique financière

2.2 Définition d'une série chronologique financière

Une série chronologique financière est une séquence de "n'' réalisations financières d'une variable X à travers le temps. Exemple : Les prix à chaque heure de l'action AAPL sur la bourse du NASDAQ, les prix et volumes d'échanges journaliers des taux de change, les taux de rendements mensuels, etc.

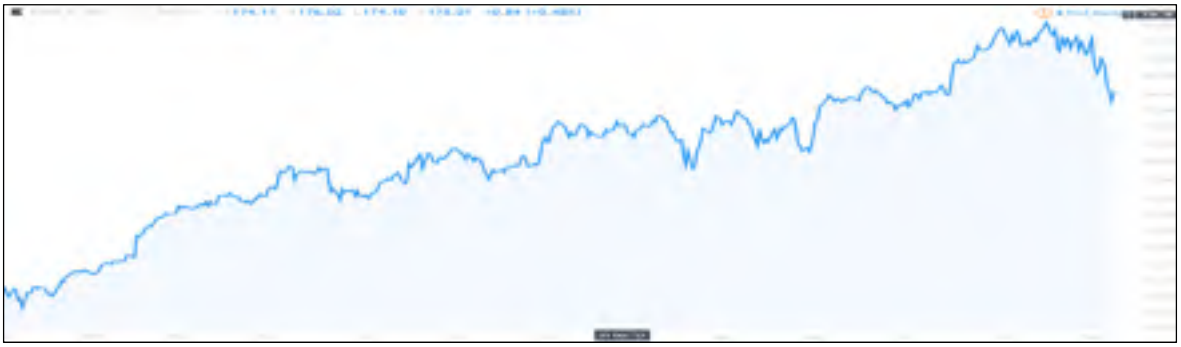


Figure 2.1 Exemple d'évolution du prix de l'action AAPL par jour
Tirée de Yahoo finance

Les séries chronologiques sont composées de données financières qui sont collectées de manière systématique à chaque instant où il y a eu un consensus sur le prix et le volume à échanger entre les offreurs et les demandeurs d'un actif financier spécifique.

Tableau 2.1 Exemple de livre d'ordres pour l'action Microsoft
Tirée de Simian Savants LLC (2019)

BTime	MMID	Cnt	Size	Bid	Ask	Size	MMID	Cnt	ATime
13:34:57	ARCA+	1409	2	52.63	52.65	7	ISLD+	2828	13:34:59
13:34:59	ISLD+	2403	6	52.63	52.67	5	BRUT+	212	13:34:59
13:34:36	EDGX+	20	10	52.63	52.67	1	ISLD+	8	13:34:59
13:34:57	ARCA+	8	1	52.63	52.67	25	NEDG+	1988	13:35:01
13:34:59	ISLD+	8	2	52.63	52.67	17	INCA+	1926	13:35:03
13:34:31	PRUS+	56	10	52.61	52.68	5	INCA+	8	13:35:01
13:34:57	ARCA+	8	1	52.61	52.68	1	SBSH+	38	13:35:00
13:34:59	BRUT+	221	10	52.60	52.70	2	ISLD+	0	13:34:56
13:35:01	NEDG+	1047	5	52.60	52.71	6	PRUS+	65	13:34:31
13:35:03	INCA-	5064	1	52.60	52.71	1	INCA-	0	13:35:01
13:34:20	SCHB-	61	1	52.59	52.72	1	SUNC-	15	13:33:45
13:34:56	BRUC+	0	5	52.59	52.72	14	NEDG+	0	13:34:56
13:34:56	LSPD+	170	5	52.59	52.72	1	BRUS+	0	13:34:56
13:35:00	SBSH+	41	2	52.59	52.73	10	LEHM+	21	13:34:36
13:35:01	INCG-	0	12	52.59	52.73	14	ARCA-	1561	13:34:57
13:33:47	WCHV+	21	1	52.57	52.75	10	FBCO+	19	13:10:45
13:34:56	ISLD+	0	2	52.57	52.75	10	PERT+	5	13:25:39
13:30:02	COWN+	0	1	52.55	52.75	1	COWN+	1	13:30:02
13:35:01	INCG-	0	1	52.55	52.75	10	SCHB-	92	13:34:20
13:34:56	BRUS+	0	2	52.50	52.76	1	ARCA+	0	13:34:57

Sur le tableau d'ordre en haut nous avons des ordres d'achat et de ventes par les offreurs et les demandeurs de l'action de la compagnie MICROSOFT. Quand le prix des offreurs (ordre de vente) égalise celui des demandeurs (ordre d'achat) sur le même marché (exemple : NASDAQ, NYSE, AMEX) l'échange peut prendre se faire, si les volumes de la demande sont pratiquement similaires à ceux de l'offre. Quand un ordre est exécuté, alors on parle de transaction.

Le déséquilibre des quantités à échanger entre l'offre et la demande fera bouger le prix du titre. En général, quand il y a plus d'offre que de demande il faudra s'attendre à une baisse du prix. Et à l'opposé, le prix augmentera quand il y a une pression des acheteurs.

Ce genre de pressions sont le fruit des nouvelles et du changement dans les attentes des investisseurs, ce qui mène à la formation des tendances. Dans le cas où il n'y a pas de déséquilibre entre l'offre et la demande, le prix du titre oscillera autour d'une valeur d'équilibre qui est déterminée par le marché.

2.3 Les composantes d'une série chronologique financière

Une série chronologique est constituée des quatre composantes suivantes : la tendance, le cycle, la saisonnalité et les fluctuations irrégulières (Bruce L. Bowerman, 1993).

La tendance reste la composante la plus importante. D'ailleurs, elle a une importance particulière en finance de marché. Car non seulement elle est l'indicateur le plus simple pour prendre une décision d'achat ou de vente, mais l'existence des tendances sur les différents marchés (actions, taux de changes, contrats à terme, etc.) est un élément important pour la répartition des actifs d'un portefeuille.

2.3.1 La tendance

La tendance (T_t): est une composante de la série chronologique qui porte sur l'évolution d'une variable (prix, volume) et qui traduit le comportement "moyen" de la série. Elle peut être croissante ou décroissante.

Pour une série de prix d'un actif boursier on peut considérer le rendement simple net sur K périodes comme une tendance (Bruce L. Bowerman, 1993)

$$\mathrm{R_{k;t}} = \frac{P_{t+k}}{P_t} - 1 \tag{2.1}$$

Où :

$R_{k;t}$ est le rendement simple d'un titre boursier entre la période t et t+k,

P_t est le prix sur le marché de l'actif au moment t (valeur marchande),

K est le nombre de périodes (ou pas) pour calculer la tendance. Il s'agit d'une variation du prix entre un moment t et t+k. cette variation pourra être positive négative ou neutre.

Hellström (1998) a suggéré la division du rendement net simple dans l'équation (2.1) par le nombre de K périodes que comporte la série dans le but de permettre une comparaison facile entre différentes séries temporelles (Hellström,1998).

$$\mathrm{Tt[k]} = \frac{R[_{k;t}]}{K} \tag{2.2}$$

Le fait de diviser le rendement par le nombre de périodes que comporte la série permet aussi de calculer une tendance moyenne sur le nombre de pas qui composent la série. L'horizon d'investissement détermine le nombre de périodes que l'on doit observer. Sur une longue période, plusieurs tendances se succèdent. En outre, une tendance pourra se poursuivre durant des semaines ou des années avant de voir un renversement et pourra évoluer de façon linéaire ou non linéaire.

Figure 2.2 L'évolution du titre SPY (S&P 500 Trust ETF)
Tirée de Yahoo finance

Le graphe en haut montre un exemple de tendance haussière linéaire pour le titre SPY (SPDR S&P 500 Trust ETF) entre Février 2016 et Décembre 2017.

2.3.2 La saisonnalité

La saisonnalité (St) correspond à l'effet périodique dans une série chronologique qui se répète à intervalles de temps réguliers. La saisonnalité peut être de fréquence intra-journalière, quotidienne, hebdomadaire, mensuelle ou annuelle.

- **saisonnalité mensuelle :**

Le plus connu de ces effets sur les marchés boursiers est l'effet de janvier ainsi nommé d'après la surperformance du mois de janvier par rapport aux autres mois de l'année. Tel qu'expliqué dans la section '1.3.2.1', c'est en fait la fiscalité qui incite les investisseurs à réaliser leurs pertes en vendant leurs positions perdantes en fin d'année pour les racheter en janvier et présenter seulement les lignes performantes dans leur clôture de comptes.

- **saisonnalité à court terme :**

Les effets de saisonnalité existent aussi sur des intervalles plus courts. L'effet « Turn of The Month » pour lequel il a été observé que les rendements lors des trois jours précédents la fin du mois et des trois jours suivants le début du mois offrent de meilleurs rendements, et l'effet « lundi » qui soutient qu'en moyenne les rendements du lundi sont inférieurs aux autres jours de la semaine. (Kayaçetin & Lekpek, 2016)

- **saisonnalité intra-journalière :**

Un autre point qui est relié d'avantage au cadre de ce projet est la saisonnalité intra-journalière. En effet, des études ont été conduites pour investiguer l'existence de la saisonnalité au cours d'une même séance de négociation sur les marchés des actions, contrats à terme et taux de change. Les résultats ont conclu à l'existence d'une saisonnalité qui est due à la microstructure des marchés, le temps d'arrivé des nouvelles et le comportement et psychologie des opérateurs boursiers.

Il est à noter aussi qu'il est facile d'observer sur les graphes boursiers intra-journaliers la saisonnalité des volumes. Pour illustrer cela, nous pouvons diviser la journée de négociation en trois plages horaires :

- de 9h30 à 12h00 : séances du matin;
- de midi à 14h00 : séances de midi;
- de 14h00 jusqu'à la clôture à 16h00 : séance de l'après-midi.

Figure 2.3 Saisonnalité intra-journalière des volumes du titre SPY

Comme on peut le constater sur le graphe en haut, le matin le volume des échanges augmente grâce aux informations publiées au cours de la nuit et avant l'ouverture de la séance. Dans la séance de midi, le volume d'échange baisse et le prix du titre ne fluctue pas beaucoup dans l'absence de nouvelles. L'après-midi le volume augmente, car les négociateurs journaliers cherchent à fermer leurs positions à court terme ou d'en créer de nouvelles avant la clôture. L'augmentation du volume fournit de la liquidité, ce qui réduit l'écart cours acheteur / vendeur (Bid/Ask spread) et réduit les risques d'exécuter des ordres à des prix démesurés. Cependant, même ces micro-saisons ont des saisons internes.

La précipitation des ordres à l'ouverture et à la clôture peut entraîner une volatilité excessive qui peut générer des risques de pertes importantes. C'est pourquoi les négociateurs professionnels recommandent d'éviter de passer des ordres au cours des 30 premières minutes et des 30 dernières minutes de la journée de négociation.

2.3.3 Le cycle

Cette composante de la série fait référence à la présence d'une certaine récurrence mais contrairement à la saisonnalité sur des durées qui ne sont pas fixes et généralement plus longues. Un exemple typique qui peut être défini comme un cycle sont les périodes où il y a eu des récessions ou des expansions.

En général, il est difficile de dissocier la tendance du cycle de la série à moins d'avoir plus de détails spécifiques. Dans le cadre de ce projet nous ne donnons pas beaucoup d'importance à cette composante du fait que notre recherche porte sur la prédiction intra-journalière.

2.3.4 Les fluctuations irrégulières

Il s'agit de la partie de la série chronologique non expliquée par la tendance ou la saisonnalité. Ces fluctuations sont en général de faible intensité et de nature aléatoire. On dit aussi des aléas, bruits ou résidus. Il est aussi important de souligner que d'autres phénomènes accidentels tels que les crashes boursiers, les attentats terroristes ou les crises politiques, peuvent notamment intervenir. Ces événements sont rares mais peuvent avoir un impact extrême sur le comportement des séries financières. Ils sont en général difficiles à prédire et nécessitent une modélisation spéciale (Piccoli, Chaudhury, & Souza, 2017).

La stationnarité est une propriété importante des séries chronologiques. La modélisation de ces séries nécessite que ces dernières soient stationnaires (TSAY, 2005). C'est-à-dire qu'elles ne contiennent pas de tendance ou de saisonnalité.

2.3.5 La stationnarité d'une série chronologique

Soit X_T une série temporelle financière. X_T est dite stationnaire (stationnarité de second degré) si :

1. $E[X_i] = \boldsymbol{\mu}$ $\quad \forall\ i = 1, 2, \ldots t$
2. Et, Var $(X_i) = \sigma^2 \neq \infty$ $\quad \forall\ i = 1,2, \ldots t$
3. Et, Cov $(X_i, X_{i-j}) = f(j)$ $\quad \forall\ i = 1,2, \ldots t\ ;\ \forall\ j = 1,2, \ldots t$

Si l'une de ces conditions n'est pas vérifiée, on parle donc de non-stationnarité.

Pour tester la non-stationnarité des séries, on procède au test de Dickey-Fuller (Mushtaq, 2011).

La stationnarité faible est une hypothèse courante dans l'analyse des séries chronologiques. Malheureusement, les séries financières ne remplissent pas souvent cette condition. Cependant, une série chronologique peut être transformée en une série temporelle faiblement stationnaire en utilisant le concept de la différenciation (TSAY, 2005).

2.3.6 Décomposition d'une série chronologique financière

Après avoir vu les principales composantes d'une série chronologique, Il sera judicieux d'étudier le comportement d'une série chronologique en isolant la tendance et la saisonnalité. Ces deux opérations s'appellent la détendancialisation et la désaisonnalisation de la série.
Une fois ces composantes éliminées, on obtient la série aléatoire ϵt des fluctuations irrégulières (résidus).

Mathématiquement :
Soit $X_1, X_2, \ldots, X_T$ une suite de variables aléatoires représentant le prix d'une action.
X_T désigne la série chronologique,
Θ est un espace de temps discret,
$t \in \Theta$ où $\Theta \subset Z$.

L'observation de la série X_T à la date t est une fonction du temps t et d'une variable ϵt qui représente la différence entre la réalité et le résultat de la décomposition

$$Xt = f(t, \epsilon t) \tag{2.3}$$

Il existe deux modèles de décomposition d'une série chronologique :

- **le modèle additif :** où l'observation de la série à la date t est une somme de la tendance (Z_t), la saisonnalité (S_t) et les fluctuations irrégulières (ϵ_t) :

$$\mathrm{Xt} = \mathrm{Zt} + \mathrm{St} + \epsilon_t \tag{2.4}$$

- **le modèle multiplicatif :** où la variable Xt est un produit de la tendance et de la saisonnalité

$$Xt = \mathrm{Zt}(1 + \mathrm{St})(1 + \epsilon_t) \tag{2.5}$$

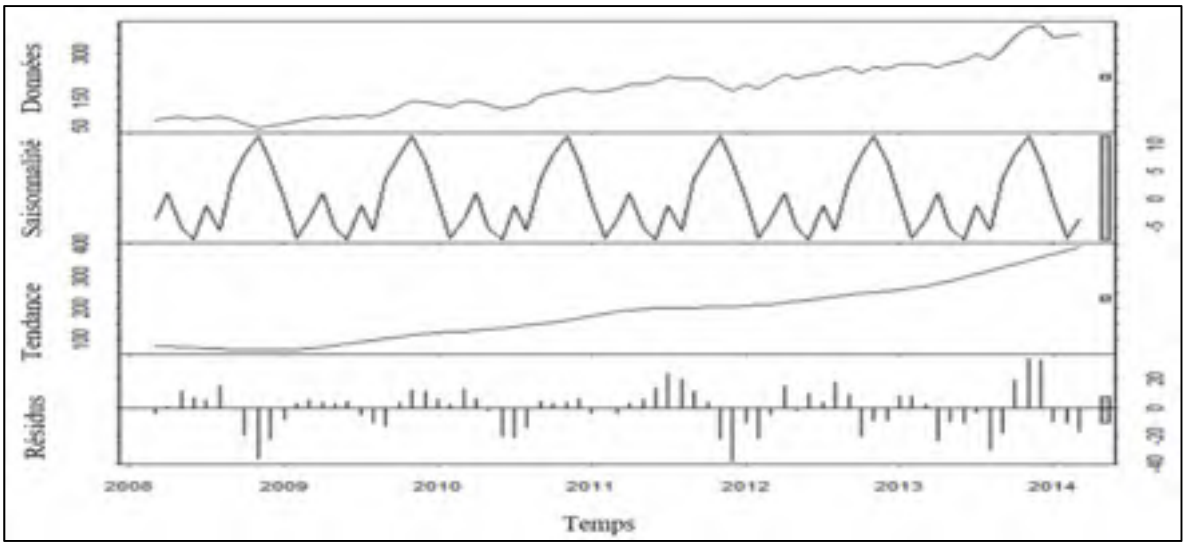


Figure 2.4 Décomposition de la série de prix de l'action Amazon entre 2008 et 2014

La figure 2.4, montre un exemple de la décomposition du cours de l'action Amazon. On y trouve les trois composantes discutées précédemment. Une tendance haussière claire et nette s'affiche.

La saisonnalité montre une baisse récurrente au cours des premiers mois de l'année et une hausse à la fin de l'année. Des fluctuations irrégulières baissières accentuées sont observées à la fin de 2008, la fin de 2011 et au milieu de 2013.

Il est à noter que les modèles de décomposition et de lissage (ex : moyennes mobiles) peuvent être utilisées dans la prédiction de la tendance. Même si nous avons exploré cette piste, L'objectif de notre projet est de tester les méthodes d'apprentissage automatique. C'est pourquoi, on va se limiter à l'analyse statistique du comportement d'une série chronologique financière et non pas à leur utilisation dans la prédiction.

2.4 Les données pour la prédiction des séries financières

Il y a une panoplie de données financières qui peuvent être utilisées pour la prédiction de la direction d'un actif financier. Tel que discuté dans la section 1.3.3 sur les méthodes de prévision, on peut utiliser soit les données financières fondamentales qui sont reliées à la situation du marché, du secteur d'activité ou de la compagnie, ou les données sur les transactions d'échange effectuées sur les marchés boursiers (données techniques : prix et volumes) pour essayer de prédire la tendance d'un titre boursier.

2.4.1 Données fondamentales

Comme nous l'avons discuté à la section (1.3.1.1) Il existe plusieurs sous catégories de l'analyse fondamentale. Les données fondamentales peuvent aussi être réparties en catégories :

- **données sur l'économie globale** : les indicateurs économiques globaux tel que le PIB, le taux d'inflation, la balance commerciale, le taux de chômage, les taux d'intérêt, la production industrielle.

- **données sur le secteur d'activité** : la distribution des parts du marché, le potentiel du secteur et son évolution.

- **données sur l'actif sous-jacent** : les ventes, les profits et les marges bénéficiaires. Il est à noter que pas toutes les données fondamentales sont disponibles au public.

2.4.2 Données transactionnelles (ou données techniques)

2.4.2.1 Les données brutes

Ce sont les données des transactions qui ont eu lieu sur les marchés boursiers. On en trouve les informations suivantes :

- le temps quand la transaction a eu lieu,
- les séries de prix (bas, haut, ouverture et fermeture),
- les volumes : les quantités qui ont été échangées au cours d'une période de temps.

Tableau 2.2 Extrait de données financières de l'indice Standard & Poors 500 Tirée de Yahoo finance (entre le 08 et le 21 Août 2018)

Date	Prix d'ouverture	Prix haut	Prix bas	Prix de ferméture	Volume
08/08/2018	2856.79	2862.44	2853.09	2857.70	2,972,200,000
09/08/2018	2857.19	2862.48	2851.98	2853.58	3,047,050,000
10/08/2018	2838.90	2842.20	2825.81	2833.28	3,256,040,000
13/08/2018	2835.46	2843.40	2819.88	2821.93	3,158,450,000
14/08/2018	2827.88	2843.11	2826.58	2839.96	2,976,970,000
15/08/2018	2827.95	2827.95	2802.49	2818.37	3,645,070,000
16/08/2018	2831.44	2850.49	2831.44	2840.69	3,219,880,000
17/08/2018	2838.32	2855.63	2833.73	2850.13	3,024,100,000
20/08/2018	2853.93	2859.76	2850.62	2857.05	2,748,020,000
21/08/2018	2861.51	2873.23	2861.32	2862.96	3,147,140,000

- **le prix d'ouverture (Open)**: c'est le prix auquel le premier échange du titre a eu lieu à l'ouverture d'une bourse (P_o) ou pour une période d'observation pour des données agrégées.

- **le prix de fermeture (Close)** : C'est le dernier prix (P_n) auquel un titre a été négocié durant une séance de négociation, ou pendant une période d'observation déterminée.

Pour la plupart des instruments financiers, il existe une séance de négociation après la fermeture des marchés. Les volumes et les niveaux de liquidité sont nettement moins élevés à cette séance. Par conséquent, le prix de fermeture à cette séance d'un titre pourra changer drastiquement au cours de cette séance. Surtout suite à la publication des résultats financiers ou à l'arrivée des nouvelles.

- **le prix haut (High)** : C'est le cours le plus élevé d'un titre au cours d'une séance de négociation ou une période d'observation.

$$P_h = Max(P_t), \qquad \text{pour } t \in [0, n] \tag{2.6}$$

- **le prix bas (Low)** : C'est le prix le moins élevé (Pl) auquel un titre a été échangé au cours d'une séance de négociation, ou pour une période d'observation pour des données agrégées.

$$P_l = Min(P_t), \qquad \text{pour } t \in [0, n] \tag{2.7}$$

Où :

Pt est la série de prix pour une séance de négociation ou une période d'observation,

n est la longueur de la série des prix ou de la période d'observation. Le nombre total d'observations que contient la série,

t est l'index temporelle où à chaque fois il y a eu un échange du titre.

- **le volume (V)** : il s'agit de la quantité totale des actions ou des contrats négociés pour un titre boursier. Il peut être mesuré pour tout type de titre négocié en bourse : actions, obligations, options, contrats à terme et tous les types de produits de base. Souvent le volume est un indicateur révélateur de la formation d'une tendance. Plusieurs stratégies utilisent le volume comme un indicateur ou une composante à pondérer avec le prix (ex : Prix moyen pondéré en volume VWAP).

En général, on assume que pour les marchés américains, il y a 252 jours de négociation par année, soit 63 jours par trimestre et 21 jours par mois.

2.4.2.2 Les données agrégées

Les séries chronologiques financières sont généralement non stationnaires, ce qui les rend peu souhaitables pour les utiliser sous leur forme brute lors de la prévision (Mushtaq, 2011).
Au lieu de cela, les informations sont dérivées de ces séries en utilisant un prétraitement, généralement sur le prix de fermeture. Par exemple, le rendement de l'actif (Hellström, 1998).
Il est aussi important d'identifier l'horizon d'investissement. En effet, il s'agit de la durée de la détention du titre. Il peut être complétement diffèrent d'un investisseur ou négociateur à un autre. Les investisseurs ont un horizon d'investissement de long terme, allant de quelques mois à plusieurs années. Les négociateurs actifs essaient de limiter leurs expositions. Ils achètent et vendent leur titre plusieurs fois le même jour, voir même en quelques secondes.

L'horizon d'investissement détermine le niveau d'agrégation qui doit se faire pour les données brutes. Un négociateur qui désire acheter et vendre des titres le même jour voudra voir les données agrégées en minutes (exemple : 5 minutes). Un investisseur passif qui voudra se positionner sur des titres à long terme aura plus à regarder les données techniques agrégées en jour ou par semaine. Donc, l'agrégation consiste à utiliser des données des séries chronologiques avec des fréquences d'observation déterminées pour calculer les mêmes variables avec des fréquences de temps plus élevées. Cela peut être le cas lors de la prévision hebdomadaire ou mensuelle. Une approche permettant d'agréger des données quotidiennes dans une série temporelle hebdomadaire consiste à utiliser la valeur de clôture par séance ou par période d'observation. Les séries de prix haut et bas peuvent ensuite être récupérées en recherchant les valeurs les plus élevées et les plus faibles.

Pour ce projet on s'intéresse à la négociation intra-journalière. Alors, nous avons élaboré notre méthode d'agrégation où nous utilisons les données brutes (données à la minute) pour la généralisation et l'application de l'agrégation à n'importe quelle autre période.

La période d'observation est à déterminer par l'utilisateur en fonction du besoin d'analyse et de prévision. Donc, Notre fonction d'agrégation contient un seul paramètre : la période d'agrégation (*K*).

Soit (Xt) La série chronologique contenant les variables suivantes :

- le prix d'ouverture;
- le prix de fermeture;
- le prix bas et le prix haut;
- le volume.

Ces données sont collectées à une période déterminée (*J*).

X^K_t est la série chronologique résultant de l'agrégation à K périodes des variables d'entrées, *K* étant la période d'agrégation (en minutes). K=5 veut dire que les sorties de l'agrégation sont les données d'entrée, agrégées à 5 minutes. À noter que K >=J,
n est la longueur de la série.

Nous avons la fonction suivante :

$$X^K t = f(X_t, K) \tag{2.8}$$

Il s'agit en fait d'une transformation à appliquer pour chaque variable des données brutes.

- **prix d'ouverture agrégé (P^K_0)** : Est le premier prix de la série des prix d'ouvertures d'entrée pendant K périodes.

$$P^K_{0(i)} = f(P_0, K) = P_{0((i-1) * k + 1)} \tag{2.9}$$

  Si nous voulons passer des données à chaque minute à des données transformées à 5 minutes. Le prix d'ouverture à 5 minutes sera le premier de la série du prix d'ouverture à la minute pendant 5 minutes.

- **prix de fermeture agrégé (P^K)** : Est le dernier prix de la série du prix de fermeture d'entrée pendant K périodes.

$$P^K_{(i)} = f(P, K) = p_{(i*k)} \quad (2.10)$$

Le prix de fermeture à 5 minutes est le dernier prix de fermeture à la minute, pendant 5 minutes d'observation.

- **prix haut agrégé (P^K_h)** : on l'obtient en cherchant le prix le plus élevé pendant K périodes entre $((i-1)*k)+1$ et $i*k$.

$$P^K_h(i) = Max(Pt((i-1)*k+1 : i*k)), \text{ pour } i \in [0, n] \quad (2.11)$$

- **prix bas agrégé (P^K_l)** : on l'obtient de la même façon que le prix haut agrégé mais en utilisant le minimum à la place du maximum.

$$P^K_l(i) = Min(Pt((i-1)*k+1 : i*k)) \text{ , pour } i \in [0, n] \quad (2.12)$$

- **volume agrégé ($V^K t$)** : c'est la somme des volumes de la série d'entrée pendant K périodes.

$$V^K t(i) = \sum_{(i-1)*k+1}^{i*k} Vt\,(i) \quad (2.13)$$

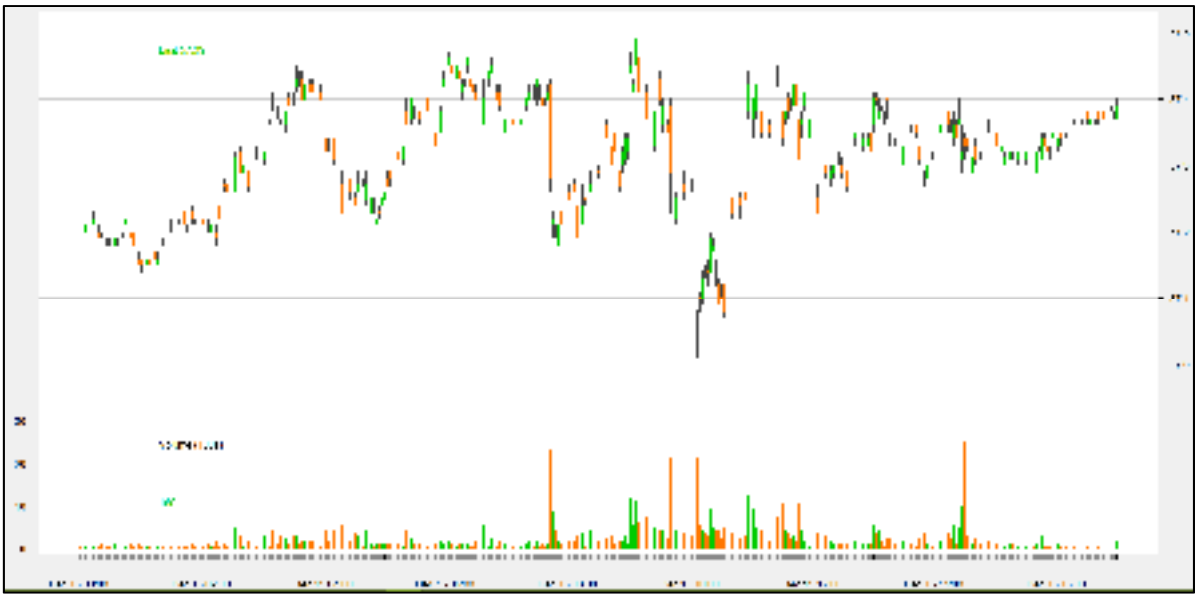


Figure 2.5 Données du contrat à terme du pétrole CLG8 à la minute

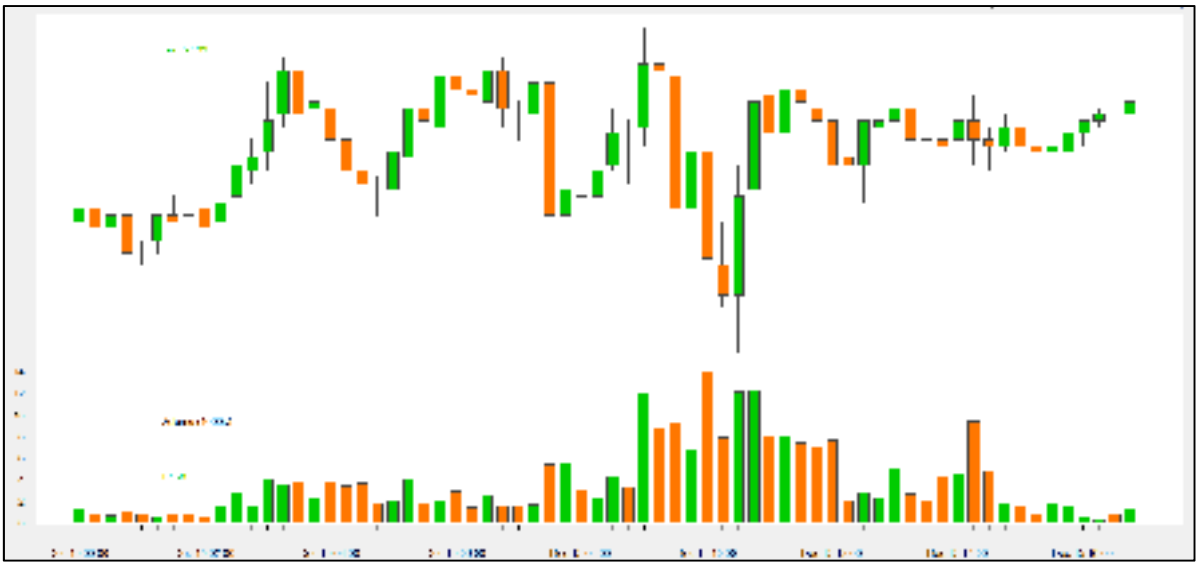


Figure 2.6 Données du contrat à terme du pétrole CLG8 à 15 minutes

Les deux graphiques en haut affichent les mêmes données. La seule différence est la plage temporelle que nous avons utilisée dans chacun d'eux. Pour le premier graphique, les données sont présentées par intervalle d'une minute. Pour le second, les données sont agrégées par 15 minutes.

L’agrégation des données permet de réduire le bruit et lisser les séries des données brutes pour les adapter à nos préférences d’analyse et les contraintes de notre stratégie d’investissement.

2.4.3 Données dérivées

À partir des séries chronologiques brutes (par exemple, les cours des actions, le volume, etc.) on peut dériver plusieurs variables qui peuvent être utilisée pour mieux représenter l'information et souligner les informations importantes contenues dans les données brutes.
Les indicateurs techniques sont des données dérivées, car ils sont calculés à partir des données brutes des séries des prix et des volumes. On abordera ça plus dans les prochains paragraphes de cette section.

2.4.3.1 Rendement du cours

Il s’agit d’un rendement qui inclut seulement de la valorisation du capital de l’actif et ne tient pas compte des revenus générés (comme les intérêts et les dividendes). Dans le cas des spéculateurs et négociateurs à court terme, on considère seulement ce gain en capital. Dans ce cas, la série des rendements est dérivée directement de la série des prix qui souffre de certaines faiblesses comme la non-stationnarité des prix. Le prix dépend toujours du prix de la période précédente. Par contre, le changement en pourcentage (ou la différence de log) enlève cet effet et permet de comparer les rendements de chaque période entre eux, ainsi que comparer le rendement des séries de prix différentes (Grothmann, 2003), ce qui donne une bonne description du comportement de l’actif entre différents points d’observation.

De plus, les rendements des actifs sont généralement traités comme des variables aléatoires indépendantes continues (Grothmann, 2003). Le rendement peut être calculé pour la plupart des actifs financiers (actions, obligations, taux de change, contrat à terme, etc.) Il suffit de connaitre le prix d’ouverture et de fermeture de la position.

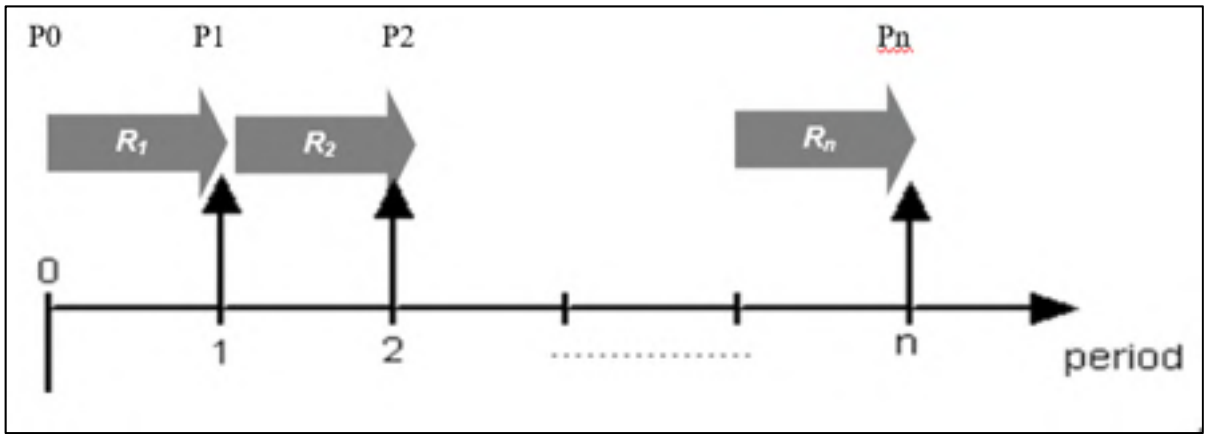
Il existe aussi différentes façons pour calculer le rendement du cours. Le rendement simple (courant dans la communauté des négociateurs) et le rendement composé continu (couramment utilisé par les universitaires) sont traités dans cette section.

- **le rendement simple :**

  Il a déjà été traité dans la section 2.1.2.1 et il s'agit tout simplement de la variation du prix entre l'instant (t) et l'instant (t+k) vu dans l'équation (2.1)

$$R_{k;t} = \frac{P_{t+k} - P_t}{P_t}$$

  Dans le cas où nous avons des positions sur plusieurs périodes, le rendement global qui sera obtenu sur différentes positions est composé de tous les rendements individuels. Ceci est présenté dans le diagramme en bas.



Dans ce cas, nous avons :

$$\hat{R}_{k;t} = \prod_{i=0}^{k-1}(1 + R_{t-i}) \qquad (2.14)$$

- **le rendement composé continue (δ)** :

  Utilisé souvent pour réduire l'impact des données aberrantes dans les séries de rendement. Pour ce rendement, on suppose la continuité de la série des prix.

$$\delta t = \log\left(\frac{P_{t+k}}{P_t}\right) \qquad (2.15)$$

  Il y a plusieurs avantages de l'utilisation de cette forme. Entre autres, la simplicité à les utiliser dans les algorithmes.

Il y a aussi le fait que la détendancialisation et la normalisation de la série chronologique, tel que vu dans la section 2.1.3, est plus facile à faire avec ce rendement à cause de l'utilisation du logarithme.

2.4.3.2 La tendance

Comme nous l'avons vu dans la section 2.1.2.1, la tendance se calcule directement à partir des prix historiques (équation 2.1 et 2.2). Souvent, il est important d'utiliser un seuil de profitabilité pour définir la tendance, au lieu de considérer juste la variation de prix entre un moment t et t + k. Ainsi, la tendance sera positive si la variation du titre entre deux moments est supérieure à un seuil prédéterminé et négatif si la variation est négative et inférieure à ce seuil.

Nous pouvons mathématiquement la définir sous l'équation suivante :

$$Tt[k] = \begin{cases} -1, si: & Rt[k] \leq -r \\ 0, si: & -r < Rt[k] < r \\ 1, si: & Rt[k] \geq r \end{cases} \quad (2.16)$$

Où :

$Rt[k]$ est la variation du titre entre k et t + k. Elle pourra être absolue ou relative, dépendamment de notre choix,

r est le seuil de rendement prédéterminé.

Si nous optons pour une variation absolue, le seuil défini devrait être aussi un nombre absolu représentant le gain minimal nécessaire pour entrer dans une position. Cependant, si nous voulons utiliser une variation relative, le seuil doit être le gain espéré en pourcentage.

En effet, notre objectif est la prévision d'un mouvement qui génèrera assez de rentabilité. Les petites variations ne sont pas importantes. Exemple : nous cherchons à entrer dans des positions sur des contrats à terme de 15 minutes seulement si la variation de prix entre le point t et (t + 15 minutes) est de plus de 5 points ou 0.25% (par exemple, le prix du contrat passe de 2000 $ à 2005 $)

Figure 2.7 Comparaison de tendance d'un contrat à terme ESM6 - sans et avec un seuil de rentabilité

Sur le graphe en haut nous avons tracé la variation du titre entre n'importe quel point et les 15 minutes suivantes. La première courbe de tendance en bas représente une simple variation entre un temps t et t + k. Tandis, que la deuxième courbe représente la tendance telle que défini dans (2.14).

On peut facilement voir que la deuxième courbe est plus lisse et contient moins de variations. Ceci est dû au fait que la tendance est non nulle seulement si la variation dépasse le seuil de rendement prédéterminé à la hausse ou à la baisse. Il est à noter aussi que plus qu'on augmente le seuil de rendement espéré, moins de signaux seront générés.

2.4.3.3 La volatilité

La volatilité est une mesure statistique de la variabilité d'une série de rendements pour un titre ou un indice de marché donné. Elle donne une bonne idée sur le risque qui est associé à un actif financier. De façon générale, plus la volatilité est élevée, plus l'actif est risqué. Elle peut être mesurée en utilisant l'écart type ou la variance des rendements logarithmiques, tel que décrit dans l'équation suivante :

$$\sigma_t = \sqrt{\frac{\sum_1^t (\delta t - \mu)}{(t-1)}} \qquad (2.17)$$

Où :

δt est le rendement logarithmique du titre (équation 2.2) $\delta t = \log\left(\frac{P_t}{P_{t-1}}\right)$;

μ est le rendement logarithmique moyen sur la période d'observation.

La volatilité du marché se mesure avec l'indice de volatilité, le VIX (Volatility Index). Elle est calculée en temps réel à partir des prix des options d'achat et de vente de l'indice boursier Standard & Poors 500, pour un horizon de 30 jours.

2.4.3.4 Les indicateurs techniques

Tel qu'introduit dans la section (1.3.3.2), l'analyse technique utilise les séries de prix, de volumes et de rendements dans le but de prédire l'évolution future des cours des actifs financiers. Pour ce faire, les techniciens (spécialistes de l'analyse technique) utilisent les indicateurs techniques. Un indicateur technique est une dérivée du prix ou du volume ou des deux, en utilisant une formule mathématique décrivant cette fonction. Les indicateurs techniques sont visualisés graphiquement dans le but de montrer l'évolution de la force et la vitesse d'un mouvement sur une période donnée. Ils sont à la base conçus pour analyser les mouvements de prix à court terme, mais les investisseurs à long terme peuvent également les utiliser pour identifier les points d'entrée et de sortie. On distingue deux types d'indicateurs techniques de base : les indicateurs de tendance et les oscillateurs.

1) Les indicateurs de tendance

Ces indicateurs utilisent la même échelle que les prix qui sont tracés en haut sur un graphique boursier. Ils permettent d'identifier la tendance en cours et déterminer des niveaux de supports et de résistances. Les moyennes mobiles et les bandes de Bollingers sont les indicateurs de tendance les plus populaires.

a) Les bandes de Bollinger

Il s'agit d'un exemple d'indicateur technique quantitatif qui pourra être visualisé graphiquement pour détecter visuellement des signaux d'entrée et de sortie mais aussi permet de programmer facilement ces règles dans un ordinateur.

Cet indicateur est composé de trois courbes. La courbe au milieu qui est une moyenne mobile des prix sur n périodes (en général 20). La bande supérieure est de "d" écart types (généralement d = 2) au-dessus de la courbe au milieu et la bande inférieure est de "d" écarts types au-dessous.

Mathématiquement, Soit P une série de prix de fermeture et σ_n est l'écart type de cette série.

Nous avons les équations suivantes :

- la bande au milieu (MM(n)): $$MM_{(n)} = \frac{1}{n} \sum_{t=0}^{n} P_t \quad (2.18)$$
- la bande supérieure B_{upper} (d, n) : $$B_{upper}(d, n) = MM(n) + d \cdot \sigma_n \quad (2.19)$$
- la bande inferieure B_{lower} (d, n) : $$B_{lower}(d, n) = MM(n) - d \cdot \sigma_n \quad (2.20)$$

L'écart entre les bandes indique le niveau de volatilité qui existe.

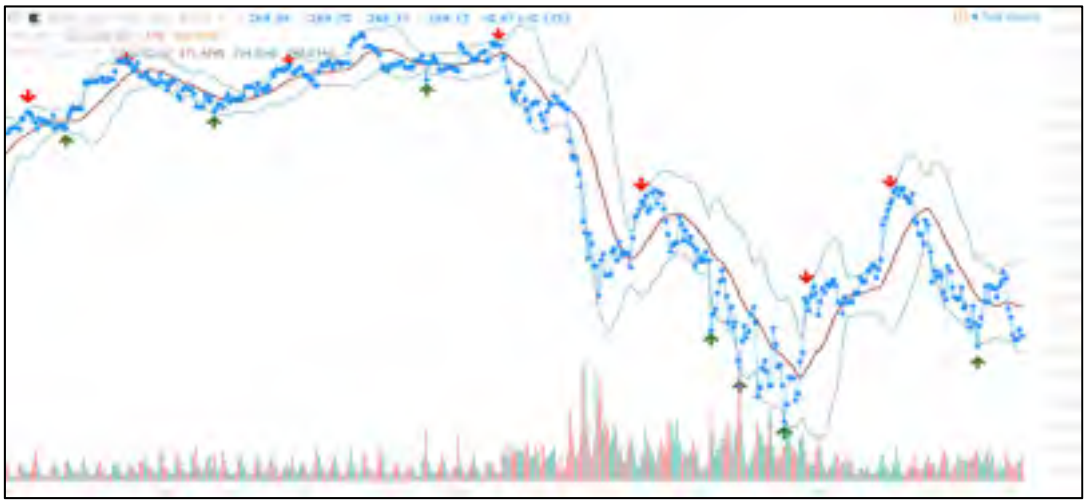


Figure 2.8 Exemple d'utilisation des bandes de Bollinger sur le titre SPY

Nous avons tracé sur le graphique en haut des signaux d'entrée et de sortie en utilisant les bandes de Bollingers. En fait, quand le prix s'approche des bandes supérieures ou inférieures, cela indique qu'une inversion est imminente.

Lorsque le prix sort de la bande supérieure et la touche continuellement on s'attend à voir une baisse, le titre est considéré surévalué. À l'inverse, lorsqu'il touche continuellement la bande inférieure, le prix est considéré comme sous-évalué, ce qui déclenche un signal d'achat.

b) Les moyennes mobiles

C'est l'un des indicateurs techniques les plus populaires et les plus simples. Une moyenne mobile n'est autre qu'une moyenne calculée sur une série de prix et qui lisse les données en créant un prix moyen qui est mis à jour périodiquement dépendamment de la fréquence utilisée. Ceci permet de réduire le bruit des variations du cours à court terme.

Les moyennes mobiles peuvent être utilisées comme niveaux de support et de résistance. Les titres dans une tendance haussière trouvent souvent un support aux moyennes mobiles longues telles que les moyennes mobiles simples couramment utilisées sur 50 et 200 jours. L'idée consiste à acheter des titres dans une tendance haussière à l'approche de leurs moyennes mobiles et vendre dans une tendance baissière à mesure qu'ils s'approchent de leur moyenne mobile inférieure. Aussi, elles sont souvent utilisées pour trouver des signaux d'entrée et de sorties sur des positions. Le signal est basé sur le croisement de moyennes mobiles de périodes différentes (courte et longues). Exemple : Lorsque la moyenne mobile simple à 50 jours (MM50) croise à la hausse, la moyenne mobile à 200 jours (MM200), il s'agit d'un signal d'achat. Et lorsque la MM50 croise à la baisse la MM200 c'est un signal de vente (Voir figure 2.9).

Figure 2.9 Utilisation de moyennes mobiles exponentielles pour trouver des signaux – Paire de devise US/CAD

Il existe plusieurs types de moyennes mobiles. Les plus fréquemment utilisées sont les moyennes mobiles simples et les moyennes mobiles exponentielles.

2) Les oscillateurs

Ce sont des indicateurs qui oscillent entre un minimum et un maximum local et permettent de reconnaitre certains déséquilibres entre les acheteurs et les vendeurs pouvant mener à une correction ou un renversement à court terme. Ils sont en général, tracés au-dessous d'un graphique de prix et des seuils sont choisis pour identifier les zones de surachat et de survente. Les exemples incluent le MACD et le RSI.

a) L'indicateur de convergence divergence des moyennes mobiles (MACD)

Sur le graphe boursier de l'action de la compagnie Microsoft dans la figure 2.1, nous avons tracé des signaux d'entrée et de sortie en utilisant l'indicateur MACD (Moving Average Convergence Divergence). Il s'agit simplement de la différence entre deux moyennes mobiles exponentielles de périodes différentes (MME).

De façon générale, on utilise des périodes de 12 et 26 jours. On emploie aussi une courbe de tendance (moyenne mobile exponentielle de la MACD) pour obtenir le signal de la MACD à la suite d'un croisement des deux.

Figure 2.10 Exemple d'utilisation du MACD sur l'action MICROSOFT

Un signal d'achat est obtenu quand la MACD traverse à la hausse la ligne de signal. Et à l'inverse, quand on a croisement à la baisse, il s'agit probablement d'un signal de vente.

b) L'indice de la force relative (RSI)

Cet oscillateur mesure la vitesse à laquelle les mouvements de prix changent et l'ampleur de ces fluctuations sur une période déterminée. Il peut être pertinent lorsqu'il est mesuré sur une période de 14 jours. De façon générale, quand l'indice dépasse un seuil de 70 points, ceci est un signal que l'actif est suracheté et c'est probablement le moment idéal pour le vendre. Si cette valeur est en bas de 30, ceci est un signal alarmant que le titre est survendu.

En effet, le RSI peut aider à savoir si une tendance se précisera ou subira un renversement. Il s'adapte en fonction des nouveaux cours intégrés dans son calcul.

Figure 2.11 Exemple d'utilisation du RSI sur l'action MICROSOFT

Conclusion sur les indicateurs techniques

Étant donné leur nature quantitative, les indicateurs techniques peuvent facilement être intégrés aux systèmes de négociation automatisés et ils sont de bons candidats pour l'utilisation comme attributs dans les algorithmes d'apprentissage automatique. Toutefois, il est nécessaire de souligner que ce sont les séries de prix et de volumes qui forment ces indicateurs et non pas le contraire. Il est souvent facile de voir de faux signaux qui sont générés par ces derniers. C'est pourquoi, de nombreux indicateurs doivent être combinés ensemble pour trouver et confirmer des signaux d'entrée et de sortie. L'une des utilités de l'apprentissage automatique est la capacité à trouver des combinaisons gagnantes de ces indicateurs ou des règles permettant de générer des signaux d'entrée et de sortie fiables.

Il existe un grand nombre d'indicateurs techniques que nous pouvons utiliser dans notre système de négociation automatisé. Dans la partie expérimentale de ce projet, nous expliquerons comment on peut utiliser ces indicateurs comme variables explicatives dans nos modèles d'apprentissage automatique et comment on fait la sélection de ces derniers pour garder seulement ceux qui sont les plus importants.

2.5 Les défis spécifiques aux séries chronologiques financières

Plusieurs défis sont rencontrés et doivent être traités quand on veut modéliser des séries temporelles financières. Entre autres, la non-stationnarité de ces séries, la distribution non équilibrée des rendements, le bruit dans les données collectées, l'existence de valeurs aberrantes et des valeurs manquantes. Dans cette section nous allons aborder ces points et essayer de donner des stratégies pouvant aider à contourner ces problèmes.

2.5.1 La distribution non équilibrée des rendements

Il est plus fréquent d'avoir une distribution non équilibrée des classes de la tendance d'une série chronologique, avec plus d'instances d'une classe particulière que d'autres (voir la figure 2.9). Faire l'apprentissage de nos algorithmes à partir d'un tel ensemble de données aboutit généralement à un model prédictif biaisé, qui aura plus tendance à favoriser la classe majoritaire. Ainsi, le model aura tendance à mal classer les instances des classes minoritaires. De façon général, les algorithmes de classification cherchent à maximiser la performance de la classification. C'est pourquoi, les instances de la classe minoritaire sont plus fréquemment mal classées.

Tableau 2.3 Exemple de distribution biaisée de la tendance d'une série chronologique financière

Tendance	Distribution
Hausse	20%
Neutre	70%
Basse	10%

Lors de la prévision de la direction des cours boursiers on s'intéresse aux mouvements qui génèrent de la profitabilité (le plus souvent, la ou les classe(s) minoritaire(s)). Les petites variations sont relativement sans importance et ne sont pas rentables.

Il y a plusieurs techniques qui peuvent être utilisées pour faire face à ce problème. Nous les discuterons dans le chapitre sur la méthodologie de notre expérimentation.

2.5.2 Les valeurs aberrantes

Les séries financières contiennent souvent des valeurs aberrantes, ce qui constitue un problème à résoudre. Elles peuvent avoir un effet négatif sur les performances des modèles de prévision. Une valeur aberrante est une valeur qui s'écarte fortement des valeurs des autres observations de la série chronologique et peut être le résultat d'une forte réaction à la suite d'une nouvelle inattendue ou d'un événement extrême (par exemple, les attentats du 11 septembre 2001).

Les valeurs aberrantes peuvent avoir un effet négatif sur les modèles de prévision. Il convient de prendre des mesures pour réduire leur impact sur les performances de prévision(TSAY, 2005). Ceci peut être fait en :

- modélisant le comportement normal du système,
- associant au couple (modèle-observation) des caractères permettant d'évaluer un écart par rapport au comportement normal,
- cherchant à décider si l'écart mesuré est significatif ou non.

Une fois identifiée, une valeur aberrante devra être traitée. Il y a plusieurs méthodes pour le remplacement et l'imputation des valeurs aberrantes.

2.5.3 Les valeurs manquantes

L'existence de valeurs manquantes dans les données est un problème très courant dans les applications du monde réel et ne doit jamais être négligé. Car ceci pourra impacter négativement la performance d'un modèle prédictif et créer un biais dans les données utilisées (Dunsmuir & Robinson, 1981). Les fins de semaines et les jours fériés sont une source de valeurs manquantes, car ce ne sont généralement pas des jours de négociation. Une autre source peut être des erreurs techniques ou humaines conduisant à des valeurs non enregistrées.

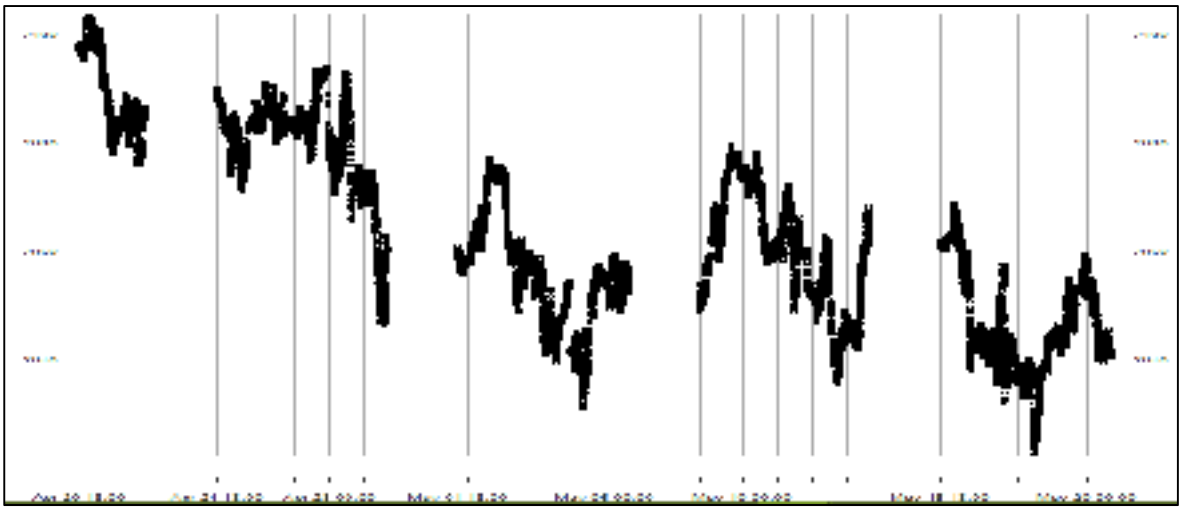


Figure 2.12 Exemple de données manquantes pour le contrat S&P 500 E-Mini

Sur le graphe en haut, on voit des données manquantes qui forment des creux durant les fins de semaines ou les jours fériés. Lors du prétraitement des données d'entrée dans un modèle de prévision, les valeurs manquantes doivent être traitées.

Trois approches différentes à ce problème sont énumérées ci-dessous :

- **remplacer les valeurs manquantes :**
  un moyen simple de gérer les valeurs manquantes dans une série temporelle consiste à les remplacer par la dernière valeur précédente connue. Un autre choix consiste à remplacer les valeurs manquantes par la moyenne de la série temporelle à laquelle elle appartient. Deux problèmes sont à considérer si on décide d'utiliser la moyenne pour remplacer les valeurs manquantes. Le premier concerne l'utilisation des données futures lors du calcul de la moyenne, ce qui est évidemment faux. L'autre problème est lié à la non-stationnarité des séries chronologiques financières tel que décrit précédemment. Ces données ont une moyenne qui change dans le temps et peut fluctuer de façons différentes pour différentes parties de la série. Il faut donc éviter d'utiliser la moyenne d'une série temporelle pour remplacer les valeurs manquantes lorsque la série est non stationnaire.

- **imputer les données manquantes en utilisant des modèles prédictifs** :
  On peut utiliser des modèles d'apprentissage automatique pour prédire la valeur des données manquantes et remplacer leur valeur par la valeur prédite. L'un des algorithmes les plus fréquents à cette fin est la méthode des K plus proches voisins qui sera abordé dans le chapitre suivant (section 3.3.1)

- **supprimer les instances qui contiennent des données manquantes** :
  L'une des façons les plus simples pour traiter les valeurs manquantes sans avoir à imputer les données manquantes, consiste à les supprimer. En général, en supprimant les lignes comportant des valeurs manquantes pour les variables utilisées à condition que les valeurs manquantes ne soient pas trop nombreuses ou n'ayant pas d'effet sur le comportement de la série globale.

CHAPITRE 3

INTRODUCTION À L'APPRENTISSAGE AUTOMATIQUE

3.1 Aperçu sur la négociation algorithmique

Dans les sections précédentes (1.3.3 et 2.2.3.4) nous avons vu des techniques de prévision pouvant être utilisées pour la prédiction des tendances des cours des actifs financiers. En effet, il est possible d'automatiser la négociation des actifs financiers en créant un système qui est basé sur des règles définies à priori. Quand ces règles sont satisfaites, le système envoi un ordre d'achat ou de vente dépendamment du signal qui est généré.

- **en utilisant l'analyse fondamentale** : si la valeur théorique du titre calculée, en utilisant les ratios fondamentaux ou les modèles d'évaluation fondamentaux, est supérieure à sa valeur marchande, alors le titre est sous-évalué et pourra représenter un bon potentiel de croissance dans le futur. À noter que ce genre de règles pourra être utilisées pour le long terme ou pour la sélection des titres sur une base fondamentale.

- **en utilisant l'analyse technique :** les indicateurs techniques tel que vu dans le chapitre précédant, pourront être utilisés pour générer des signaux d'entrée et de sortie sur le marché. Un bon exemple est celui des bandes de Bollinger (figure 2.6).

Le grand défi d'un tel système est la capacité à trouver des règles gagnantes pour n'importe quel titre ou marché et la capacité d'adaptation aux changements dans les propriétés des séries chronologiques financières (ex. : volatilité, non stationnarité, dérive conceptuelle). L'autre point est qu'un système qui est basé sur des règles prédéfinies est limité par le nombre de variables et de règles pour lesquelles il a été conçu. Il est alors extrêmement difficile, sans avoir recours à des techniques quantitatives avancées, d'avoir un bon nombre de combinaisons

de variables et de règles qui peuvent être testés pour démontrer leur performance et leur efficacité.
D'où vient la nécessité d'avoir un système dynamique évolutif qui est capable d'apprendre de façon continue, au fur que de nouvelles données sont intégrées, et d'optimiser la stratégie d'investissement en conséquence.

Dans ce chapitre on va aborder les techniques d'apprentissage automatique et comment des algorithmes peuvent être utilisés pour apprendre à partir des données collectées. On va ensuite énoncer les algorithmes les plus populaires, tel que les arbres de décision, les machines à support vectorielles, les réseaux de neurones et les méthodes d'ensemble.

3.2 Les types d'apprentissage automatique

L'apprentissage automatique ou Machine Learning (ML) est une branche de l'intelligence artificielle qui consiste à programmer des algorithmes permettant d'apprendre automatiquement à partir des données et d'expériences passées ou par interaction avec l'environnement. Ce qui rend l'apprentissage machine vraiment utile est le fait que l'algorithme peut "apprendre" et adapter ses résultats en fonction de nouvelles données sans aucune programmation à priori.

Il existe plusieurs façons d'apprendre automatiquement à partir des données dépendamment des problèmes à résoudre et des données disponibles. La figure 3.1 donne un sommaire des types d'apprentissage automatique les plus connus.

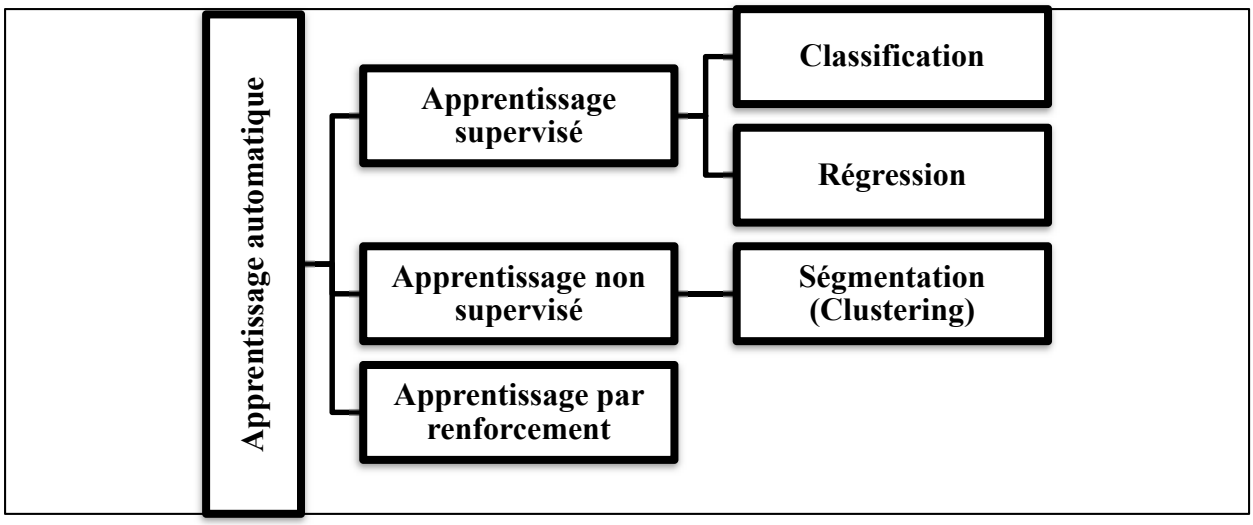


Figure 3.1 Les grandes classes d'apprentissage automatique

3.2.1 L'apprentissage supervisé

L'algorithme est entrainé en utilisant une base de données d'apprentissage contenant des exemples de cas réels traités et validés. L'objectif est de trouver des corrélations entre les données d'entrée (variables explicatives) et les données de sorties (variables à prédire), pour ensuite inférer la connaissance extraite sur des entrées avec des sorties inconnues.

Chaque exemple, appelé aussi instance, est un couple d'entrée-sortie (x_n, y_n) $1 <= n <= N$, avec $x_n \in X$ et $y_n \in Y$ et Où :
X est l'ensemble d'attributs (discrets ou continues),
Y est l'ensemble des valeurs de sortie (la variable cible ou dépendante),
Y peut être discrète ou continue (Lo & MacKinlay, 2002).

En apprentissage supervisé, on distingue entre deux types de tâches :

- **la classification** : quand la variable cible (à prédire) est discrète, Y = {1, ...I}. Ce qui revient à attribuer une classe (ou étiquette) à chaque entrée. C'est le cas si on cherche à prédire la tendance d'un mouvement futur d'un actif (haut, neutre, bas).

Tableau 3.1 Exemple d'un ensemble de données d'apprentissage supervisé (Classification)

Données d'entrée								Cible
Date	**VWAP**	**SMA 20**	**Croisement de MM (20,50)**	**MACD**	**RSI**	**BB %**	**Sentiment**	**Tendance**
01-08-2018	51.54	51.95	105%	-30%	-40%	45%	Positif	Baisse
02-08-2018	52.37	53.31	102%	-30%	-39%	45%	Positif	Baisse
03-08-2018	51.57	53.03	91%	-30%	-40%	46%	Négatif	Baisse
04-08-2018	50.9	52.86	86%	-31%	-41%	47%	Négatif	Hausse
05-08-2018	50.33	52.79	88%	-31%	-41%	48%	Négatif	Hausse
06-08-2018	50.03	53	87%	-32%	-42%	49%	Positif	Hausse
07-08-2018	50.66	53.67	79%	-31%	-41%	49%	Positif	Neutre
08-08-2018	49.75	53.13	80%	-32%	-39%	49%	Positif	Baisse
09-08-2018	48.86	52.07	73%	-33%	-39%	51%	Neutre	Neutre
10-08-2018	48.5	51.99	73%	-32%	-38%	53%	Neutre	Neutre

- **la régression** : quand la variable cible à prédire est continue $Y \subset \mathbb{R}$. Exemple : prédire le prix futur en dollars de l'actif en question.

Tableau 3.2 Exemple d'un ensemble de données d'apprentissage supervisé (Régression)

Données d'entrée								Cible
Date	**VWAP**	**SMA 20**	**Croisement de MM (20,50)**	**MACD**	**RSI**	**BB %**	**Sentiment**	**Prix (en $ CAD)**
01-08-2018	51.54	51.95	105%	-30%	-40%	45%	Positif	52.3
02-08-2018	52.37	53.31	102%	-30%	-39%	45%	Positif	54.1
03-08-2018	51.57	53.03	91%	-30%	-40%	46%	Négatif	51.5
04-08-2018	50.9	52.86	86%	-31%	-41%	47%	Négatif	51.9
05-08-2018	50.33	52.79	88%	-31%	-41%	48%	Négatif	51.05
06-08-2018	50.03	53	87%	-32%	-42%	49%	Positif	52
07-08-2018	50.66	53.67	79%	-31%	-41%	49%	Positif	50.45
08-08-2018	49.75	53.13	80%	-32%	-39%	49%	Positif	50.25
09-08-2018	48.86	52.07	73%	-33%	-39%	51%	Neutre	49.61
10-08-2018	48.5	51.99	73%	-32%	-38%	53%	Neutre	49.03

Pour les deux exemples dans le tableau 3.1 et 3.2 on utilise les mêmes attributs. La seule différence et que pour le premier exemple on cherche à prédire une classe (la tendance de l'action pour le jour suivant) tandis que pour le deuxième, on cherche à prédire le prix de l'action.

Nous avons utilisé 6 indicateurs techniques (variables numériques) et un indicateur de sentiment (variable catégorique). La date est utilisée comme un index pour notre tableau de modélisation mais il ne sera pas inclus comme une variable explicative dans notre modèle.

3.2.2 L'apprentissage non supervisé

Pour ce type d'apprentissage la base de données d'apprentissage ne contient pas de variable cible (comme on l'a vu en apprentissage supervisé). Il y a seulement un ensemble de données collectées en entrée. L'algorithme doit découvrir par lui-même la structure en fonction des données. On utilise cette technique pour partitionner les données en groupes d'éléments homogènes. La distance est souvent la plus utilisée comme mesure de similarité entre les groupes.

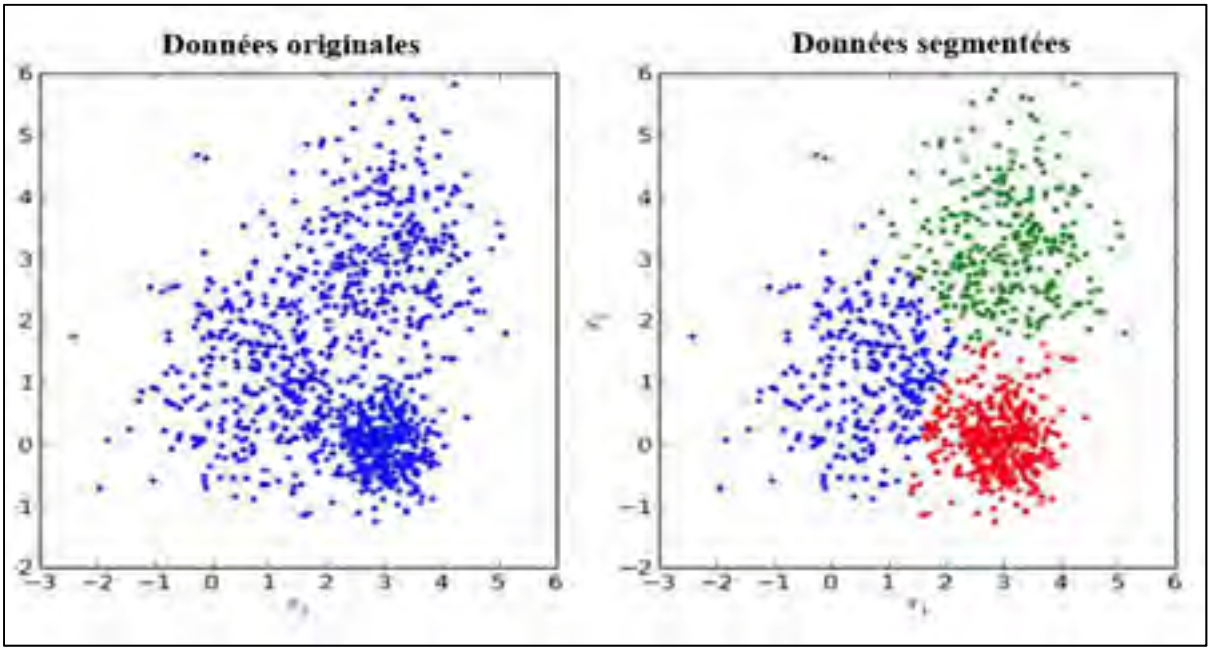


Figure 3.2 Exemple d'apprentissage non supervisé

3.2.3 L'apprentissage semi supervisé

Il s'agit d'un mixe entre l'apprentissage supervisé et non supervisé en utilisant des données étiquetées et non-étiquetées pour le même ensemble de données.

L’avantage d’utiliser cette approche réside dans le fait que l'étiquetage de données peut être couteux et prend souvent beaucoup de temps. En plus, il pourra entrainer un biais humain dans les données étiquetées. Dans ce cas, l'apprentissage semi-supervisé, qui ne nécessite que quelques étiquettes, est très pratique. Et le fait d’inclure un grand nombre de données non étiquetées au cours du processus d’entrainement a tendance à améliorer la performance du modèle final tout en réduisant le temps et les coûts consacrés à sa construction.

3.2.4 L’apprentissage par renforcement

L’apprentissage se fait sans supervision, par interaction avec l’environnement (principe d’essai / erreur) et, en observant le résultat des actions prises. Chaque action de la séquence est associée à une récompense. Le but est de déterminer la stratégie comportementale optimale afin de maximiser la récompense totale. Pour cela, un simple retour des résultats est nécessaire pour apprendre comment la machine doit agir. Ceci est appelé le signal de renforcement. Il peut être très avantageux pour la prévision financière à haute fréquence où l’environnement est dynamique et en conséquence, il est difficile de trouver ou d’automatiser manuellement des stratégies efficaces.

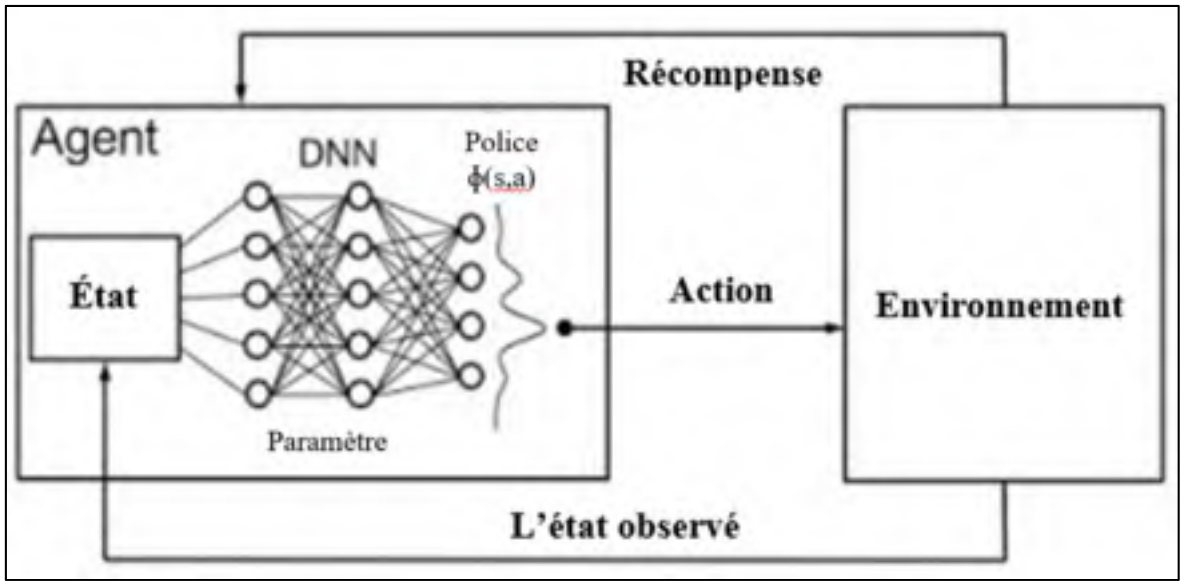


Figure 3.3 schéma descriptive de l’apprentissage par renforcement

Pour ce projet on va se limiter aux algorithmes d'apprentissage supervisé que nous allons détailler dans les sections suivantes.

3.3 Les algorithmes d'apprentissage supervisé

3.3.1 La méthode des K plus proches voisins

En abrégé KNN, de l'anglais 'k-nearest neighbors', est un algorithme de reconnaissance des formes qui peut être utilisé autant pour la classification que pour la régression. C'est l'une des techniques non paramétriques fréquemment utilisée en prédiction financière non linéaire. Cette préférence est dû principalement à deux raisons :

- premièrement, la simplicité algorithmique de la méthode comparée aux autres méthodes globales telles que les réseaux de neurones ou les algorithmes génétiques.
- deuxièmement, la méthode KNN a démontré empiriquement une importante capacité de prédiction.

L'idée de la méthode est de prédire le futur d'une série temporelle en analysant comment elle a évolué dans une situation similaire dans le passé. Ainsi, pour faire une prédiction on prend les données historiques les plus récentes disponibles et on cherche parmi ces données, les K plus proches instances appelés aussi les K plus proches vecteurs.

La méthode KNN peut être décrite comme suit :
Soit X_t , t=1…T une série chronologique d'un actif financier donné. On commence par construire la matrice suivante :

$$\begin{pmatrix} x_1 & x_{1+\mu} & \cdots & x_{1+(m-1)\mu} \\ x_2 & x_{2+\mu} & \cdots & x_{2+(m-1)\mu} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ x_{T-(m-1)\mu} & x_{T-(m-1)\mu+\mu} & \cdots & x_T \end{pmatrix} = \begin{pmatrix} X^1 \\ X^2 \\ . \\ . \\ X^T \end{pmatrix} \tag{3.1}$$

Où :

$$X^i = (x_i, x_{i+\mu}, x_{i+2\mu}, \dots, x_{i+(m-1)\mu}) \tag{3.2}$$

Avec $i = 1 \ldots T - (m-1)\mu$. m et μ sont appelés respectivement la dimension et le retard de reconstruction.

La série originale est reconstruite par le théorème de Takin (Takens, 1981) et (J.P Beckman et D.Ruelle 1985 et A.M Fraser 1989). Par la suite, il faudra trouver comment déterminer les deux paramètres m et μ. Souvent, dans le cas de prédiction des séries financières, on considère que μ=1. Trois raisons justifient ce choix : d'une part la simplification de l'analyse. D'autre part, la réduction de la complexité et le temps de calcul.

Concernant le choix du paramètre de la dimension m, il y a la méthode des faux plus proches voisins (Kannel et al.1992). Pour μ=1 on a :

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ x_{T-m} & x_{T-m+1} & \cdots & x_{T-1} \end{pmatrix} = \begin{pmatrix} X^1 \\ X^2 \\ . \\ . \\ X^{T-m} \end{pmatrix} \tag{3.3}$$

Pour ce faire, on choisit les K vecteurs X^i qui minimisent la distance par rapport a X^{T-m+1}. Concernant le choix de la distance, la distance Euclidienne a été amplement utilisée dans la littérature. Cependant, la distance de déformation temporelle dynamique "Dynamic Time warping" est plus adaptée aux séries temporelles. Cette distance est basée sur la programmation dynamique et cherche un appariement optimal entre deux séries temporelles en alignant les formes similaires. Contrairement, à la distance euclidienne qui apparie les points des deux séries linéairement (nième élément de la série 1 avec le nième élément de la série 2) indépendamment de leur similarité. L'idée est représentée dans la figure 3.4.

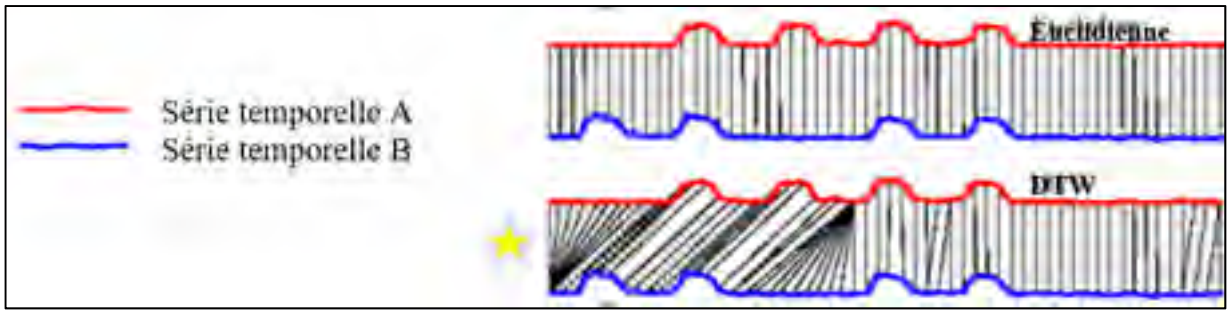


Figure 3.4 Différence entre la distance euclidienne et la distance de déformation temporelle dynamique

Pour une nouvelle observation, la classe est déterminée en se basant sur les classes des K plus proches voisins. C'est-à-dire les observations minimisant la distance avec les entrées de l'observation à prédire. Il s'agit simplement de la classe la plus fréquente parmi ses voisins. Soit :

$$\hat{Y} = Argmax \sum_{k=0}^{n} 1\{yj = y\} \tag{3.4}$$

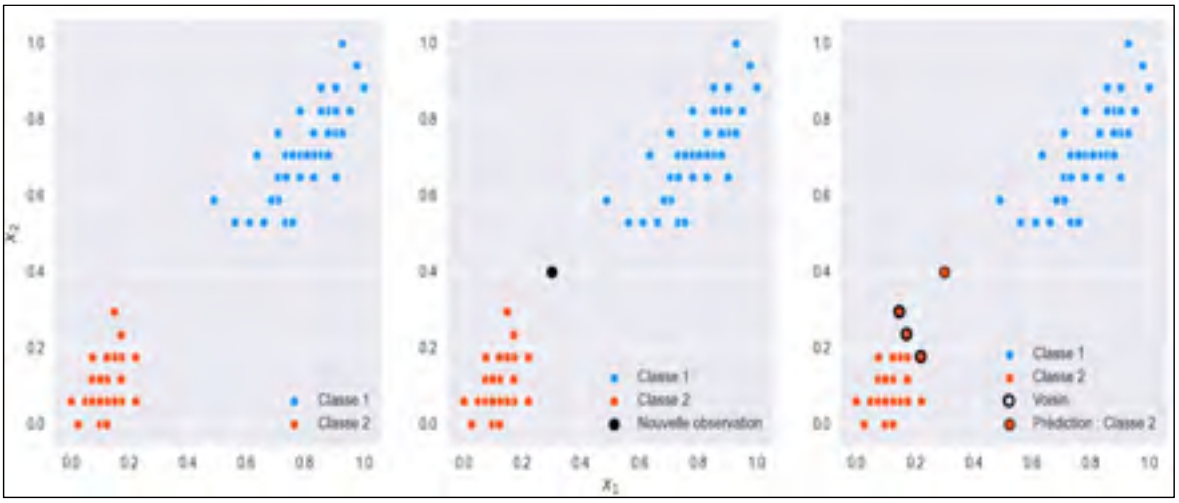


Figure 3.5 Visualisation d'un exemple de classification en utilisant la méthode KNN (K=3)

3.3.2 La régression logistique

La régression logistique est un modèle prédictif qui cherche à trouver des associations entre un vecteur de variables explicatives aléatoires, de nature numérique ou catégoriques, ($x_1, x_2 ..., x_k$), et une variable dépendante catégorique binaire ou multinomiale Y. Il s'agit tout simplement une transformation non linéaire de la régression linéaire, où on cherche à prédire une classe à la place d'une valeur numérique continue. Pour ce faire, on utilise une fonction logistique appelé aussi 'Sigmoïde', pour retourner les probabilités qui sont utilisées pour séparer les classes à prédire. Exemple : si la probabilité générée par une régression logistique, pour prédire la tendance d'un titre, est supérieure à 0.5, alors la tendance prédite sera haussière, si non, elle est baissière.

Ce modèle produit une courbe logistique, qui est limitée aux valeurs comprises entre 0 et 1.

Soit :

- Y la variable indépendante (à prédire),
- X = $(X_1, X_2, ...X_k)$ les variables dépendantes ou explicatives.

Comme le modèle est linéaire, on peut écrire l'équation de régression comme suit :

$$Y^{(i)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n \tag{3.5}$$

Les β_i sont les paramètres du modèle que nous cherchons à estimer pour obtenir notre fonction de prédiction. On utilise le logarithme naturel des « cotes » de la variable dépendante pour construire la courbe de la régression logistique, plutôt que la probabilité conditionnelle :

$$\ln\left(\frac{p(Y=1 \mid X)}{p(Y=0 \mid X)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n \tag{3.6}$$

$$p(Y=1 \mid X) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n)}} \tag{3.7}$$

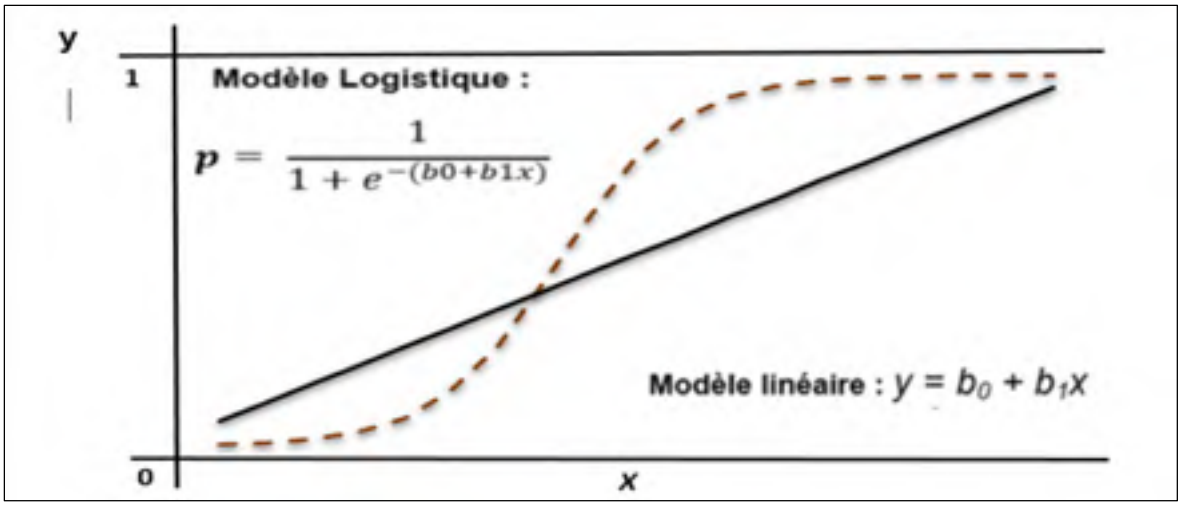


Figure 3.6 Comparaison entre une régression linéaire et une régression logistique

3.3.3 Les machines à support vectoriel

Les machines à support vectoriel ou 'Support Vector Machine' en anglais (SVM) est un algorithme d'apprentissage supervisé qui peut être utilisé pour des problèmes de classification ou de régression. Cependant, il est principalement utilisé pour faire de la classification. Dans cet algorithme, dans un espace à N dimensions (N est le nombre d'attributs dans les données), on cherche l'hyperplan de marge optimale qui sépare correctement les données tout en étant le plus éloigné possible de toutes les observations. L'objectif étant de minimiser l'erreur de généralisation (Vapnik, 2013). La marge correspond à la plus petite distance entre le plan séparateur et un des exemples d'entrainement.

Description de l'algorithme SVM

Soit X_t un exemple d'entrainement, $X = \{x^t, r^t\}$ où $r^t = \begin{cases} +1 \; si \; xt \in C1 \\ -1 \; si \; xt \in C2 \end{cases}$

la distance entre cet exemple et le plan de paramètres w et w_0 est :

$$dist(x_t;\, w, w_0) = \frac{|w_0 + w^T x^t|}{||w||} \tag{3.9}$$

La marge correspond donc à :

$$marge(w, w_0) = min_t \; dist(x_t;\, w, w_0) \tag{3.10}$$

On trouve L'hyperplan séparateur optimal en cherchant w, w_0 tel que :

$$r^t(w_0 + w^T x^t) \geq +1 \qquad \textit{(Engelbrecht, 2007)}$$

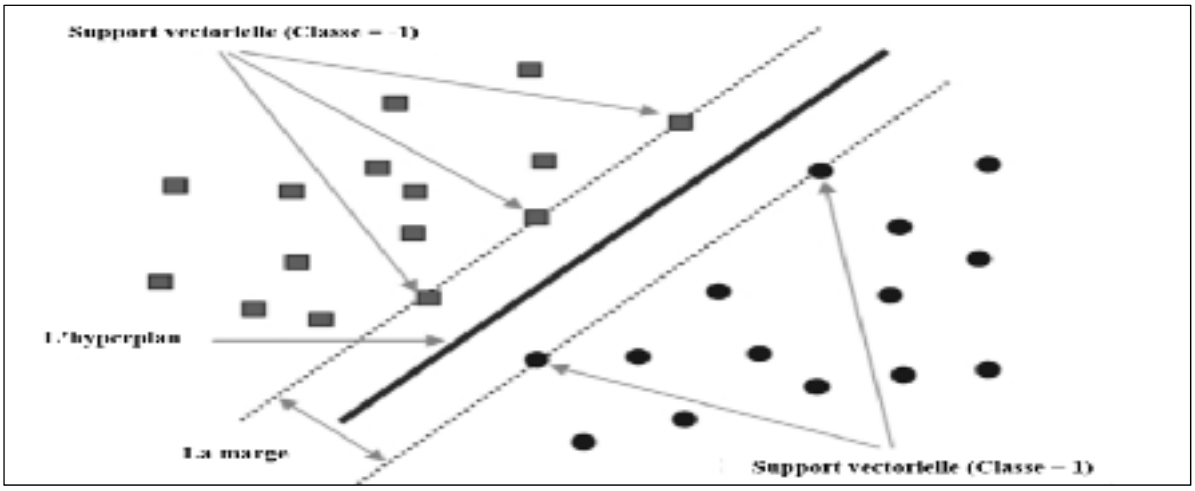


Figure 3.7 Exemple d'un SVM avec un noyau linéaire

3.3.4 Les réseaux de neurones

Le réseau de neurones artificiel (RNA) contient un ensemble de neurones artificiels fortement interconnectés inspirées des neurones biologiques du cerveau. Le but est d'imiter certaines fonctions du cerveau humain, tel que la mémorisation par association, l'apprentissage par exemple, etc.

3.3.4.1 Réseau de neurones biologiques

Le cerveau est composé d'un grand nombre de neurones biologiques, chacun composé d'un corps cellulaire, de dendrites et d'un axone (voir Figure 3.5). Ces neurones sont interconnectés et créent ainsi un système neuronal plus vaste. Les connexions, appelées synapses, sont établies entre la dendrite d'un neurone et l'axone d'un autre. Lorsqu'un neurone reçoit un signal via l'axone, il active ou non ce signal, dépendamment de sa force, et le transmet via ses dendrites à l'ensemble des neurones connectés (Touzet, 1992).

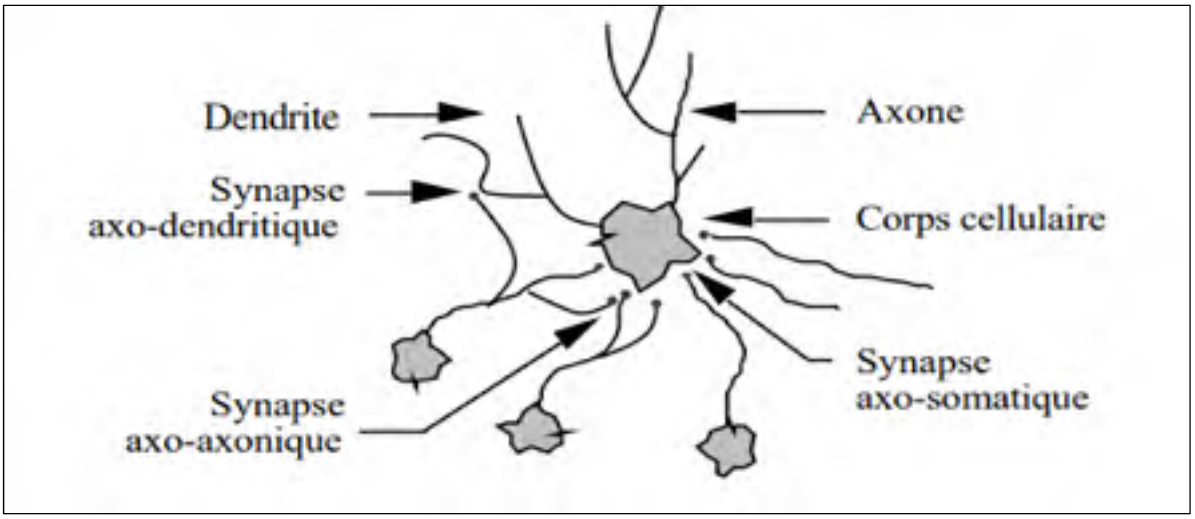


Figure 3.8 Un neurone avec son arborisation dendritique
Tirée de Touzet (1992, p.11)

3.3.4.2 Réseau de neurones artificielles

Les réseaux de neurones artificielles (RNA) sont constitués de couches d'entrée et de sortie, ainsi que (dans la plupart des cas) d'une ou plusieurs couches cachées composées d'unités qui transforment l'entrée en quelque chose que la couche de sortie peut utiliser. L'objectif du RNA est de déterminer un ensemble de poids Θ (entre les nœuds d'entrée, masqués et de sortie) qui minimisent la somme totale des erreurs. Le calcul se fait en utilisant la somme des entrées pondérés par les poids (fonction de la force des connexions).

$$n = \sum_{i=0}^{n} \theta i \, . xi \tag{3.11}$$

n est appelé le signal d'entrée.

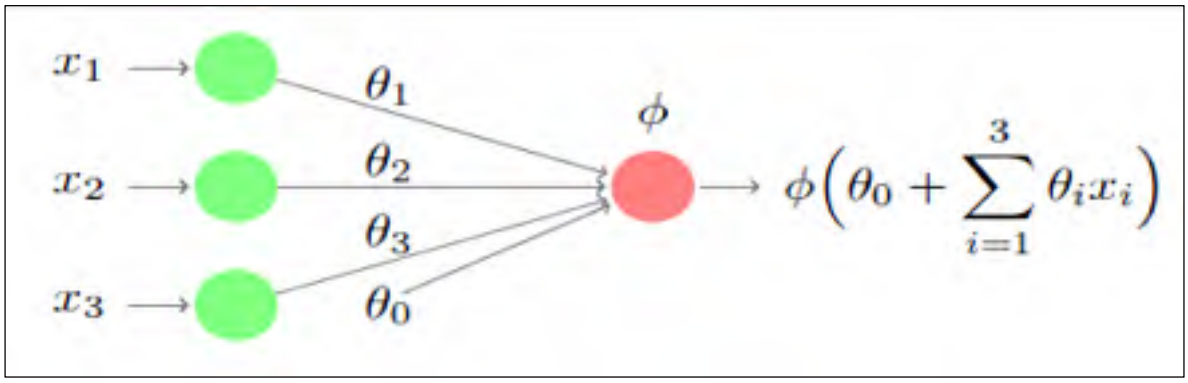


Figure 3.10 Exemple d'un neurone artificiel avec trois entrées
Tirée de Touzet (1992, p.11)

Pendant l'entraînement, les poids θi sont ajustés en fonction d'un paramètre d'apprentissage λ ∈ [0, 1] jusqu'à ce que les sorties deviennent cohérentes avec la sortie.

Chaque neurone a une fonction d'activation qui va calculer la valeur du signal du neurone. Le neurone compare ensuite la somme pondérée des entrées à une valeur de seuil et fournit alors une réponse en sortie (Haykin & Network, 2004).

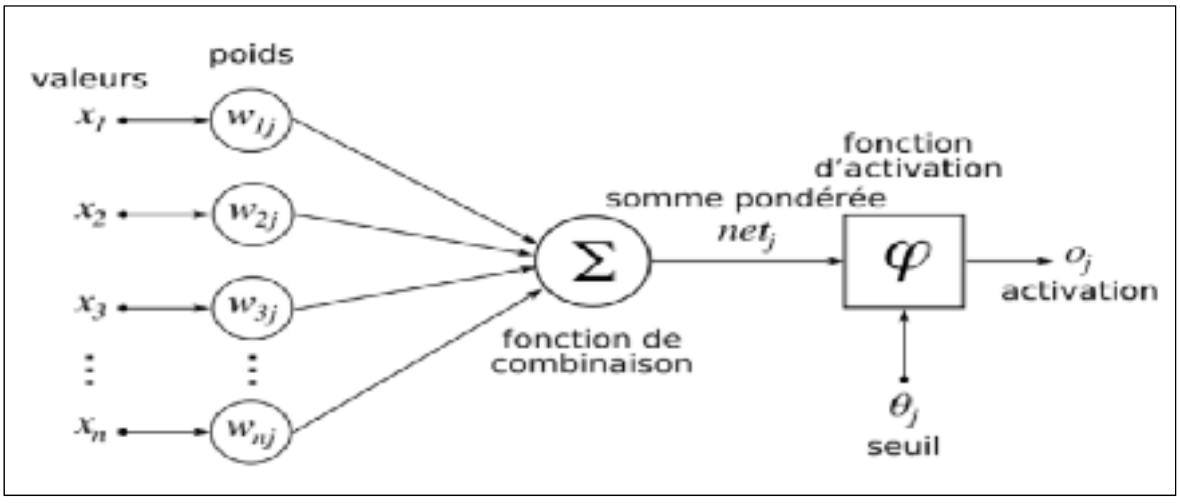
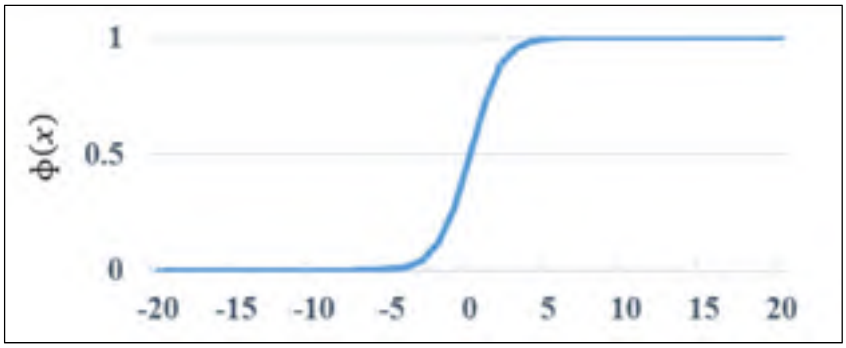


Figure 3.11 Exemple d'un réseau de neurones avec une fonction d'activation
Tirée de Voix schématique d'un neurone artificiel avec un index j (wikipedia, 2005)

Il y a plusieurs types de fonctions d'activation. La plupart sont continues et leurs valeurs possibles sont définies sur les intervalles [0, +1] ou [-1, +1]. On peut citer quelques exemples de fonctions d'activation pour le neurone artificiel :

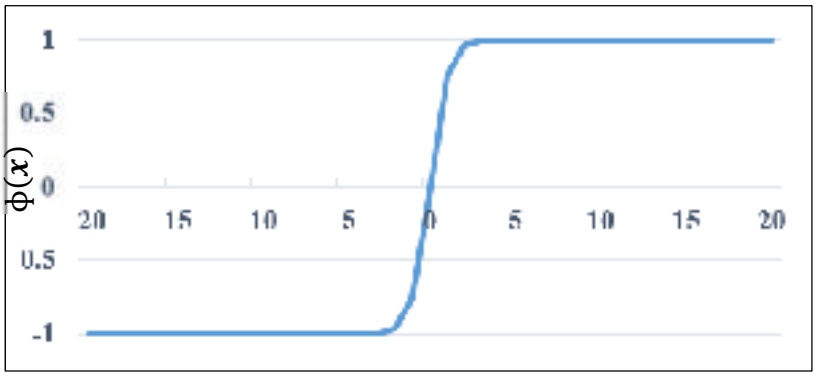
- fonction sigmoïde (logistique) :

$$\phi(x) = \frac{1}{(1 + e^{-x})}$$



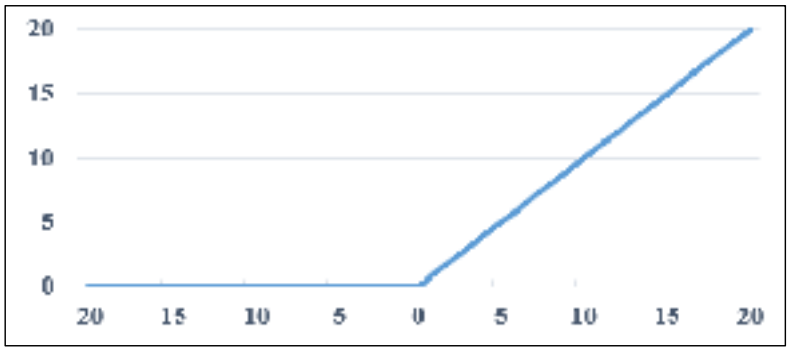
- fonction tangente hyperbolique (tanh) :

$$\phi(x) = \frac{2}{(1 + e^{-2x})} - 1$$



- fonction d'unité linéaire rectifiée (relu) :

$$\phi(x) = \begin{cases} 0, & \text{si } x < 0 \\ x, & \text{si } x \geq 0 \end{cases}$$



En combinant plusieurs neurones on obtient un perceptron multicouche (MLP). Il contient au moins trois couches : une d'entrée, une couche cachée et une couche de sortie. Chaque nœud a sa propre fonction d'activation non linéaire, à l'exception du nœud d'entrée.

L'algorithme de la Descente de gradient :

Soit h(θ) : X -->Y un modèle paramétrique avec une fonction du coût J(θ) continue et dérivable à mesurer, et un seuil de tolérance ϵ >=0.

L'objectif est de minimiser la fonction J(θ) par essai erreur jusqu'à convergence de :

$$\theta_{t+1} \leftarrow \theta_t - \alpha . \frac{\partial J(\theta_t))}{\partial \theta_t}$$

Où $\alpha \; est \; un \; taux \; d'apprentissage.$

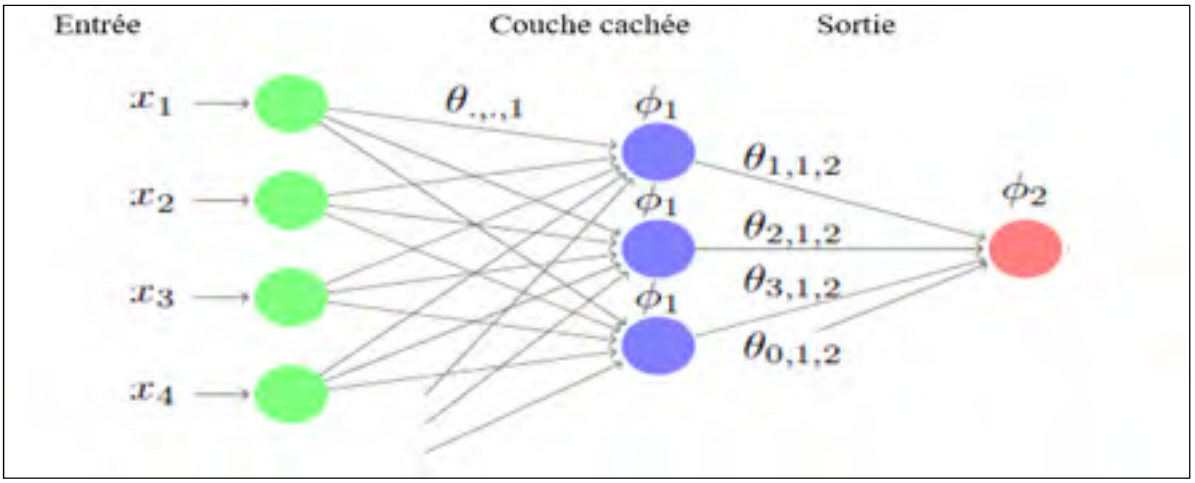


Figure 3.12 Exemple d'un perceptron multicouche
avec quatre entrées et une couche cachée
Tirée de Ali *et al.* (2002, p. 3)

L'algorithme utilise une technique qui s'appelle la rétropropagation (Engelbrecht, 2007). L'optimisation des paramètres est faite en utilisant l'algorithme de la Descente de gradient.

3.3.5 Les arbres de décision

Les arbres de décision font partie des méthodes d'apprentissage supervisés non paramétriques les plus utilisées en classification et en régression. D'une part à cause de leur simplicité algorithmique et d'une autre part, à cause de la facilité à les interpréter et expliquer les résultats générés.

Les arbres de décision sont construits via une approche algorithmique et peuvent être visualisées sous forme d'arbre avec des règles qui identifient les façons de fractionner un ensemble de données. L 'objectif est de créer un modèle qui prédit la valeur d'une variable cible en apprenant les règles de décision.

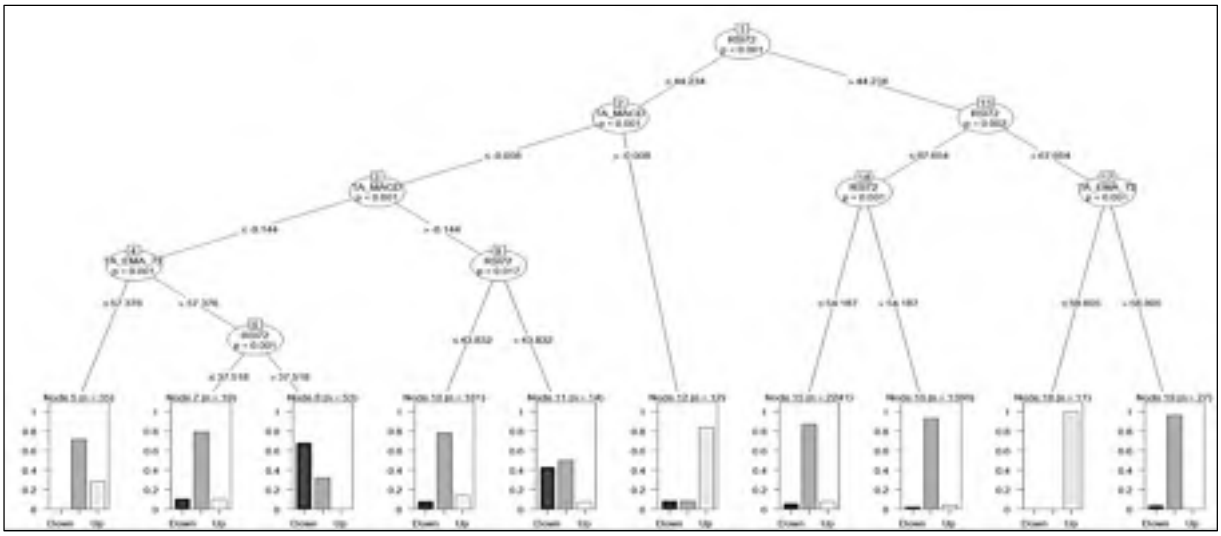


Figure 3.13 Exemple d'arbre de décision pour prédire la tendance du contrat à terme E-Mini- S&P

L'exemple précédant montre comment nous avons utilisé un arbre de décision pour facilement extraire et visualiser les règles qui peuvent être valides et utilisées dans la négociation. Dans cet exemple, nous avons utilisé trois indicateurs techniques : le RSI, le MACD et une moyenne mobile exponentielle pour prédire la tendance du prix d'un contrat futur à chaque cinq minutes. Il est à noter qu'il existe plusieurs versions algorithmiques d'arbres de décision. Pour ce travail, nous utiliserons les deux versions suivantes : les arbres d'inférence conditionnelle et la version C5.0 (Salzberg, 1994). Ce choix s'explique par l'efficience de ces deux algorithmes, la robustesse au bruit et nécessitent moins de mémoire pour traiter les données.

3.3.6 Les méthodes d'ensemble

Ces techniques sont des méta-algorithmes qui consistent à combiner plusieurs modèles uniques de base, comme les arbres de décision, dans un même modèle prédictif. L'objectif est d'améliorer la généralisation et la robustesse de nos modèles. En effet, statistiquement parlant, la moyenne d'un ensemble d'échantillons est plus fiable que celle d'un seul échantillon. Exemple : si nous avons un groupe d'experts financiers qui doivent se prononcer sur la tendance future d'un titre. Alors, il est plus probable que la décision du groupe à majorité soit correcte que l'avis d'un seul expert.

Les méthodes d'ensemble peuvent être divisées en deux catégories : les méthodes d'ensemble parallèles et les méthodes d'ensemble séquentielles.

3.3.6.1 Méthodes d'ensemble parallèles (Bagging)

Pour ces méthodes, les modèles de base (exemple : les arbres de décision) sont générés de façon indépendante et en parallèle (par exemple, les forêts aléatoires). La motivation derrière ces méthodes est que l'erreur de prédiction peut être réduite de manière significative en réduisant la variance.

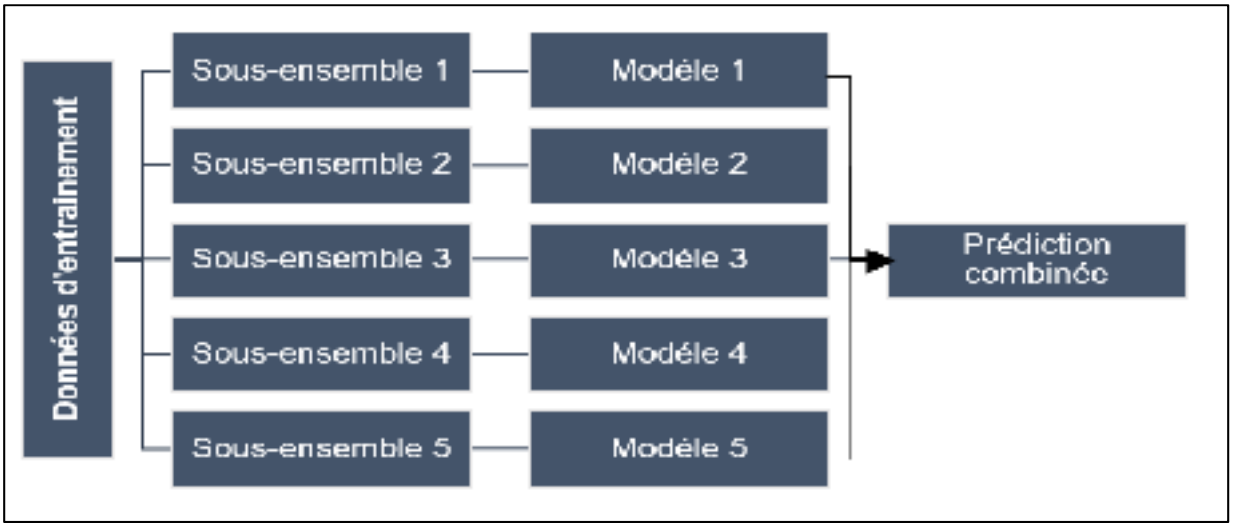


Figure 3.14 Schéma illustrant le fonctionnement des méthodes d'ensembles parallèles (Bagging)

- **la forêt aléatoire**

  La forêt aléatoire est un algorithme d'apprentissage supervisé d'ensemble qui est construit à partir de plusieurs modèles de base (en général des arbres de décision). Ils sont ensuite fusionnés pour obtenir une prédiction plus précise et plus stable.

  Chaque arbre de l'ensemble est construit à partir d'un échantillon généré aléatoirement avec remplacement à partir d'un ensemble de données d'apprentissage. Le diagramme suivant fournit une représentation visuelle de la logique de l'algorithme.

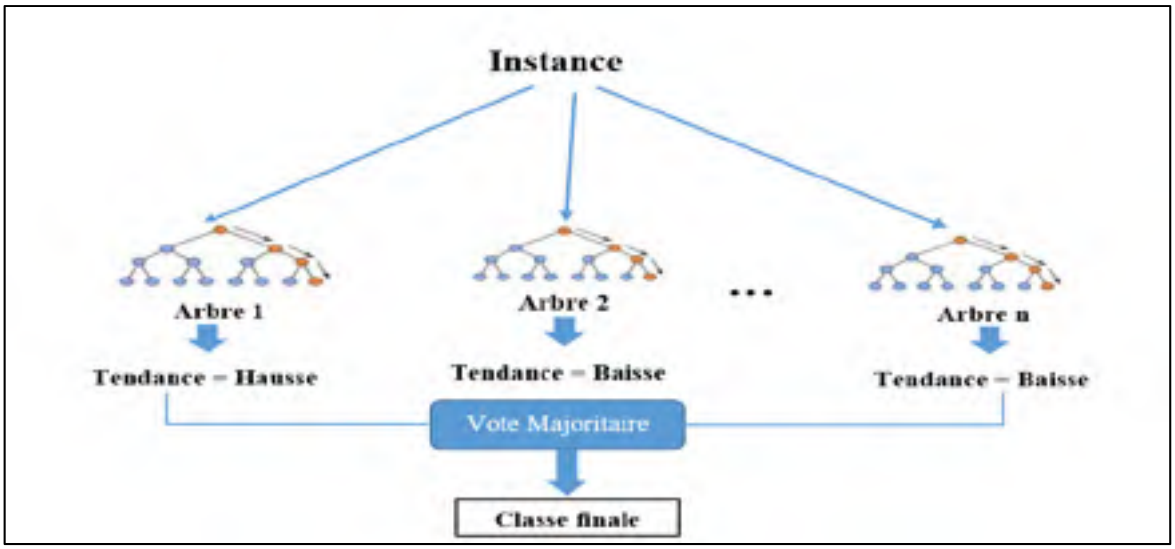


Figure 3.15 Description de l'algorithme de la forêt aléatoire

- **les arbres extrêmement aléatoires**

  'Extremely Randomized (Extra-trees)' en anglais. C'est une autre méthode d'ensemble conçu spécialement pour utiliser les arbres de décision. Le caractère aléatoire est relié à la façon dont les fractionnements sont calculés. À l'opposé des forêts aléatoires, au lieu de rechercher les seuils les plus discriminants, des seuils sont tirés au hasard pour chaque attribut et le meilleur de ces seuils générés aléatoirement est choisi comme règle de fractionnement (Bühlmann, 2012).

3.3.6.2 Méthodes d'ensemble séquentielles (Boosting)

Pour ces méthodes, les classifieurs de base sont générés de manière séquentielle (par exemple AdaBoost) et dépendante, contrairement aux méthodes parallèles. À chaque fois qu'un classifieur de base est entrainé, les instances mal classées précédemment sont pondérées avec un poids plus élevé, dans le but que lors des prochaines itérations, les nouveaux modèles corrigent les erreurs des modèles précédents, ce qui devra améliorer la performance globale.

- **l'algorithme AdaBoost**

  C'est un algorithme de classification qui vise à utiliser des classifieurs de base pour construire un ensemble puissant. À la fin, le modèle génère une classification agrégée. En effet pour chaque modèle de base l'algorithme attribue un poids en fonction de la performance individuelle de ce dernier. L'idée est que les classifieurs devront se concentrer sur les observations difficiles à classer correctement. Tout algorithme d'apprentissage automatique peut être utilisé comme classifieur de base s'il accepte les poids sur l'ensemble d'entrainement.

  Description de l'algorithme AdaBoost :

  Considérons un ensemble d'entrainement S = $(x1, y1) \dots (x_n, y_n), xi \in X, yi \in \{-1, 1\}$

  L'algorithme fonctionne comme suit :

  1. Initialiser les probabilités de chaque donnée $p_1^t = \frac{1}{N}, t = 1, \dots N$
  2. Entrainer le classifieur h_j avec X_j (généré à partir de X selon la probabilité p_j^t)
  3. Calculer l'erreur de $h_j : \epsilon_j = \sum_{j=1}^{N} P_j^t \, \delta(yi \neq hj(xi))$ (δ *est une fonction indicatrice*)
  4. Choisir un facteur $\alpha t = \frac{1}{2} \ln(\frac{1-\epsilon t}{\epsilon t})$
  5. Mettre à jour la distribution des poids à travers des exemples :
  6. $D^{(t+1)}(i) = \frac{D^{(t)}(i).e^{-\alpha t.yi.ft(xi)}}{Z^{(t)}}$ $\forall\, i \in \{1, 2, \dots m\}$

  Résultat = la classe voté $\forall\, x$, *C(x) sign* $\sum_{t=1}^{T} \alpha_t \,.\, f_t(x)$

- **l'algorithme de gradient boosting machine :**

  En abrégé GBM, de l'anglais 'Gradient Boosting Machine'. C'est un autre algorithme d'ensemble séquentielle similaire au 'AdaBoost' avec la seule différence que les poids sont optimisés en utilisant la descente du gradient pour minimiser la fonction de perte. Il est similaire à la manière dont les réseaux de neurones l'utilisent pour optimiser les poids à apprendre.

Il existe une variété d'implémentations du GBM. On peut citer XGBoost qui utilise des approximations plus précises pour trouver les meilleurs modèles de base. Il y a aussi le GBM léger (LGBM), qui est une version améliorée du GBM et qui est conçue pour être plus efficace, car il utilise moins de mémoire et génère une performance meilleure et capable de gérer des données à grande échelle.

3.3.6.3 Le stacking

Il s'agit d'un procédé qui consiste à combiner plusieurs modèles très différents dans le but d'améliorer la qualité de la prédiction finale. Le fonctionnement de cette technique est décrit dans le schéma en bas.

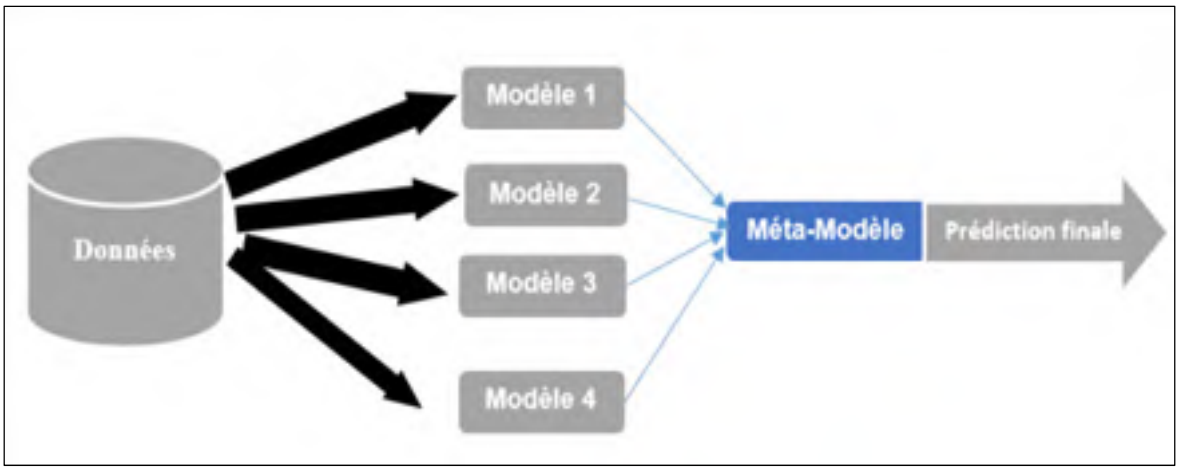


Figure 3.16 Description du fonctionnement du Stacking

On peut par exemple utiliser une combinaison de modèles, réseau de neurone artificielle, une régression logistique et un SVM pour construire un seul modèle. On attribuera un poids à chaque modèle pour aboutir à notre modèle agrégé.

3.3.6.4 Conclusion sur les méthodes d'ensemble

Les méthodes d'ensemble ont connu du succès dans une variété de problèmes et elles ont démontré qu'ils sont capables de faire face à plusieurs défis d'apprentissage automatique comme le biais et la variance de prédiction. Toutefois, certains désavantages de ces méthodes méritent d'être soulignés. Cela inclut la mémoire supplémentaire et les coûts de calcul en combinant plusieurs classifieurs et en les gardant prêts à être traités. La perte d'interprétabilité peut également être une source de préoccupation, mais à échelle moins important pour des prédictions financières, vu qu'on s'intéresse plus à la maximisation de la performance des prédictions.

Conclusion

Dans ce chapitre nous avons introduit et expliqué le fonctionnement des algorithmes d'apprentissage automatique les plus populaires qui peuvent être utilisé pour faire de la classification. Dans le chapitre 5 et 6 nous allons montrer comment nous les avons utilisés dans le cadre de ce projet afin de prédire la tendance des actifs financiers. Ça inclus l'entrainement de ces algorithmes avec des données historiques et le choix des paramètres. Quand ces deux tâches sont faites, on parle d'un modèle qui prend des données d'entrée dans le but de générer des prédictions (données de sortie).

Une autre tâche qui est très importante au moment où nous cherchons à choisir entre plusieurs modèles ou à savoir si notre modèle génère des résultats qui font du sens, consiste à évaluer la qualité des sorties de nos modèles, en les comparant avec des résultats réels. Nous allons consacrer le prochain chapitre à cette tâche, où nous allons présenter les principales techniques et métriques d'évaluation des modèles prédictifs dans le contexte de la classification.

CHAPITRE 4

L'ÉVALUATION DE LA PERFORMANCE DES MODÈLES D'APPRENTISSAGE AUTOMATIQUE

La performance des modèles prédictifs dépend de plusieurs facteurs. Entre autres, les données utilisées pour construire et entrainer ces modèles et le choix des paramètres d'optimisation. Si nous n'utilisons pas des métriques qui mesurent correctement la performance par rapport à notre problème, ou des méthodes qui ne sont pas robustes, nous pouvons conclure à tort que nous avons construit de bons modèles. En apprentissage automatique, l'évaluation de la performance des modèles est une tâche cruciale et complexe en même temps. C'est pourquoi, il faut l'effectuer soigneusement pour que les résultats reportés soient fiables.

Ce chapitre aborde les métriques et les méthodes qui peuvent être utilisés pour tester et évaluer la performance prédictive des modèles d'apprentissage automatique, pour les problèmes de classification. Dans le contexte des métriques d'évaluation, on commence par définir c'est quoi une matrice de confusion. Ensuite, on présente les mesures qui sont directement calculées à partir de cette matrice (précision, rappel, spécificité et taux des faux négatifs) ainsi que la courbe ROC (Receiver Operating Characteristic en anglais).

Plusieurs techniques d'estimation d'erreur ou d'échantillonnage appartenant à la famille de la validation croisée, ainsi que d'auto-amorçage (Bootstraping), seront abordées dans le contexte des méthodes d'évaluation.

4.1 Les métriques de mesure de la performance des modèles de classification

Il est très important de tester, mesurer et surveiller la performance d'un modèle prédictif avant et après l'avoir déployé en production. On doit alors définir les mesures à utiliser pour l'évaluation de cette performance.

Les métriques d'évaluation à utiliser dépendent de plusieurs facteurs. Telle que, la tâche de modélisation (classification, régression ou segmentation), le contexte du problème que nous essayons de résoudre ainsi que la distribution des données. Ces métriques servent aussi à comparer la performance des modèles et sélectionner ceux qui donnent la meilleure performance.

Étant donné que la tâche principale de ce travail est de prédire la classe de la tendance d'un actif financier. Alors, on va se limiter aux métriques de classification.

4.1.1 La matrice de confusion

Il s'agit d'un tableau de taille n x n pour visualiser les résultats des modèles prédictifs pour les problèmes de classification. Où 'n' est le nombre de classes dans nos ensembles de données. Dans cette matrice on croise les classes cibles réelles avec les classes prédites obtenues. Ceci nous donne le nombre d'instances qui sont correctement classées et le nombre d'instances mal classées.

Tableau 4.1 Exemple de matrice de confusion pour une classification binaire

		Classes actuels		
		Positif	**Négatif**	
Classes prédites	**Positif**	VP	FP	Valeur prédictive positive (VPP)
	Négatif	FN	VN	Valeur prédictive négative (VPN)
		Rappel	Spécificité	Taux de succès = (VP + VN) / (VP+FP+FN+VN)

On choisit une classe comme la classe positive. Exemple : la tendance haussière.

VP : c'est le nombre des vrais positifs, le nombre d'instances positives correctement classifiées,
FP : c'est le nombre des faux positifs, le nombre d'instances qui ne sont pas positives et qui sont prédites comme positives,
FN : c'est le nombre des faux négatifs, le nombre d'instances non négatives classifiées comme négatives,
VN : c'est le nombre des vrais négatifs, le nombre d'instances négatives correctement classifiées.

À partir de la matrice de confusion on peut calculer plusieurs métriques.

- **le taux de succès (Accuracy ou Hit ratio)** : C'est la proportion des instances qui sont correctement classifiées.

$$A = \frac{(\mathrm{VP} + \mathrm{VN})}{(\mathrm{VP} + \mathrm{FP} + \mathrm{FN} + \mathrm{VN})} \tag{4.1}$$

- **le taux d'erreur (E)**: c'est l'erreur global de la classification.

$$E = 1 - taux\ de\ succés \tag{4.2}$$

- **la valeur prédictive positive**: c'est la proportion des instances actuelles positives parmi les instances prédites comme positives

$$\mathrm{VPP} = \frac{\mathrm{VP}}{(\mathrm{VP} + \mathrm{FN})} \tag{4.3}$$

- **la valeur prédictive négative** : c'est la proportion des instances actuelles négatives parmi les instances prédites comme négatives

$$VPN = \frac{VN}{(VN + FN)} \tag{4.4}$$

- **le rappel** : appelé aussi sensitivité: c'est le pourcentage des instances positives correctement identifiées.

$$Rappel = \frac{VP}{(VP + FN)} \tag{4.5}$$

- **la spécificité** : c'est le pourcentage d'instances négatives correctement identifiées.

$$Spécificité\ = \frac{FP}{(FP + VN)} \tag{4.6}$$

- **la mesure F** : Il s'agit de la moyenne harmonique de la précision et le rappel. Il prend des valeurs entre 0 et 1. Cette mesure est utilisée quand on cherche une balance entre la précision et le rappel. Plus que la valeur de cette mesure s'approche de 1, plus que notre modèle fait une bonne décision, et le contraire est vrai, quand elle s'approche de zéro.

$$F - mesure\ = 2.\frac{(Précision\ .\ rappel)}{(Précision + rappel)} \tag{4.7}$$

Le taux de succès (Accuracy en anglais) est la métrique la plus utilisée pour évaluer la performance des modèles de classification. Toutefois, cette métrique peut nous induire en erreur si elle n'est pas combinée avec d'autres mesures, surtout quand les classes des données utilisées ne sont pas uniformément distribuées (équilibrées).

Pour la prédiction de la tendance des séries temporelles financières, la distribution non équilibrée des classes est un grand défi.

Dans ce contexte, le taux de succès n'est pas une mesure appropriée pour évaluer la performance de nos modèles, car la classe majoritaire pourra dominer les prédictions générées par nos modèles. Exemple : Si on suppose que nous voulons faire une classification binaire pour un problème de classification avec des classes non équilibrées, Où la classe majoritaire (B) représente 95% des instances. Si on suppose que la taille de l'ensemble test est de 1000 instances et que notre modèle classe toutes les instances comme (B), Nous avons la matrice de confusion suivante :

		Classes actuels	
		A	B
Classes prédites	A	0	0
	B	50	950

En utilisant la formule dans 4.1 on obtient un taux de succès de 95% même si notre modèle a mal classé toutes les instances positives (A). Si on s'intéresse à la détection des cas positifs, comme c'est le cas de la majorité des problèmes en pratique, le ratio utilisé pour évaluer la performance recherchée ne reflète pas le besoin de la modélisation. Dans ce cas, nous devrons utiliser une autre métrique qui prend en considération la distribution des classes.

4.1.2 La statistique de Cohen Kappa

Une autre métrique qui est utile surtout dans les cas où les données ne sont pas équilibrées ou pour des cas de classification multiple, est la statistique de Cohen Kappa. Cette statistique mesure l'accord entre deux classifieurs qui classent chacun n éléments dans C classes mutuellement exclusives. Elle inclut la distribution marginale de la variable cible dans le calcul de la performance.

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \tag{4.8}$$

$$p_0 = Taux\ de\ succés\ (A) \tag{4.9}$$

Où :

p_0 est l'accord observé (la même chose que le taux de succès dans 4.1),

p_e est'accord attendu. Il s'agit d'une probabilité hypothétique que l'accord soit dû au hasard.

$$p_e = \frac{(VN + FP) * (VN + FN) + (FN + VP) * (FP + VP)}{N^2} \quad (4.10)$$

Cette statistique permet de comparer le taux de succès observé avec celui qui serait obtenu si on suppose que les classifications sont générées de façon aléatoire. En gros, elle indique à quel point notre classifieur est plus performant que les classifications générées au hasard en fonction de la fréquence de chaque classe.

Interprétation :

- lorsque le kappa = 1, Nous pouvons conclure qu'il s'agit d'un accord parfait entre les classifications générées et les vraies observations.
- si le kappa a pour valeur 0, L'accord obtenu est le même que celui qui pourrait être obtenu par hasard.
- lorsque le kappa < 0, l'accord est pire que celui qui pourrait être obtenu par hasard.

4.1.3 La courbe ROC

La courbe ROC (Receiver Operating Characteristic) a été utilisée dans le traitement de signal pour faire la distinction entre le signal et le bruit. Elle est très utilisée en apprentissage automatique pour évaluer la performance des classifieurs. Il s'agit d'une courbe où on croise le taux des vrais positifs (TVP) avec celui des faux négatifs (TFN) pour tous les seuils de classification. On utilise un classifieur aléatoire comme une ligne de base. Si on classe aléatoirement nos instances, on est censé avoir cette ligne.

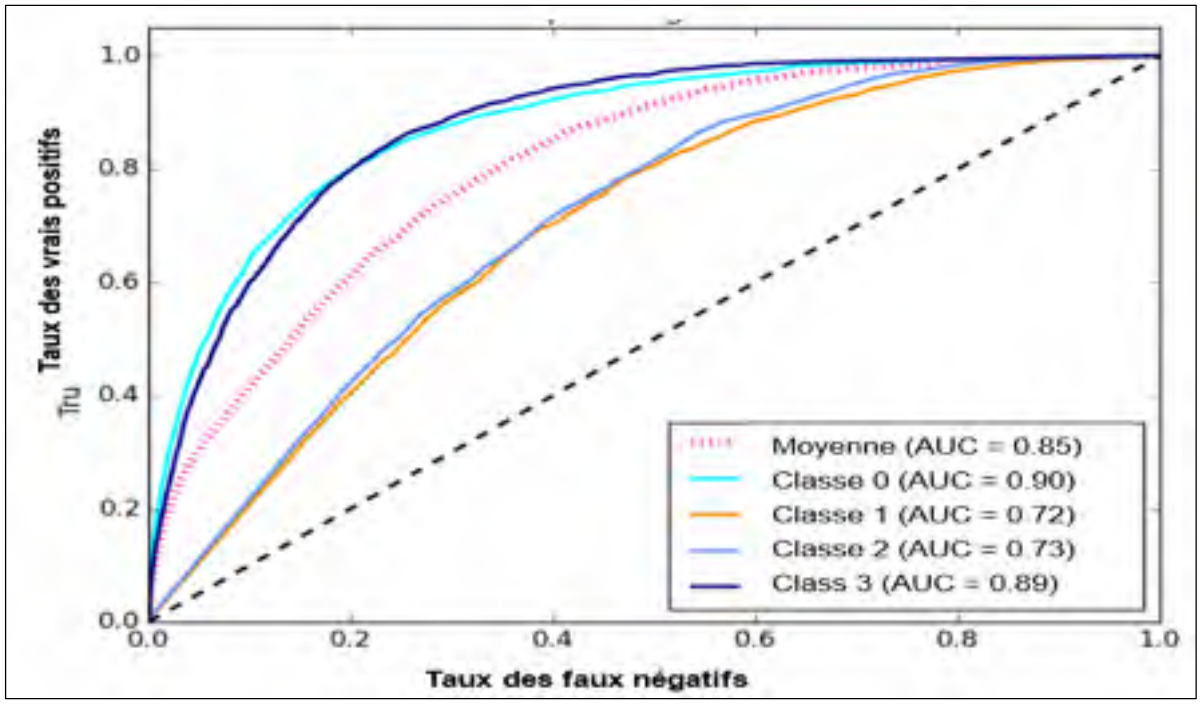


Figure 4.1 Exemple de courbes ROC pour des classes multiples

Un modèle avec une discrimination parfaite (sans chevauchement entre les classes) passe par le coin supérieur gauche. Plus que la courbe ROC est proche du coin supérieur, meilleure est la performance de la classification.

L'aire sous la courbe ROC (AUC : Area Under the Curve) représente une mesure qui permet de quantifier numériquement la performance de nos classifieurs :

- si AUC = 1, Il s'agit d'un modèle qui fait une séparation parfaite entre nos classes. Il permet de classer toutes les instances positives correctement et fait la même chose avec les autres instances.

- si AUC = 0.5, la classification n'est pas meilleure que celle qui serait obtenue si nous générons aléatoirement nos instances. Le modèle dans ce cas, ne fait aucune distinction entre nos classes. Chaque instance a une probabilité de *1/n* d'être bien classée en utilisant ce modèle. Où n est le nombre de classes.

- si AUC < 0.5, notre modèle fait pire qu'une classification aléatoire. Il vaut mieux deviner aléatoirement, qu'utiliser ce modèle.

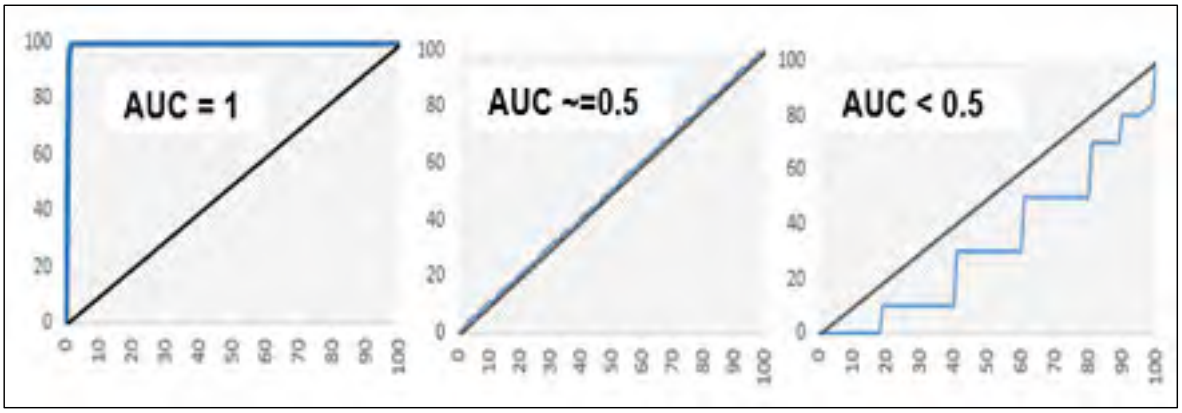


Figure 4.2 L'aire sous la courbe ROC pour 3 scénarios

L'AUC est très utile quand on veut faire une comparaison entre différents modèles. Le meilleur modèle est celui qui a l'AUC le plus élevé.

4.2 Les problèmes d'une faible généralisation des modèles prédictifs

Les causes principales de la faible performance des modèles d'apprentissage automatiques sont le sousapprentissage ou le surapprentissage.

Le but d'un algorithme d'apprentissage automatique supervisé est de trouver une fonction d'approximation, qui utilise des variables d'entrée X dans le but de d'approcher une variable de sortie Y. Cette fonction sert à capturer toutes les propriétés et les corrélations présentes dans l'ensemble des données d'entrainement. Ces dernières sont propres à l'ensemble d'entrainement. C'est pourquoi, Il est important de savoir comment cette fonction généralise la connaissance acquise sur de nouvelles données (non vues).

En effet, il est important d'avoir un ensemble d'entrainement qui est assez représentatif pour une bonne qualité d'inférence.

La généralisation est importante car les données que nous collectons et nous utilisons dans l'apprentissage ne représentent qu'un échantillon, et elles peuvent être incomplètes et contenir du bruit.

En apprentissage automatique, lorsque nous parlons de la capacité d'un modèle d'apprentissage automatique à apprendre et à généraliser à de nouvelles données, on utilise les termes surapprentissage et le sousapprentissage.

4.2.1 Le surapprentissage

Un surapprentissage survient lorsqu'un modèle apprend les détails et le bruit dans les données d'apprentissage de telle sorte que ça impacte de façon négative les performances du modèle pour de nouvelles données. Cela signifie que les fluctuations aléatoires dans les données d'apprentissage sont capturées et apprises en tant que concepts par le modèle. Le problème est que ces concepts ne s'appliquent pas aux nouvelles données et dégradent la capacité du modèle à généraliser correctement.

Dans le cas d'une présence de surapprentissage, le modèle prédictif pourra générer de très bons résultats sur les données d'entrainement, mais à l'opposé, les prédictions qui sont générées sur des données qu'il n'a pas encore vues, ne seront pas de bonne qualité. Dans ce cas, on dit que le modèle souffre de surapprentissage (Overfitting, en anglais).

4.2.2 Le sousapprentissage

On parle de sous-apprentissage ou 'Underfitting' en anglais, quand un modèle ne peut ni modéliser les données d'apprentissage ni se généraliser à de nouvelles données. Dans ce genre de cas, les erreurs de prédiction vont être importantes.

Contrairement au surapprentissage, Il est plus facile de détecter le sousapprentissage. Il suffit d'observer la performance du modèle sur différents ensembles d'entrainement et de

validation. une faible performance sur les données d'entrainement est un signal fort que le modèle souffre de ce biais.

Pour faire face à un tel évènement, il faut tester d'autres algorithmes, modifier les paramètres du modèle, augmenter la taille des données d'entrainement ou ajouter plus attributs à la liste des variables d'entrée, jusqu'à l'obtention d'une performance suffisante, qui est évaluée par une bonne métrique de performance.

Il est essentiel de s'assurer que les modèles prédictifs que nous désirons déployer ne souffrent ni de surapprentissage ni de sousapprentissage. Le modèle à privilégier est celui qui n'a pas une grande variance, et ne souffre pas d'un grand biais.

4.3 Les méthodes d'évaluation des modèles prédictifs de classification

Dans le but de quantifier la fiabilité d'un modèle d'apprentissage automatique, il est nécessaire d'évaluer la performance de ce modèle en utilisant des données qui reflètent la tâche de prédiction et le problème pour lequel le modèle a été construit.

Ce qui est intuitif est de construire et tester la performance d'un modèle en utilisant un seul ensemble qui contient toutes les données disponibles. Dans ce cas, on se trouve à évaluer des prédictions à partir des données qui ont déjà été vu par le modèle utilisé. Il est possible d'utiliser cette approche dans le cas où on s'intéresse plus à construire des modèles descriptifs et non pas prédictifs. C'est-à-dire, que nous cherchons à modéliser la structure des données observées pour une meilleure compréhension de la façon dont les prédictions sont générées.
Exemple, utiliser un arbre de décision pour extraire les règles qui définissent la relation entre les attributs et la variable cible. Ces règles peuvent être utilisées pour expliquer pourquoi un résultat est obtenu.

Dans le contexte où nous cherchons à généraliser la connaissance acquise sur des données non vues, il ne sera pas judicieux de se baser sur les résultats de performance obtenus en testant nos modèles sur les mêmes données utilisées pour l'entrainement, afin de choisir nos meilleurs modèles. Une surestimation de la performance pourra résulter de ce processus.

On peut diviser les méthodes d'évaluation de la performance d'un modèle en deux grandes catégories : validation croisée et échantillonnage de réserve (Hold-out).

4.3.1 La validation croisée

Il s'agit d'une méthode d'évaluation populaire et facile à comprendre. Il en résulte généralement un modèle moins biaisé par rapport aux autres méthodes. Parce qu'il garantit que chaque observation de l'ensemble de données d'origine a la chance d'apparaître dans les ensembles d'entrainement et de test. C'est l'une des meilleures approches si les données d'entrée sont limitées.

4.3.2 Échantillonnage de réserve

C'est une méthode simplifiée de la validation croisée. L'ensemble de données est séparé en deux ensembles, appelés ensemble d'apprentissage et ensemble de test. Le modèle apprend à partir des données d'entrainement uniquement pour ensuite prédire les valeurs de sortie pour les données de l'ensemble de test (il n'a jamais vu ces valeurs de sortie auparavant). Les erreurs qu'il commet s'accumulent pour permettre de calculer le taux d'erreur global, qui est utilisée pour évaluer le modèle. L'avantage de cette méthode est qu'elle est simple et ne nécessite pas beaucoup de calcul. Cependant, son évaluation peut avoir une grande variance. L'évaluation peut dépendre énormément des points de données qui se retrouvent dans l'ensemble d'entrainement et ceux qui se retrouvent dans l'ensemble test. Par conséquent, l'évaluation peut être très différente selon la manière dont la division est effectuée.

Conclusion

Dans ce chapitre nous avons vu les métriques et techniques qui peuvent être utilisés pour l'évaluation de la performance des modèles d'apprentissage automatique. Dans les chapitres suivants nous utiliserons ces métriques et techniques pour l'évaluation de la performance statistique de nos modèles. Ceci nous permettra de voir si nos modèles génèrent des résultats meilleurs que des choix purement aléatoires et comparer les modèles entre eux, dans le but de sélectionner celui ou ceux qui génèrent la meilleure performance.

Le fait d'utiliser des techniques de validation croisée et de glissement de fenêtre temporelle, nous permettra d'améliorer la généralisation de nos modèles et construire des modèles robustes qui résistent mieux au problème de surapprentissage. Dans la phase d'entrainement, nous allons analyser la performance de chaque modèle. Ceci nous permettra de voir s'il existe des modèles qui souffre de sousapprentissage et qu'il faudra éliminer de la liste de nos choix.

CHAPITRE 5

ANALYSE EXPÉRIMENTALE - MÉTHODOLOGIE

Ce chapitre a pour but d'expliquer la méthodologie utilisée pour expérimenter nos modèles d'apprentissage automatique. On commence par une brève description des données utilisées. Ensuite, on expliquera le traitement effectué aux données, tel que le nettoyage, la création d'attributs, la normalisation et l'agrégation des données. Par la suite, chaque ensemble de données collecté sera divisé en deux parties. Une qui sera utilisée pour entrainer et trouver les paramètres de nos modèles et l'autre partie, qui est indépendante de la première, sera utilisée pour tester leurs performances statistiques et la rentabilité générée.

5.1 Les données des actifs utilisées

5.1.1 Les contrats à terme

Les données utilisées pour entrainer et évaluer la performance et la rentabilité des modèles utilisés dans ce projet sont des contrats à terme d'indices boursiers et de matières premières.
Les contrats à terme ou « Futures » en anglais, sont des produits financiers qui servent à prendre un engagement d'achat ou de vente d'une quantité déterminée d'un actif sous-jacent à un prix et une date définie à l'avance.

5.1.2 Les caractéristiques des contrats à terme

Les contrats à terme sont composés des éléments suivants :

- **le sous-jacent** : c'est le produit financier qu'on s'engage à acheter ou à vendre à l'échéance. Exemple : pétrole, or, actions, indices boursiers, devises

- **le terme** : c'est la date à laquelle le contrat n'est plus valide. À cette date, l'opération d'achat ou de vente doit être faite.

À cette date le contrat expire et si la position n'est pas fermée avant cette date, on devient responsable de la marchandise et le contrat devient livrable physiquement.

- **le prix** : Il s'agit du prix auquel le sous-jacent sera acheté ou vendu à l'échéance.

- **la valeur d'un contrat** : c'est la valeur d'une unité du sous-jacent multiplié par le cours du contrat à terme. Exemple : le contrat à terme portant sur l'indice boursier S&P 500, E-Mini s'échange à 2,840$ (US). Chaque point est d'une valeur de 50$ (US). Donc, détenir un contrat E-mini est équivalent à contrôler une valeur de 2840 x 50 = 142,000$ (US).

- **les horaires de négociation des contrats à terme** : du lundi au vendredi De : 18:00. À : 16: 00.

- **le dépôt de garantie ou marge de dépôt** : C'est le montant minimum requis sur un compte pour pouvoir négocier un contrat à terme déterminé. Ce dépôt peut évoluer en fonction du risque de la volatilité des marchés.

- **la marge de maintenance** : Il s'agit du montant nécessaire pour qu'une opération sur marge reste ouverte. Le montant disponible sur le compte de négociation doit toujours être supérieur ou égale à la marge de maintenance. Dans le cas d'une position perdante, si la valeur du compte baisse en bas de cette marge, l'investisseur reçois un appel de marge. C'est-à-dire, qu'il doit déposer plus de fond pour couvrir le montant en bas de la marge. Si non, les positions seront liquidées, en partie ou en totalité, par le courtier.

  En général cette marge représente entre 5% à 10% de la valeur du contrat. Nous utiliserons cette information dans nos tests rétroactifs pour s'assurer que notre marge de maintenance est toujours respectée.

l'effet de levier : Les contrats à terme disposent d'un effet de levier important car les dépôts de garantie requis pour les négocier sont relativement faibles comparées à leur valeur de rachat.

Figure 5.0 Cours du contrat S&P 500 E-mini expirant en juin 2019
Tirée de barchart.com

Exemple : un investisseur achète en date du 08 Mars 2018 un contrat E-mini S&P 500 (ES) au prix de 2,730 $, avec une marge de 6,000$. La valeur de ce contrat est donc de 50$ x la valeur de l'indice S&P 500. Soit 136,500. Pour chaque point, quand le prix bouge, on gagne ou on perd 50$. Si l'investisseur décide de détenir le contrat jusqu'au 17 Mars, où le prix du sous-jacent touche la barre du 2,830$, il réalisera un profit de 50$ x100 = 5000$. Dans cet exemple, on voit que l'investisseur avait besoin de seulement 6000$ pour générer un profit de 5000$, soit un retour sur investissement de 83%. Ceci n'est qu'un exemple, mais en réalité, il est beaucoup plus difficile de réaliser ce genre de profit.

5.1.3 Description des données utilisées

Les contrats à terme expirent à une date prédéterminée. Et, contrairement à d'autres actifs financiers où le nom de l'actif ne change pas, le nom du contrat à terme comprend le symbole du sous-jacent, le mois d'expiration et les deux derniers numéros de l'année d'expiration.

Exemple : le nom d'un contrat à terme sur l'or qui expire en février 2018, sera [Symbole = « GC » & Code du mois = « G » & Code pour l'année d'expiration « 9 » en 2019] = **GCG9**.

Nous avons créé une base de données de 31 tables qui contiennent ces données. Chaque tableau est nommé selon la nomenclature décrite en haut, et comprend les informations suivantes :

- le temps où la transaction a eu lieu (chaque minute),
- le prix de fermeture (Close) à la minute,
- le prix d'ouverture (Open) à la minute,
- les prix bas (Low) et haut (High) à la minute,
- le volume des transactions à la minute.

Nous avons limité la liste d'actifs à tester aux contrats dans le tableau en bas. Ceci est dû à la qualité des données collectées, le temps nécessaire pour pouvoir tester tous les contrats et le nombre de scénarios à expérimenter.

Tableau 5.0 Liste des actifs utilisés

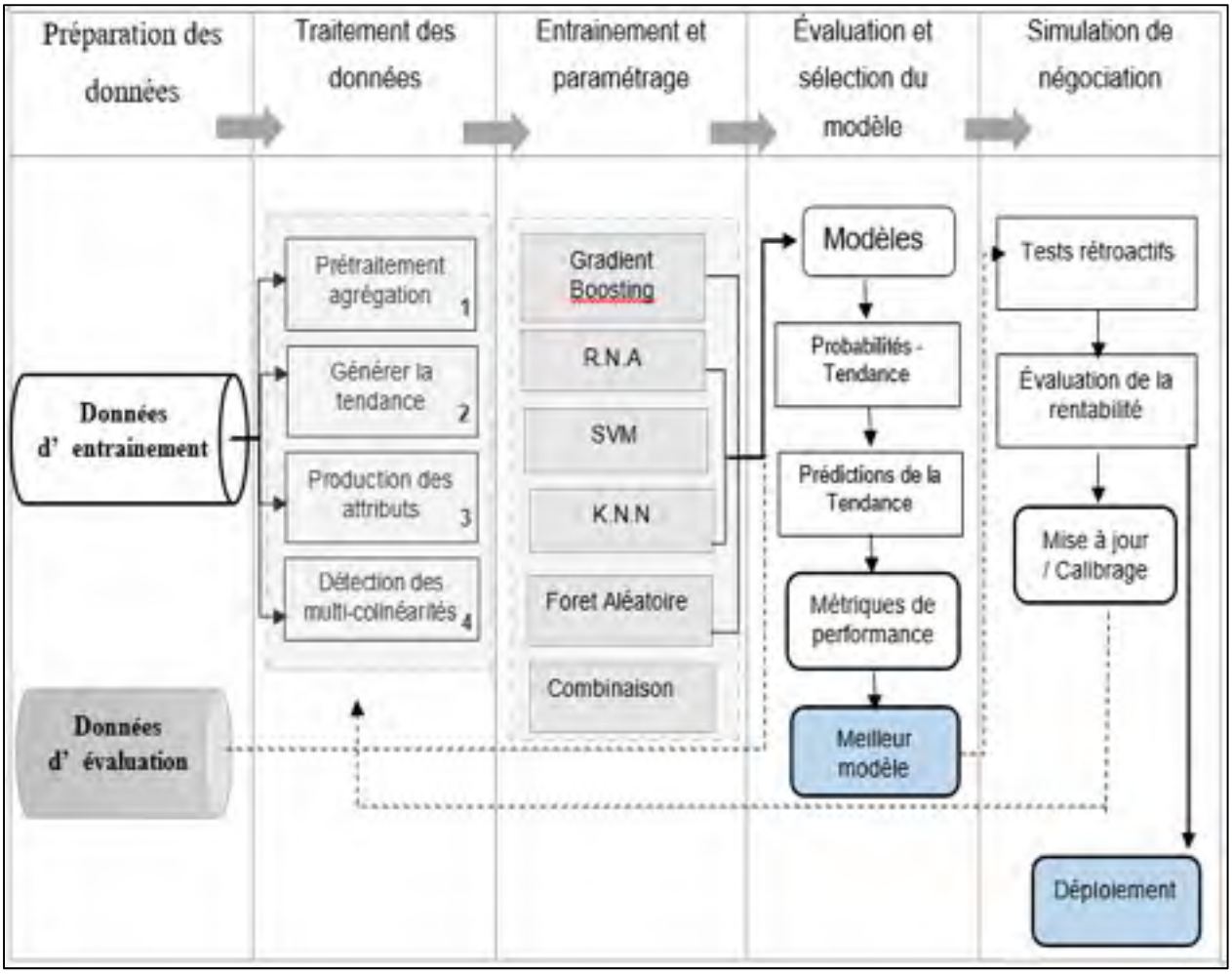
Actif	**Description**	**Période couverte**	**Remarques**
Futures E-mini S&P (ES)	Contrat à terme sur sur l'indice boursier Standard & Poor's 500	De : Déc. 2016 À : Déc. 2017	Données séparées dans plusieurs tables dépendamment de la date d'expiration
Crude Light (CL)	Contrat à terme sur le pétrole	De : Nov. 2016 À : Fév. 2018	Données séparées dans plusieurs tables dépendamment de la date d'expiration
Gold Futures (GC)	Contrat à terme sur l'or	De : Nov. 2016 À : Fév. 2018	Données séparées dans plusieurs tables dépendamment de la date d'expiration

Il est à noter qu'il est possible de fusionner les données d'un même sous-jacent avec des dates d'expiration différentes. Ceci aura comme effet, l'élargissement de la série des données.
Pour ce travail nous n'avons pas été en mesure de faire ça, car les données proviennent de plusieurs périodes et ne représentent pas une suite temporelle continue. En plus, on n'est pas certain si la date d'expiration du contrat impacte la volatilité du prix. Une chose qui nécessite plus de recherches et de tests pour invalider.

5.2 Méthodologie

5.2.1 Processus de modélisation

Le diagramme dans la page suivante décrit les étapes de modélisation de ce projet.



5.2.2 Préparation des données

Dans cette étape, on a utilisé les données collectées et placées dans notre base de données pour créer des échantillons. On a aussi nettoyé les données et on les a validés pour s'assurer qu'il n'y a pas de données manquantes ou de valeurs aberrantes.

On a aussi normalisé la structure des fichiers contenant ces données pour qu'ils soient tous structurés de la même façon et ayant les mêmes noms de champs.

5.2.3 Traitement des données

5.2.3.1 Création de la variable cible et agrégation des données

Cette étape comprend :

- **l'agrégation des données** : où on passe des données d'entrée disponibles (à la minute), à un niveau déterminé par l'utilisateur tel qu'expliqué dans la section (2.2.2.2 Agrégation et prétraitement des données). L'agrégation sera utilisée pour comparer la performance prédictive de nos modèles en variant la granularité de nos données.

- **calcul de la tendance (étiquetage des données)** : nous faisons le calcul pour les données d'entrainement et les données d'évaluation, tel que mentionnée à la section 2.2.3.2. La tendance est calculée en fonction d'un seuil de profitabilité prédéterminé (r) et pour un intervalle de temps prédéterminé.

$$Tt[i] = \begin{cases} -1, si : & Rt[i] \leq -r \\ 0, si : & -r < Rt[i] < r \\ 1, si : & Rt[i] \geq r \end{cases}$$

1. Pour chaque point de la série des prix de fermeture, on calcule la différence de prix entre ce point (t) et chaque point subséquent (t+i) :

   $$Rt[i] = \text{Pt} + \text{i} - \text{Pt}$$

   pour i allant de 1 à k. Où k est la fenêtre temporelle d'exposition.

2. On compare $Rt[i]$ à notre seuil de profitabilité (r). S'il est égal ou supérieur à cette cible, alors la tendance est haussière. S'il est négatif et inférieure à (–r), la tendance est baissière. Si non, elle est neutre.

La figure en bas montre ces deux points.

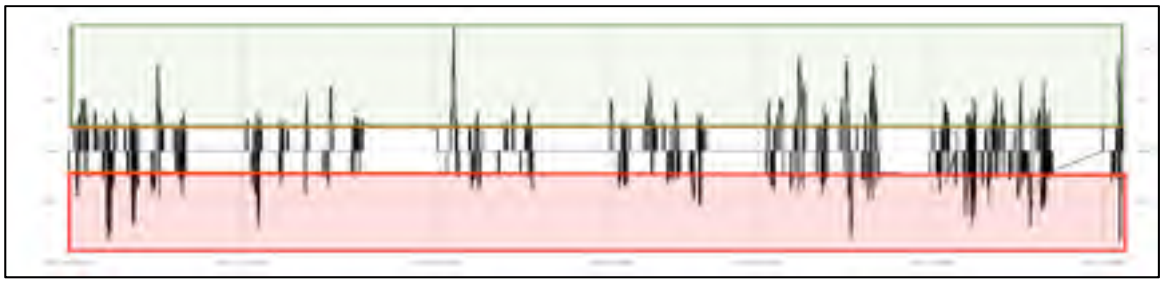


Figure 5.1 Calcul du rendement sur une période d'exposition de 60 minutes

Exemple : pour des données du contrat de pétrole de 10 minutes. La tendance est présentée dans le tableau ci-dessous, Où nous avons utilisé des barres de 10 minutes, un temps d'exposition de 10 barres et un seuil de rendement = 0.15 point (150$)

Tableau 5.1 Calcul des rendements pour générer la tendance

temps (t)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
62	0.06	0.10	0.10	0.09	0.22	0.25	0.30	0.26	0.29	0.20
63	0.04	0.04	0.03	0.16	0.19	0.24	0.20	0.23	0.14	0.10
64	0.00	-0.01	0.12	0.15	0.20	0.16	0.19	0.10	0.06	0.05
65	-0.01	0.12	0.15	0.20	0.16	0.19	0.10	0.06	0.05	0.15
66	0.13	0.16	0.21	0.17	0.20	0.11	0.07	0.06	0.16	0.10
67	0.03	0.08	0.04	0.07	-0.02	-0.06	-0.07	0.03	-0.03	-0.02
68	0.05	0.01	0.04	-0.05	-0.09	-0.10	0.00	-0.06	-0.05	-0.12
69	-0.04	-0.01	-0.01	-0.14	-0.15	-0.05	-0.11	-0.10	-0.17	-0.11
70	0.03	-0.06	-0.10	-0.11	-0.01	-0.07	-0.06	-0.13	-0.07	-0.02
71	-0.09	-0.13	-0.14	-0.04	-0.10	-0.09	-0.16	-0.10	-0.05	-0.04
72	-0.04	-0.05	0.05	-0.01	0.00	-0.07	-0.01	0.04	0.05	0.05
73	-0.01	0.09	0.03	0.04	-0.03	0.03	0.08	0.09	0.09	0.43

Pour cet exemple, on remarque l'existence de plusieurs séquences où la tendance était haussière.

Dans le cas où nous serons intéressés par ce genre de séquences, il est possible de modifier la définition de la tendance pour aller les chercher. Le but sera dans ce cas d'améliorer le signal de la prédiction. En d'autres mots, on ne veut pas entrer dans des positions que si nous pensons que le prix sera supérieur à un certain seuil de profitabilité (r), *x* fois subséquentes durant la période d'exposition. En faisant ça, on risque de réduire le nombre de fois où la tendance est haussière ou baissière. La classe neutre deviendra majoritaire et les autres classes deviennent rares.

Le graphique en bas illustre l'idée derrière le calcul et la génération de la tendance pour le 1er vecteur du tableau en haut (t=62). La couleur verte de la variation du prix représente la tendance haussière.

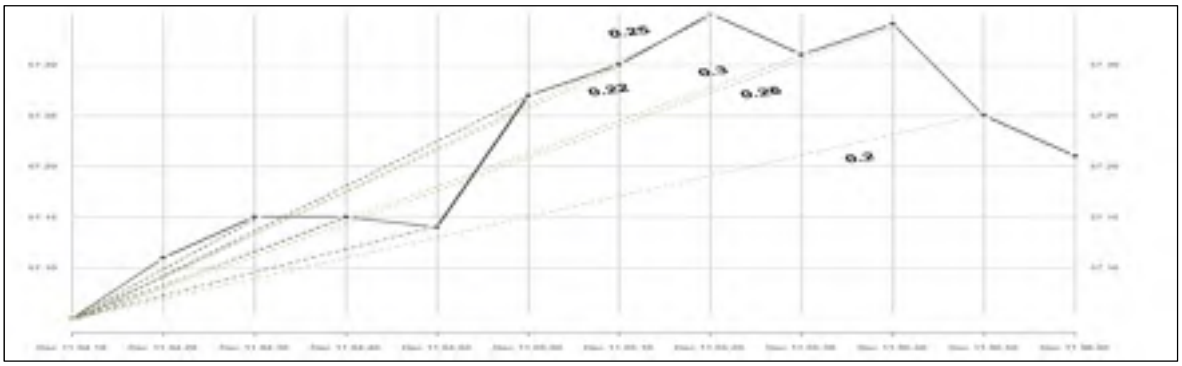


Figure 5.2 Description de l'étiquetage des données pour la tendance

Toujours pour le contrat de pétrole, nous avons généré la tendance tout en variant la fréquence d'agrégation des données (1 minute, 5 minutes et 10 minutes). Le but est de voir comment les distributions changent et voir ultérieurement, s'il existe une corrélation entre l'agrégation des données et la performance de nos modèles. Le seuil de profitabilité utilisé est r = 0.15 points (150$) et le temps d'exposition est k = 60 minutes.

Tableau 5.2 Distribution de la tendance pour le contrat à terme CLG8

Tendance / plage de temps utilisée	**1 minute**		**5 minutes**		**10 minutes**	
Baissière	6,218	16%	1117	14%	529	13%
Neutre	25,269	64%	5322	67%	2748	69%
Haussière	8,173	21%	1493	19%	689	18%
Total	39,660	100%	7,932	100%	3966	100%

On voit comment la distribution de la tendance calculée est non équilibrée. Dans plus de 64% des instances, le prix du contrat ne bouge pas assez pour que la différence dépasse le seuil de profitabilité.

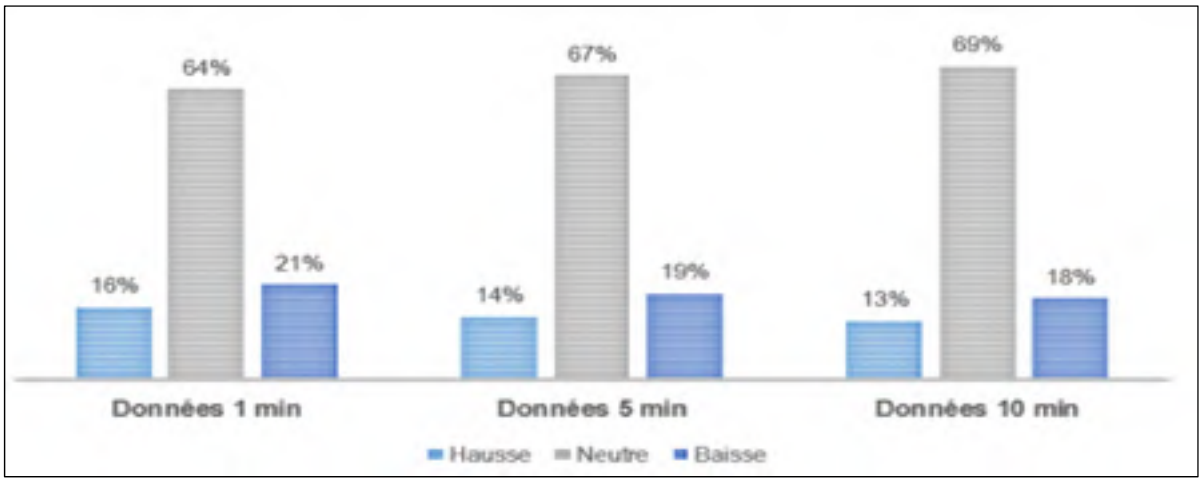


Figure 5.3 Distribution de la tendance pour le contrat CLG8

On voit aussi sur le graphe en haut comment la classe ‘neutre’ est majoritaire et comment l’agrégation des données a généré plus de déséquilibre au niveau de la distribution de nos classes. Plus qu’on augmente la fréquence d’agrégation, plus que la classe ‘neutre’ prend plus de poids. Le contraire est vrai pour les deux autres classes. Ce problème provient du processus du calcul de la tendance où on commence par agréger les données. Ensuite, on calcule la tendance à partir des données agrégées.

Pour corriger ce problème on propose de changer le processus, en calculant la tendance pour les données non agrégées à la place, et en considérant la tendance comme un élément dans le processus d'agrégation. Vu que la tendance est calculée en fonction du prix de fermeture, la nouvelle agrégation devra utiliser la même formule. La formule d'agrégation de la tendance à l'instant (*i)* devient :

$$T^{K}_{(i)} = f(T, K, i) = T_{(i*k)} \tag{5.2}$$

Où T représente la tendance des données granulaires et K la fréquence d'agrégation.

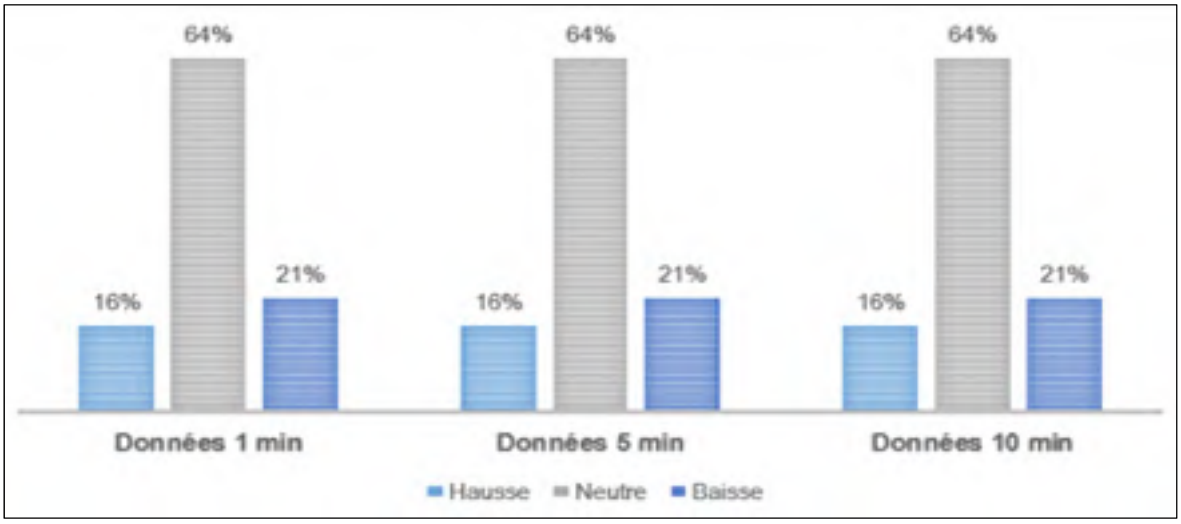


Figure 5.4 Distribution de la tendance pour le contrat CLG8 en utilisant la nouvelle définition

Sur le graphe en haut, on voit comment la nouvelle formule a éliminé le biais d'agrégation. En effet les distributions sont maintenant les mêmes quand on varie la fréquence d'agrégation. Toutefois, les classes entre elles restent non équilibrées.

Entrainer les modèles d'apprentissage automatique avec des données non équilibrées va mener à générer des prédictions biaisées en faveur de la classe majoritaire. Nous proposons en bas trois techniques pour faire face à ce problème.

a) Balancer les données d'entrainement pour avoir des données où la distribution est assez similaire pour les trois classes

L'idée ici consiste à refaire l'échantillonnage pour que la distribution des trois classes soit pratiquement semblable. Ceci peut se faire en :

- réduisant le nombre d'instances de la classe majoritaire (sous-échantillonnage),
- augmentant le nombre d'instances de la classe minoritaire (sur-échantillonnage) (Khun, 2019),
- générer artificiellement de nouveaux exemples synthétiques (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

Nous avons utilisé la dernière option où nous avons fait un sur-échantillonnage synthétique de nos deux classes minoritaires. Chaque classe est traitée à la fois au même compte que la somme du compte de toutes les autres classes. Il en résulte un léger déséquilibre dans le nombre total d'étiquettes, comme on le voit sur les deux figures en bas.

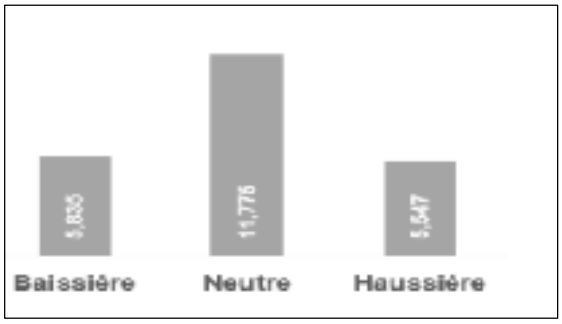


Figure 5.6 Distribution des données originales (non équilibrées)

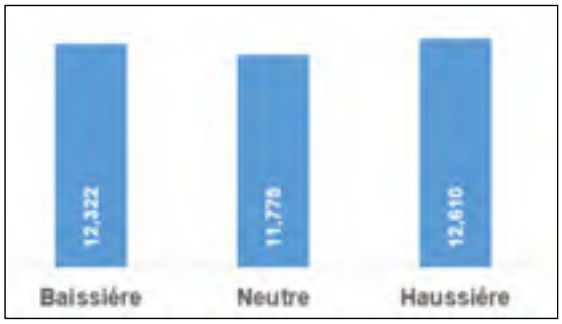


Figure 5.7 Distribution des données équilibrées

b) Modifier les seuils de probabilités utilisés par les modèles pour générer la classification

Par défaut, les modèles génèrent une classification en assumant l'équiprobabilité des classes. Donc, par défaut chaque classe a un seuil qui est égale à (1/m) où m est le nombre de classes.

Nos modèles calculent la probabilité suivante : $p(classe = i \mid X = x1, x2 \dots xn)$. Ensuite, la classe prédite sera celle où cette probabilité est supérieure ou égale à notre seuil de probabilité.

Pour notre cas, où nous avons trois classes, nous avons par défaut, les seuils de probabilité suivants : (Hausse = 1/3, Baisse = 1/3, neutre = 1/3). La classe générée par notre modèle pour une observation est celle qui a la probabilité maximale et qui est supérieure au seuil prédéfini. Exemple : si nous supposons que le modèle M a généré les probabilités suivantes {0.4, 0.35,0.25} pour les classes {hausse, neutre, baisse} pour l'observation « *i* ». Si nous avons des seuils égaux. Alors, la classe générée sera 'Hausse'.

Maintenant, si on suppose que nous avons les seuils suivants : {0.5, 0.3, 0.2} à la place des seuils par défaut. Dans ce cas, la classe générée sera 'neutre', car c'est la seule classe pour laquelle la probabilité générée par le modèle dépasse le seuil prédéfini. Même si la probabilité la plus élevée est celle de la classe 'hausse', cette dernière ne dépasse pas le seuil prédéfini pour cette classe.

Notre intervention à ce stade dans ce projet consiste à calibrer les seuils de probabilités, en discriminant la classe majoritaire au détriment des autres classes. Autrement dit, au lieu d'utiliser un seuil égal pour toutes les classes, on va l'augmenter pour la classe majoritaire et le baisser pour les autres classes. La somme de ces seuils doit être égale à 1. Pour ce travail, nous utiliserons les fréquences empiriques et la courbe ROC pour déterminer ces seuils (Perkins & Schisterman, 2006). En bas nous avons un exemple illustrant comment on peut changer ces seuils.

Tableau 5.3 Seuils de probabilités par défaut et seuils basé sur la fréquence de la distribution des classes

	Hausse	**Neutre**	**Basse**
Seuil par défaut = 1/3	1/3	1/3	1/3
Nouveau seuil	0.27	0.5	0.23

c) Utiliser des facteurs de coûts pour pénaliser la classe majoritaire et améliorer les sorties des modèles

Ici on cherche à calculer un coût global qui est la somme, des coûts unitaires des mauvaises classifications par classe multiplié par le nombre des mauvaises classifications.

$$coût\ total = \sum_{i=1}^{n} \#\left(Y_i \neq Y_i^{Pred}\right) * \ CU(Y_i \neq Y_i^{Pred}) \qquad (5.3)$$

Où :

Y est la variable cible (la tendance),

Y^{pred} est la valeur prédite de Y,

CU est le coût unitaire d'une mauvaise classification pour une classe spécifique.

Une façon simple pour améliorer la qualité des prédictions pour des classes non équilibrées, consiste à attribuer un coût plus élevé pour les mauvaises classifications de la classe majoritaire.

		Classes actuels		
		Hausse	**Neutre**	**Basse**
Classes prédites	**Hausse**	0	C1	C2
	Neutre	**C3**	0	**C4**
	Basse	C5	C6	0

Le modèle est donc sensible aux coûts des mauvaises classifications et tendra à calibrer ses classifications générées en minimisant le coût total.

Nous avons testé cette approche avec l'algorithme des arbres de décision C5 et nous avons constaté une légère amélioration des résultats.

5.2.4 Production des attributs

5.2.4.1 Sélection des variables dépendantes

Dans cette étape nous avons commencé par générer des attributs qui peuvent être utilisés pour la prédiction de la tendance. Ces attributs peuvent être des données fondamentales, des indicateurs techniques, des nouvelles, des scores de sentiments ou autres attributs susceptibles d'améliorer le pouvoir prédictif de nos modèles.

Pour ce projet nous nous sommes limités aux indicateurs techniques comme variables explicatives. Au total, on a généré 22 indicateurs techniques. Parmi ces indicateurs on trouve 6 RSI avec des périodes différentes allant de 12 à 72 en augmentant à chaque fois de 12 unités de temps. La même chose a été faite avec la moyenne mobile exponentielle mais en augmentant le pas de 24 unités. Nous avons aussi calculé les variations des moyennes mobiles exponentielles et du prix de fermeture par rapport au VWAP.

Les autres indicateurs comme le MACD, les bandes de Bollinger sont directement calculés et intégrés dans nos tables de modélisation. Nous avons exclu toutes les données brutes (prix et volumes). La liste des indicateurs techniques avec leur définitions et formules se trouve dans l'annexe 1.

5.2.4.2 Analyse de multicolinéarité

En statistique, La multicolinéarité est un terme qui fait référence à l'utilisation du même type d'information plus d'une fois. En général, ce phénomène est présent quand les variables explicatives sont très corrélées entre elles. La présence de ce phénomène impacte l'inférence statistique faite à partir des données contenant de la multicolinéarité.

Il s'agit d'un problème commun dans l'analyse technique. C'est pourquoi il faut être prudent à ne pas utiliser d'indicateurs techniques qui reflètent la même information.

La matrice de corrélation en bas montre comment certains indicateurs techniques sont très corrélés entre eux, surtout les indicateurs de la même famille. Les moyennes mobiles exponentielles sont parfaitement corrélées entre elles. Les indicateurs RSI avec des périodes différentes montrent une corrélation trop élevée. Le MACD aussi est très corrélé avec la variation relative des deux moyennes mobiles exponentielles. D'ailleurs, ce dernier n'est qu'une différence entre deux moyennes exponentielles. Donc, le fait d'utiliser des indicateurs dérivés à partir de la même série de données pour confirmer chacun l'autre, est une mauvaise approche.

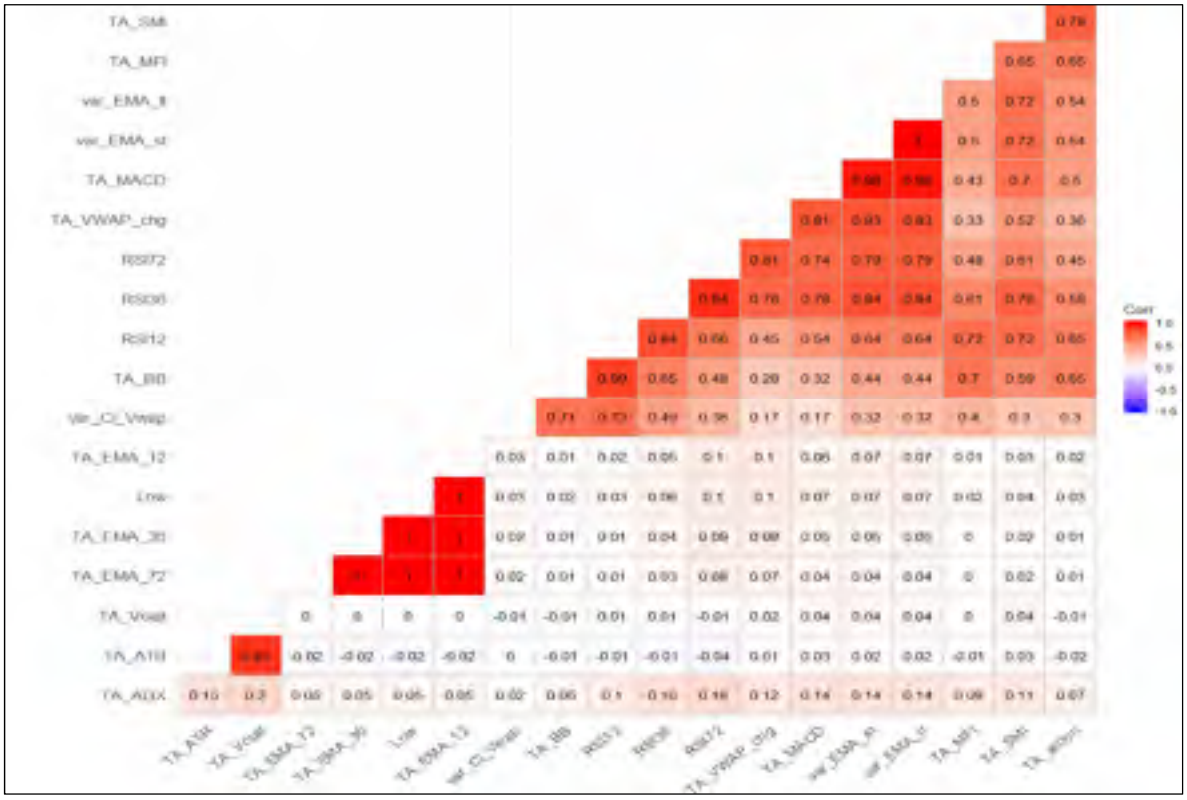


Figure 5.5 Matrice de corrélation des indicateurs techniques pour le contrat à terme E-mini S&P

Un grand danger de cette approche est le surapprentissage. C'est-à-dire que les résultats obtenus en utilisant des données d'entrainement vont être très bons mais ça risque de produire des résultats moins bons sur l'ensemble d'évaluation ou lors du déploiement du modèle en production.

Le graphe dans la figure 5.7 montre comment des RSI de période 12, 36 et 72 affichent la même information. Dans ce cas, le fait d'utiliser ces trois indicateurs en même temps, va juste ajouter du bruit à nos modèles. Un seul indicateur parmi les trois, celui qui donne le plus d'information, devra être utilisé.

Nous verrons dans ce qui suit comment on calcule la valeur de l'information.

Figure 5.6 Comparaison entre des indicateurs RSI avec des paramètres différents

Pour faire face à ce problème, on cherchera à réduire le nombre d'attributs utilisés par le modèle et limiter les interactions entre ces variables explicatives.

On procède de la manière suivante :

1. On classe nos indicateurs dans des catégories :
   - indicateurs de tendance : AROON, MACD, les moyennes mobiles, etc.
   - indicateurs de volatilité : ATR, les bandes de Bollinger, l'indice de volatilité relative, etc.
   - indicateurs Momentum : RSI, l'indice Momentum stochastique, taux de changement du prix, etc.
   - indicateurs de la force du marché : volume, volume d'équilibre, Prix moyen pondéré en volume.

2. On calcule les facteurs d'inflation de la variance (Bollinger, 1981)[5] pour chaque indicateur dans le but de quantifier la sévérité de la multicolinéarité. Ceux avec des valeurs élevées sont retirée de la liste des variables explicatives. La définition de « élevé » est quelque peu arbitraire, mais des valeurs comprises entre 5 et 10 sont couramment utilisées.

3. Si deux ou plusieurs indicateurs sont corrélés, on utilise la valeur de l'information pour sélectionner celui qui a la plus grande valeur et on exclue les autres. Exemple : Entre le VWAP36 et le VWAP12 qui sont parfaitement corrélés, on ne sait pas lequel choisir. Ceci est vrai aussi pour la série des RSI (12, 24, 36, 60) et le ATR avec l'index de volatilité qui montrent une corrélation trop élevée. Si on prend le cas des VWAP, seul le VWAP36 sera sélectionné. Entre le ATR et la volatilité, le ATR sera utilisé. La même chose pour le RSI60. Ceci s'explique par le fait que ces indicateurs ont la valeur de l'information (l'importance de l'attribut) la plus élevée. Ceci sera discuté dans la prochaine section.

---

[5] Les facteurs d'inflation de la variance mesurent combien la variance d'un coefficient de régression fluctue si les variables explicatives sont corrélées.

Une alternative à la méthode décrite précédemment consiste à utiliser l'analyse par composante principale (ACP), une procédure statistique qui utilise une transformation linéaire pour convertir des variables corrélées en un ensemble de variables linéairement indépendantes appelées composantes principales. Le nombre de composantes générées est inférieur ou égal au nombre de variables d'origine(Roweis & Saul, 2000).

5.2.4.3 Mesure de l'importance des variables explicatives

L'importance d'une variable se mesure en fonction de la contribution de cette dernière dans l'amélioration de la performance du modèle.

Il existe plusieurs techniques pour faire la sélection des variables les plus importantes et enlever celles qui ne contribuent pas ou peuvent même diminuer la performance du modèle. Les deux catégories d'algorithmes pour la sélection des attributs que nous avons explorés sont les suivantes :

- **les méthodes de filtrage** : Ces méthodes utilisent des mesures statistiques pour assigner un score à chaque attribut.

  On peut par la suite, ordonner nos attributs en utilisant ce score et déterminer un seuil de rejet pour exclure les variables avec un faible score. On peut citer comme exemple, l'entropie, la valeur de l'information et le test de Chi carré.

- **les méthodes de recherche itérative « Wrapper »**: ces méthodes utilisent des sous-ensembles de variables, à chaque fois, jusqu'à trouver l'ensemble qui donne la meilleure performance. Exemple : GBM, foret aléatoire.

  Le graphe dans la figure 5.8 montre l'importance des indicateurs pour la prédiction de la tendance en utilisant un modèle GBM. La mesure de l'importance se fait par élimination récursive. C'est-à dire, à chaque fois on enlève une variable de notre ensemble de variables explicatives et on mesure la performance, jusqu'à obtention de la meilleure performance en utilisant toutes les combinaisons possibles.

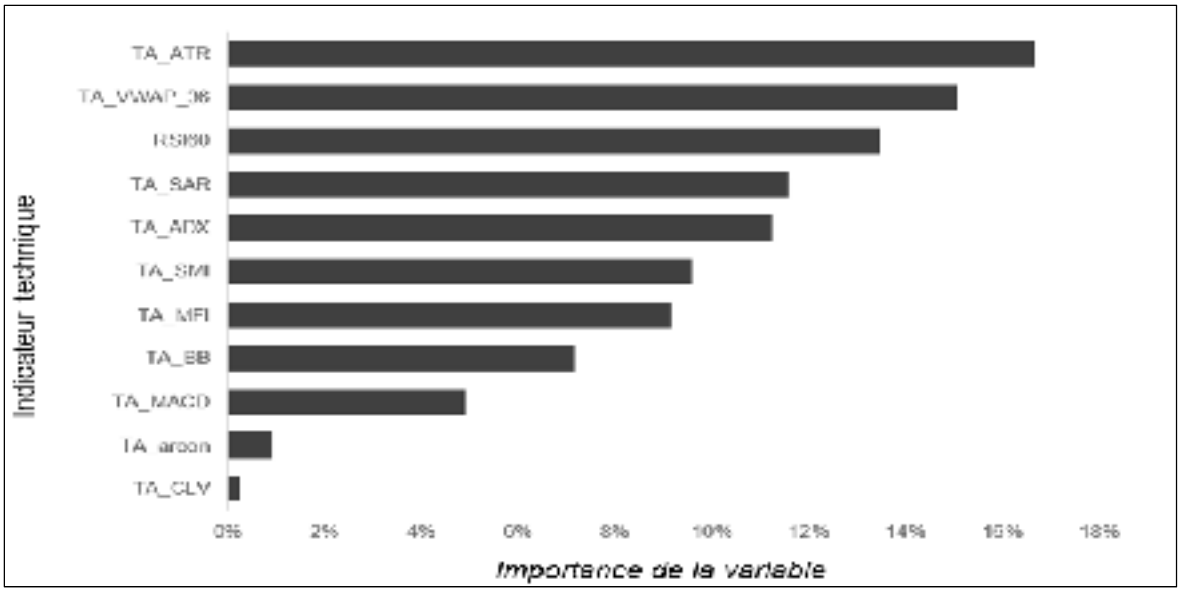


Figure 5.7 Valeurs de l'information pour les indicateurs techniques du contrat S&P E-mini

La figure en haut montre l'importance des indicateurs les plus importants. Huit indicateurs ont été suffisants pour atteindre la performance maximale. Une illustration des distributions des meilleurs indicateurs par classe de la tendance est en bas.

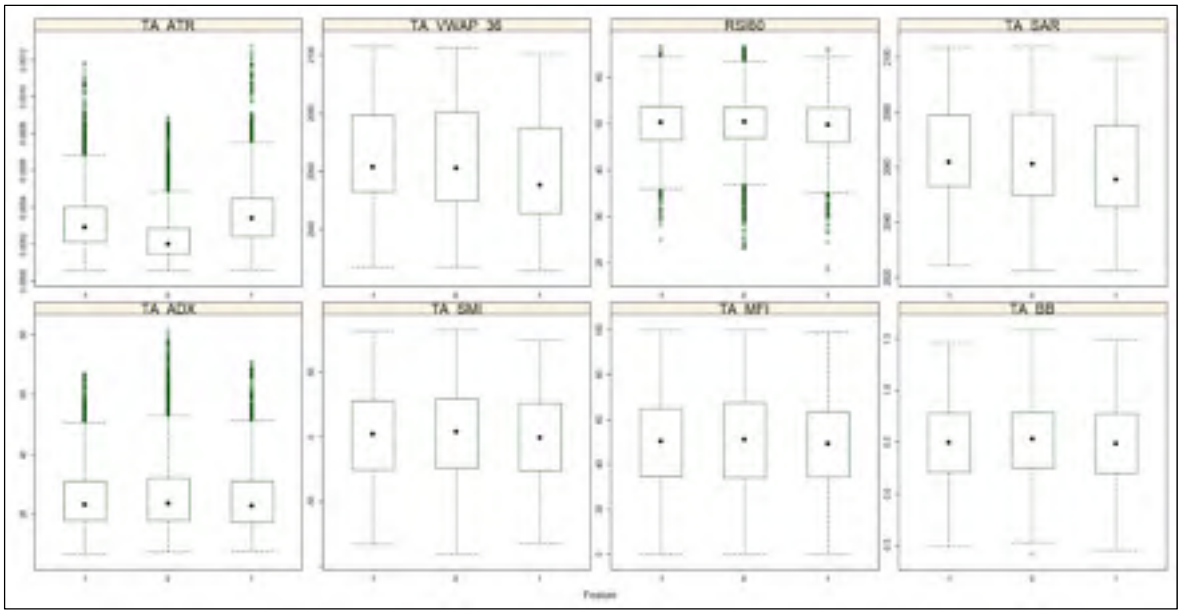


Figure 5.8 Distribution des indicateurs les plus importants par rapport à la tendance

La figure en haut montre comment certaines corrélations peuvent être observées entre la tendance et les indicateurs techniques.

On peut voir que quand le ATR est moins élevé, la tendance a plus de chance d'être neutre et le contraire aussi est vrai. Il est important de mentionner ici que nos indicateurs techniques, y compris le ATR, sont calculées à partir des instances passées et non pas celles qui sont utilisées pour calculer la tendance. On peut alors dire qu'il y a probablement des autocorrélations dans les séries des prix de nos données, car pour les instances où la tendance était neutre, le ATR a bougé moins comparer aux autres cas où la tendance était différente. Ce qui veut dire que quand le prix du titre bouge, il a plus de chance de continuer dans la même direction.

Pour le VWAP, on voit que nous avons plus tendance à voir un mouvement haussier quand cet indicateur baisse. Pour le RSI 60 minutes, on voit plus une tendance haussière ou neutre quand il est en bas de 30. On voit aussi une légère différence entre la moyenne du RSI par type de tendance, mais elle n'est pas significative (RSI_{moy} = 49.28 pour C = 1, RSI_{moy} = 50.34 pour C=0 et RSI_{moy} = 50 pour C = -1) où 1, 0 et -1 représentent la classe de la tendance haussière, neutre et baissière.

Sur la figure 5.9 nous avons tracé les fonctions de densité de chaque indicateur par rapport à la tendance calculée. On a donc la densité conditionnelle de l'indicateur technique (IT) étant donnée la classe de la tendance : $f(IT(i)|C = j)$ où $j = -1, 0, 1$.

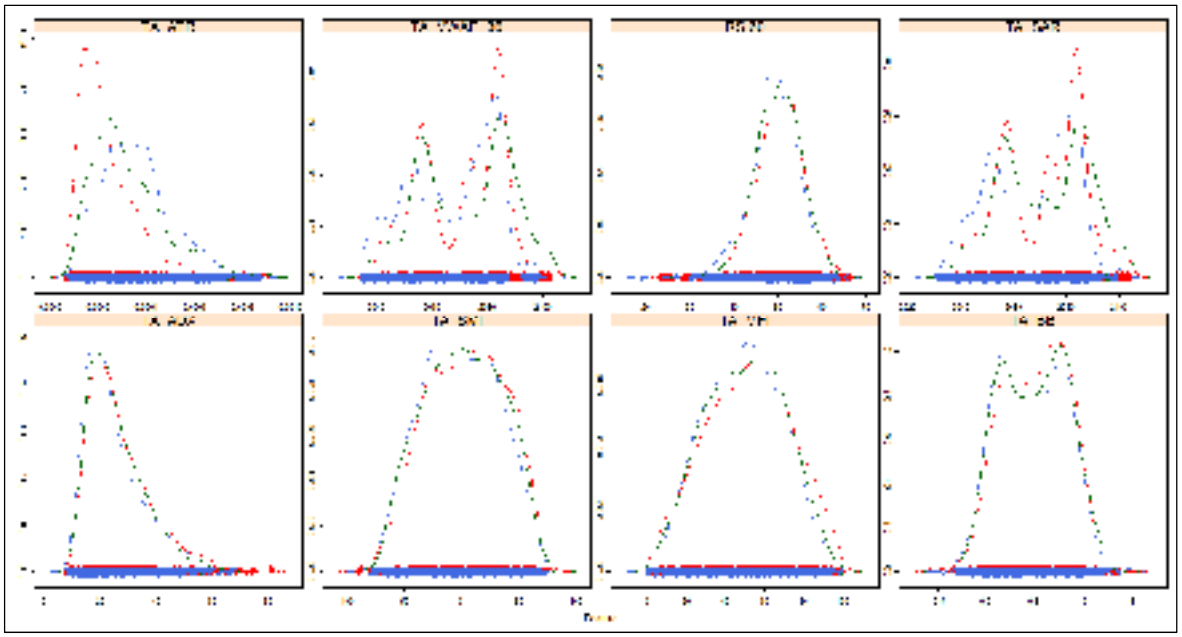


Figure 5.9 Densités des indicateurs techniques les plus importants par rapport à la tendance

Plus que les densités conditionnelles d'un même indicateur sont similaires et colle l'une sur l'autre, plus que l'indicateur n'est pas important pour prédire la tendance. Par contre, une divergence, tel que le montre le ATR et le VWAP_36, est un signe que l'indicateur est un facteur discriminant de la tendance et peut avoir un pouvoir prédictif élevé.

5.2.5 Entrainement et paramétrage des modèles

Dans la première partie de cette section on va expliquer comment, à travers les techniques d'échantillonnage, on peut générer des ensembles de données pour être utilisés dans l'entrainement de nos modèles et d'autres ensembles indépendants qui vont être utilisés seulement pour l'évaluation de la qualité des prédictions générées par nos modèles prédictifs. Dans la deuxième partie, on discutera les étapes qui ont été nécessaires pour le choix et l'optimisation des paramètres de nos modèles.

5.2.5.1 Sélection des ensembles de données d'entrainement et d'évaluation

Cette étape consiste à créer des échantillons pour entrainer, valider et évaluer la performance de nos modèles. Il est nécessaire de prendre en considération la dépendance temporelle des prix dans notre échantillonnage, étant donné qu'on travaille avec des séries temporelles. Pour ce faire nous allons diviser nos données en trois ensembles :

- **ensemble d'apprentissage** : représente les 60% premières instances de l'ensemble des données pour chaque contrat. Elles sont utilisées pour entrainer les modèles.

- **ensemble de validation** : représente les 20% suivantes des données. Il est utilisé pour optimiser le choix des paramètres à utiliser.

- **ensemble d'évaluation** : sous-ensemble destiné à l'évaluation de la performance prédictive des modèles et la rentabilité du système. Il est constitué des 20% des données restantes.

Figure 5.10 Exemple d'échantillonnage temporel simple du contrat S&P E-mini

Sur le graphe en haut, on voit comment l'ensemble d'entrainement est différents des deux autres ensembles. Il est alors possible qu'un biais relié aux données soit introduit et nos modèles peuvent mal généraliser la connaissance apprise à partir des données d'entrainement.

Une autre approche plus robuste consiste à déplacer l'ensemble d'entrainement au fur et à mesure qu'on avance dans le temps. Ceci est expliqué dans le schéma en bas. L'apprentissage et la prédiction se font de la même façon.

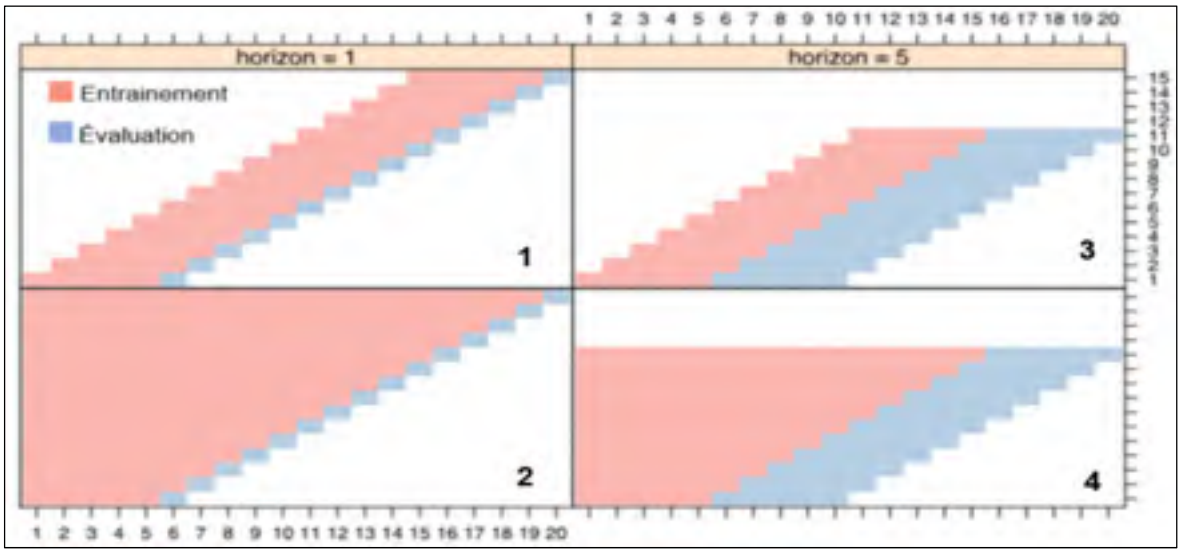


Figure 5.11 Échantillonnage temporel par roulement
Tirée de (Khun, 2019)

L'avantage de cette technique est de s'assurer que le modèle intègre la nouvelle connaissance et s'adapte aux nouveaux changements.

Sur la figure 5.12, on voit que pour les quatre scénarios, l'ensemble d'entrainement ainsi que l'ensemble d'évaluation ou de prédiction (dépendamment de la tâche) bouge simultanément. Au total nous avons 20 points de données dans ces exemples.

- pour le scénario 1, nous voulons prédire la classe du prochain point en utilisant les cinq derniers points. À chaque fois, on prend les cinq derniers points pour entrainer notre modèle et faire la prédiction du point suivant. Ceci aura pour effet, la mise à jour de notre modèle avec les données récentes. Le scénario 1 est similaire au scénario 3. La seule différence est l'horizon.

- Pour le scénario1, on prédit une instance à chaque fois, tandis que pour le scénario 3, la taille de l'ensemble d'entrainement est la même que celle de l'ensemble de prédiction (égale à 5 pour l'exemple affiché).

- pour le scénario 2, on ajoute à chaque fois les nouvelles données à notre ensemble d'entrainement tout en gardant l'historique des données passées. Ce qui résulte de ça, une mise à jour et une augmentation de la taille des données d'entrainement. Mais contrairement au scénario 1, les données récentes auront moins de poids. L'un des grands problèmes de cette approche est le temps de calcul et la mémoire qui doit être allouée pour exécuter l'ensemble des itérations. Nous avons essayé de tester nos modèles en utilisant cette approche (scénario 3 et 4).

Nous avons utilisé un ensemble de données de 23,880 points. 2388 ont été utilisé comme données d'entrainement de départ et un horizon de 2388 a été fixé. À chaque fois, le modèle utilisé fera la prédiction pour le point suivant et il compare le résultat obtenu avec les données réelles, et ainsi de suite. Ceci fera au total 21492 itérations. Si on suppose que ça prend 10 secondes pour changer l'échantillon, rouler notre modèle et générer une prédiction. Avec un calcul simple, on peut trouver que le temps que ça prendra pour générer les prédictions pour la série des points à prédire, est de 60 heures.

Nous avons dû abandonner et refaire l'expérience avec moins d'itérations. La technique a montré de bons résultats. Toutefois, on n'a pas pu la tester sur un échantillon de données de grande taille. Mais si elle est déployée en production avec un modèle pré-entrainé, le temps d'exécution sera négligeable.

5.2.5.2 Entrainement et paramétrage des modèles

Nous entrainons nos modèles en utilisant un ensemble de données que nous avons préparé. Le choix des paramètres est fait sur plusieurs sous-ensembles de données de validation que nous générons en utilisant la validation croisée.

Voici les étapes suivies pour l'optimisation des paramètres de nos modèles :

On commence par définir l'ensemble des paramètres à optimiser pour un modèle M_i

Pour chaque paramètre δi :

1. Pour chaque itération, on génère des échantillons de données.
2. On se sert d'une partie de ces échantillons pour l'apprentissage.
3. On fait de la prédiction avec les échantillons non utilisés (validation).
4. On calcule la moyenne de la performance avec les données de validation.
5. On détermine le paramètre optimal. Celui qui donne la meilleure performance.
6. On utilise le paramètre optimal sur l'ensemble des données d'entrainement et d'évaluation.

Nous avons entrainé et testé sept classifieurs. dans le tableau 5.4, nous avons la liste des modèles testés avec les paramètres sélectionnés.

Tableau 5.4 Liste des modelés utilisés pour prédire la tendance des actifs financiers sélectionnés

Classifieur	**Description**
K plus proches voisins (KNN)	Le nombre de voisins les plus proches a été déterminé en faisant une recherche de grille 'Grid Search'. La valeur qui donne la meilleure performance était égale à 5.
Régression logistique	Nous avons utilisé une recherche de grille (grid search) pour trouver le paramètre C, l'inverse de la force de régularisation.
Arbres de décision C5	Nous avons choisi 10 comme nombre d'itérations de Boosting. Nous avons aussi utilisé une matrice des coûts dans une autre version pour des fins de comparaison
Forêt aléatoire	• nombre d'arbres dans la forêt : 200. Ce choix est basé sur la convergence de l'erreur. On a remarqué que l'erreur se stabilise et converge vers une valeur constante aux alentours de 200 arbres. • nombre de variables à utiliser pour partitionner les arbres : 6

Gradient Boosting Machine (GBM)	Nous avons trouvé que la combinaison des paramètres suivants qui donne le meilleur score de Kappa : • nombre d'arbres : 500 • profondeur d'interaction : 10 • taux de rétrécissement (Shrinkage) : 0.1 • nombre minimum d'observations par nœud : 10
Réseaux de neurones artificielles (PMC)	• nombre de couches : 10 • taux d'apprentissage : 0.05 • perte de poids (Weight decay): 0.1
Machines à support vectorielle (SVM)	• méthode : Linéaire • paramètre de pénalité (C) : 1 • le nombre maximum d'itérations : 1000

CHAPITRE 6

RÉSULTATS DE L'ANALYSE EXPÉRIMENTALE

Dans ce chapitre on va exposer les résultats obtenus en appliquant les différentes techniques et modèles décrits précédemment sur les trois contrats utilisés, en se basant sur des tests rétroactifs de validité (Backtesting). Il peut être répartie en deux grandes parties :

- la première consiste à tester le pouvoir prédictif de nos modèles en utilisant les métriques qui ont été discutés au chapitre IV. Nous commençons par définir un temps d'exposition et un seuil de rendement cible. Ensuite, on teste nos modèles sur nos ensembles de données et on compare nos résultats à une stratégie qui suppose que les changements dans la série des prix sont aléatoires.

- la deuxième partie sert à évaluer la rentabilité obtenue en utilisant les modèles sélectionnés dans la première partie. Ensuite, on compare la rentabilité obtenue à celle qui serait générée à une stratégie de gestion passive.

6.1 Évaluation de la performance statistique des modèles

6.1.1 Contrat à terme sur l'indice boursier Standard & Poor's 500

Le premier actif que nous utilisons pour nos évaluations est un contrat à terme sur l'indice boursier Standard & Poor's 500 (ESM6).

Tableau 6.1 Données et critères utilisés pour le contrat à terme ESM6

	Données d'apprentissage	**Données d'évaluation**
Période couverte	Du : 20 Avr. 2016 Au : 26 Fév. 2016	Du : 12 May. 2016 Au : 20 May. 2016
Granularité des données brutes	1 minute	1 minute
Périodes d'agrégation utilisées	5, 10, 15 minutes	5, 10, 15 minutes
Nombre de points (données brutes)	21030	8675
Valeur d'un point	50 $	50 $
Temps d'exposition	120 minutes	120 minutes
Seuil de profitabilité minimum	4 points	4 points

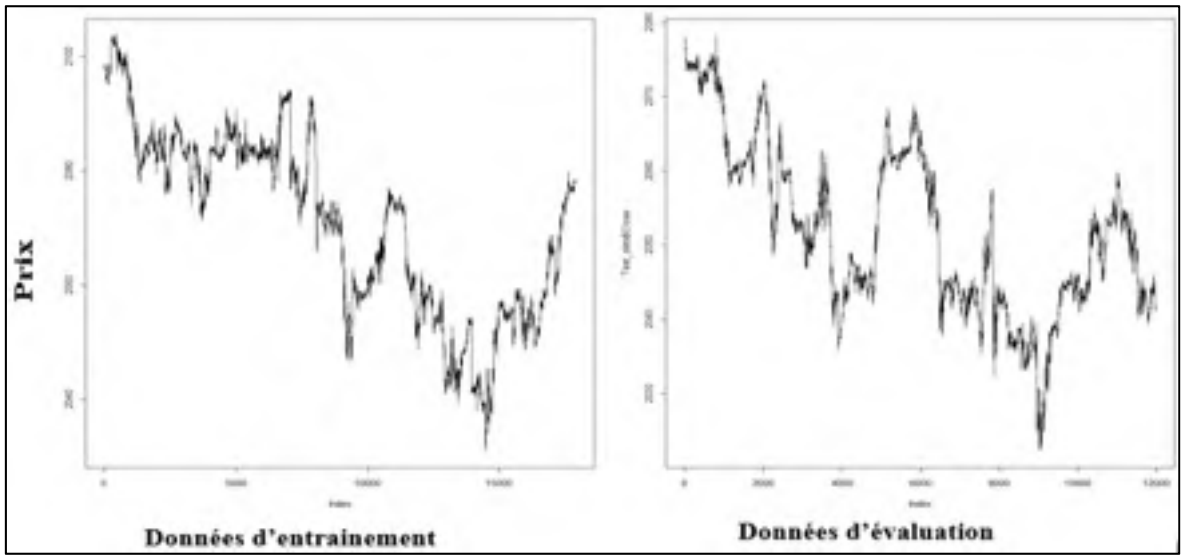


Figure 6.1 Comparaison entre les données d'entrainement et les données d'évaluation - Contrat : ESM6

Tel qu'expliqué au chapitre 5, nous avons proposé trois techniques pour faire face au problème de la distribution non équilibrée des classes de la tendance. Pour ce projet nous avons opté pour la technique de suréchantillonnage synthétique pour les données d'entrainement.

L'idée consiste à balancer les distributions de nos données d'entrainement tout en gardant les données d'évaluation inchangées. Pour tous les actifs utilisés dans ce chapitre nous avons utilisé la même technique.

6.1.1.1 Résultats de l'évaluation sur des données d'entrainement

Le tableau en bas montre les résultats de la performance statistique de nos modèles en utilisant la validation croisée sur l'ensemble d'entrainement.

Tableau 6.2 Résultats d'évaluation des modèles sur l'ensemble d'entrainement - Contrat : ESM6

Modèle	Min.	1er quartile	Médiane	Moyenne	3 éme quartile	Max.
Forêt aléatoire	93.2%	93.7%	94.2%	94.1%	94.5%	95.3%
GBM	77.2%	77.9%	78.3%	78.4%	78.9%	79.9%
KNN	73.1%	73.6%	74.2%	74.1%	74.4%	74.9%
Réseau de Neurones perceptron multicouches	64.9%	65.6%	65.8%	66.1%	66.8%	67.5%
Régression logistique	56.2%	62%	63.8%	64.5%	67.3%	70.4%
SVM	55.5%	55.8%	56.0%	56.1%	56.4%	56.9%

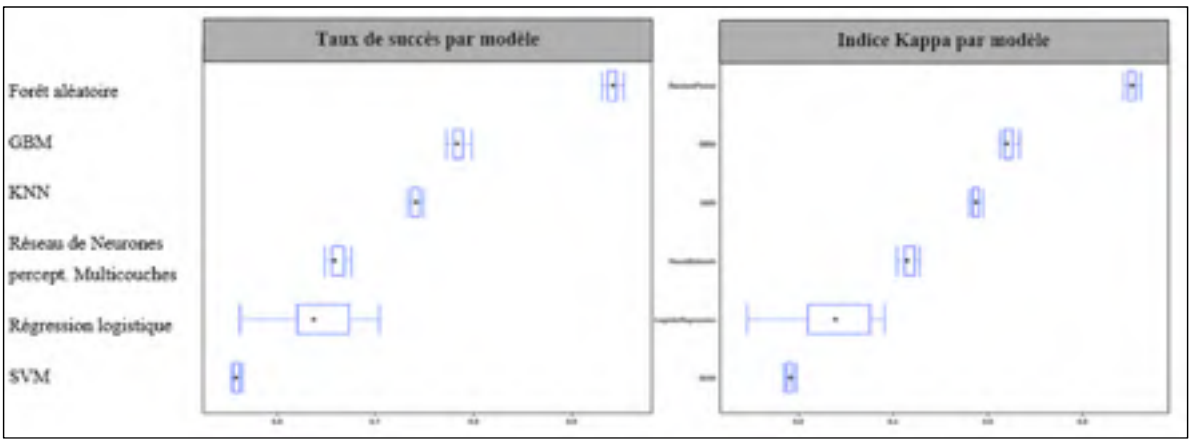


Figure 6.2 Résultats de performance des modèles pour des données d'entrainement avec 10 échantillons aléatoires Contrat : ESM6

- les résultats en haut ont été obtenus en testant nos modèles sur 10 échantillons de données aléatoires sélectionnées à partir des données d'entrainement.

- nous avons utilisé la validation croisée, où à chaque fois 9 échantillons sont utilisés pour entrainer nos modèles et l'échantillon restant pour évaluer la performance. Cet exercice est répété pour chacun des dix échantillons. À la fin on calcule la moyenne et la variance des performances obtenues sur l'ensemble des échantillons testés, pour avoir une estimation plus robuste sur la performance des modèles et voir comment elle change à la suite des changements dans les données.

- on peut voir que la variance de la régression logistique est beaucoup plus élevée comparé aux autres modèles. Ceci n'est pas un bon signe, car ce modèle varie davantage à mesure que les données à tester change, ce qui veut dire que le modèle manque de robustesse et pourra probablement souffrir d'un surapprentissage.

- comme on le voit sur la figure 6.2, le taux de succès et la statistique Kappa sont très bons, surtout pour les méthodes d'ensemble. Le taux de succès moyen pour le modèle forêt aléatoire tourne autour de 94% et il est de 78% en moyenne pour le modèle GBM.

- il est important de savoir qu'il s'agit ici d'un test pour voir comment nos modèles apprennent à partir des données d'entrainement. Compte tenu des résultats obtenus, on peut conclure que la majorité de nos modèles, à l'exception du SVM et de la régression logistique, ne souffrent pas de problème de sousapprentissage. Toutefois, pour avoir des résultats fiables sur la performance espérée, On doit utiliser de nouvelles données qui n'ont pas été utilisées dans l'entrainement de nos modèles.

- les faibles résultats du SVM (avec une fonction du noyau linéaire) et de la régression logistique s'explique par le fait que les deux sont des séparateurs linéaires. c'est-à-dire qu'ils cherchent à séparer linéairement des classes qui ne sont pas distribuées de façon linéaire.

6.1.1.2 Résultats de l'évaluation sur des données de validation

a) Résultats pour des données non agrégées

Le tableau 6.3 contient les résultats obtenus en testant nos modèles sur l'ensemble d'évaluation pour des données d'une minute, pour le contrat à terme E-mini S&P 500.

Tableau 6.3 Résultats de la performance des modèles pour des données d'évaluation - fréquence = 1 minute

Modèle	**% bonnes classifications Tendance = 'Baisse'**		**% bonnes classifications Tendance = 'Neutre'**		**% classés correctement Tendance = 'Hausse'**		**Taux de succès global (Accuracy)**	**AUC ROC**
	Non Équilibrée	Équilibrée[6]	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée		
Forêt aléatoire calibré[7]	36%	60%	69%	65%	63%	65%	**60%± [0.9%]**	**58%**
Gardient Boosting Machine (GBM)	37%	60%	68%	67%	41%	59%	56%±[0.9%]	56%
K plus proches voisins (Knn)	25%	52%	55%	57%	46%	59%	45%±[0.9%]	55%
Arbre de décision C5	27%	54%	60%	62%	47%	59%	47%±[0.9%]	54%
Régression logistique	2%	51%	73%	68%	73%	70%	55%±[1.2%]	57%
Machine à support vectoriel (SVM)	2%	50%	84%	61%	45%	62%	52%±[0.9%]	55%
Réseau de neurones	2.5%	50%	82%	62%	47%	62%	51%±[0.9%]	53%

[6] :Ici on calcule la performance par classe en prenant en considération la distribution non équilibrée des données. Elle est définie comme la moyenne du, rappel et de la spécificité, obtenue sur chaque classe.
Taux de succès balancé = (rappel + spécificité) / 2

[7] : Nous avons utilisé des seuils de probabilités équivalentes à la distribution de nos classes des données d'entrainement pour calibrer les classifications de notre modèle.

On voit comment la performance de nos modèles a baissé significativement en testant nos modèles sur des données d'évaluation qui non jamais été vues par nos modèles. Par contre, on voit toujours que ce sont les méthodes d'ensemble qui donnent les meilleurs résultats.

Il est important de mentionner encore que nous sommes en face d'un problème de classification multiple (3 classes). Et en conséquence, si nous essayons aléatoirement de deviner la tendance pour chaque point des données de l'ensemble d'évaluation, la probabilité d'avoir une bonne classification devra converger vers 1/3 et devra être la même pour chaque classe, si la distribution des classes est uniforme. Si non, pour chaque classe elle devra se rapprocher de la probabilité empirique calculée à partir des données utilisées. C'est similaire à lancer une pièce de monnaie. Si la pièce n'est pas biaisée, on a 50% de chance d'avoir pile ou face. En revanche, si elle est biaisée avec des probabilités de 1/3 pour pile et 2/3 pour face. On va s'attendre à avoir le 1/3 des lancers piles et les 2/3 des faces, en répétant l'expérience pendant un grand nombre de fois[8]. Toutefois, pour toutes les classes de tendance combinées, la probabilité moyenne d'avoir une bonne prédiction[9] devra se situer aux alentours de 1/3.

Interprétation des résultats :

Si on suppose que face à nos modèles, nous avons une méthode d'investissement purement aléatoire. Nous aurons pour chaque point de l'ensemble d'évaluation une probabilité d'avoir raison qui est égale à 1/3. Il s'agit d'une distribution binomiale avec p = 1/3 et n : le nombre des points à prédire = 8,675.

Pour notre meilleur modèle nous avons été capable de prédire correctement 5,205 points. Si on suppose que nous avons une stratégie qui génère des prédictions au hasard, on a la probabilité d'avoir ce nombre de bonnes classifications[10] : $P(X = 5{,}205) <= 0.00001$.

[8] En utilisant la loi faible des grands nombres

[9] On peut le prouver en utilisant le théorème de bayes et en supposant qu'aucune autre information n'est utilisée pour la prédiction. $P[Bonne\ classification] = \sum_{i=1}^{n} P\left[\hat{Y}_i = C_j \mid Y_i = C_j\right] . P\left[Y_i = C_j\right]$ Où Cj représente la classe j.

[10] En utilisant la formule (1.2) : $Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ et en appliquant le théorème central limite, vu que n est grand. Notre distribution binomiale pourra être approximée par une distribution normale centrale réduite.

Autrement dit, on sait que nous avons une probabilité de plus de 99 % d'avoir un nombre de succès inférieur à celui qui a été prédit correctement par notre modèle.

D'après (1.3) nous avons $E(x) = n.p = 8{,}675 * \frac{1}{3} = 2{,}892$, l'espérance mathématique du nombre de bonnes classifications attendues en utilisant une stratégie aléatoire.

Le taux de succès moyen d'une méthode de prédiction aléatoire est égal à 1/3 (toute classe combinée). Et donc, avoir 60% de succès en utilisant notre meilleur modèle est un très bon résultat. Il est aussi important d'analyser ce taux par classe. On peut alors constater que nos modèles, à l'exception des méthodes d'ensemble, ont plus de difficulté à prédire correctement la tendance baissière. Les résultats de certains modèles, tel que la régression logistique et le SVM ne sont pas meilleurs qu'une stratégie aléatoire pour prédire la tendance baissière. Il est possible que les méthodes d'ensemble résistent davantage aux distributions non équilibrées comparés autres modèles.

Dans notre sélection du meilleur modèle, nous allons exclure tous les modèles ayant un taux de classifications correctes, inférieure à la fréquence empirique par classe. Seuls les modèles ayant des taux de succès plus élevés pour chaque classe seront retenus.

D'après les résultats de la régression logistique, on peut voir que le taux de succès pour la classe baissière est presque nul (2%). Donc, il s'agit d'un modèle qui est biaisé et que nous devons exclure de nos choix de modèles.

Le SVM prédit correctement 84% des instances neutres, mais seulement 2% des instances baissières, ce qui est inférieur à la fréquence empirique pour la classe baissière. Et en conséquence, ce modèle doit aussi être exclu.

En optant pour cette approche, on voit que la forêt aléatoire calibrée et le GBM sont les modèles qui donnent les meilleures performances.

En bas nous avons tracé la courbe ROC pour les modèles, forêt aléatoire et GBM.

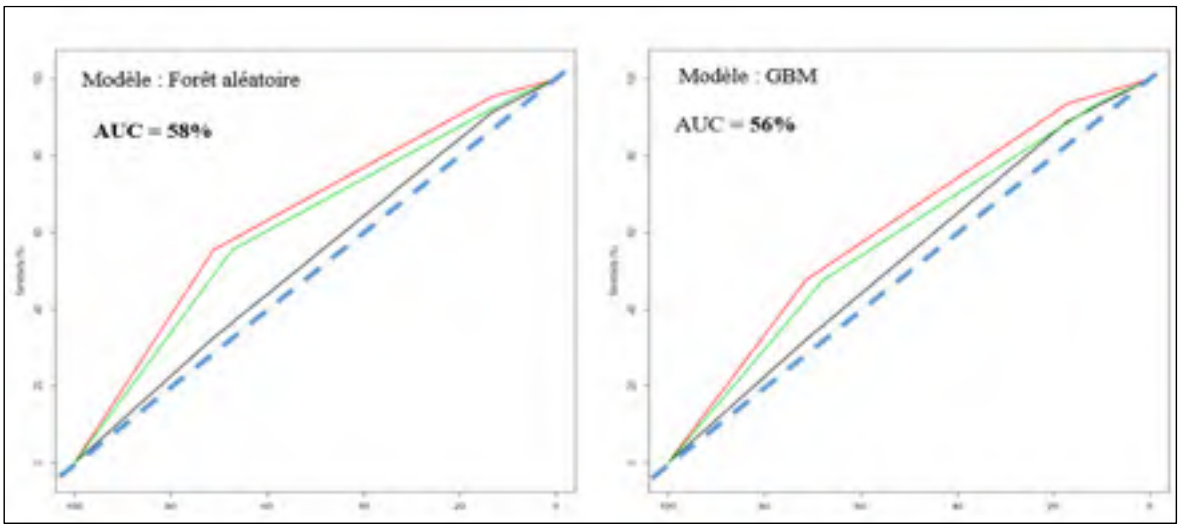


Figure 6.3 La courbe ROC des modèles GBM et foret aléatoire pour les données d'une minute

Sur les deux courbes ROC en haut nous avons le taux des vrais positifs (la sensitivité) versus le taux des faux positifs (la spécificité). Chacune des trois courbes représente une classe. La ligne oblique en bleu au milieu représente les résultats d'une classification aléatoire. L'aire au-dessous de la courbe représente le degré de distinction que le modèle fait entre les classes.

On voit comment les résultats pour les trois classes sont au-dessus de la ligne représentant la classification aléatoire. Ce qui veut dire que nos deux modèles sont meilleurs que des choix purement aléatoires, pour chacune des classes.

Pour notre meilleur modèle 'forêt aléatoire calibré', et en tenant compte de la distribution non équilibrée de nos données, on peut conclure qu'on est à peu près **80%** meilleur qu'une stratégie basée sur le hasard (marche aléatoire)[11].

[11] (Taux de succès obtenu / taux de succès d'une classification aléatoire = 1/3) - 1

b) Résultats pour des données agrégées à 5 minutes

En bas nous avons les résultats de nos tests pour des données que nous avons agrégées à 5 minutes, toujours pour le contrat E-mini sur l'indice S&P 500.

Tableau 6.4 Résultats de la performance des modèles pour des données d'évaluation - fréquence = 5 minutes, Contrat : E-mini S&P 500

Modèle	% bonnes classifications Tendance = 'Baisse'		% bonnes classifications Tendance = 'Neutre'		% classés correctement Tendance = 'Hausse'		Taux de succès global	AUC ROC
	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée		
GBM	13%	53%	67%	55%	39%	56%	**46%±[2.4%]**	**54%**
Forêt aléatoire	10%	52%	65%	52%	33%	52%	44%±[2.4%]	52%
KNN	20%	51%	55%	49%	25%	49%	39% ±[2.6%]	49%
Régression logistique	8%	51%	77%	52%	25%	54%	49%±[2.8%]	53%
Arbre de décision C5	13%	53%	64%	56%	35%	53%	44%±[2.3%]	53%
Réseau de neurones	13%	50%	52%	50%	41%	54%	40%±[2.3%]	50%

En comparant les résultats obtenus pour les données non agrégées avec les résultats des données de 5 minutes, Il est clair que la performance de nos modèles s'est dégradée de façon significative. Les prédictions pour la classe baissière sont maintenant pire qu'un choix aléatoire. Le graphe de la courbe ROC montre aussi la même chose.

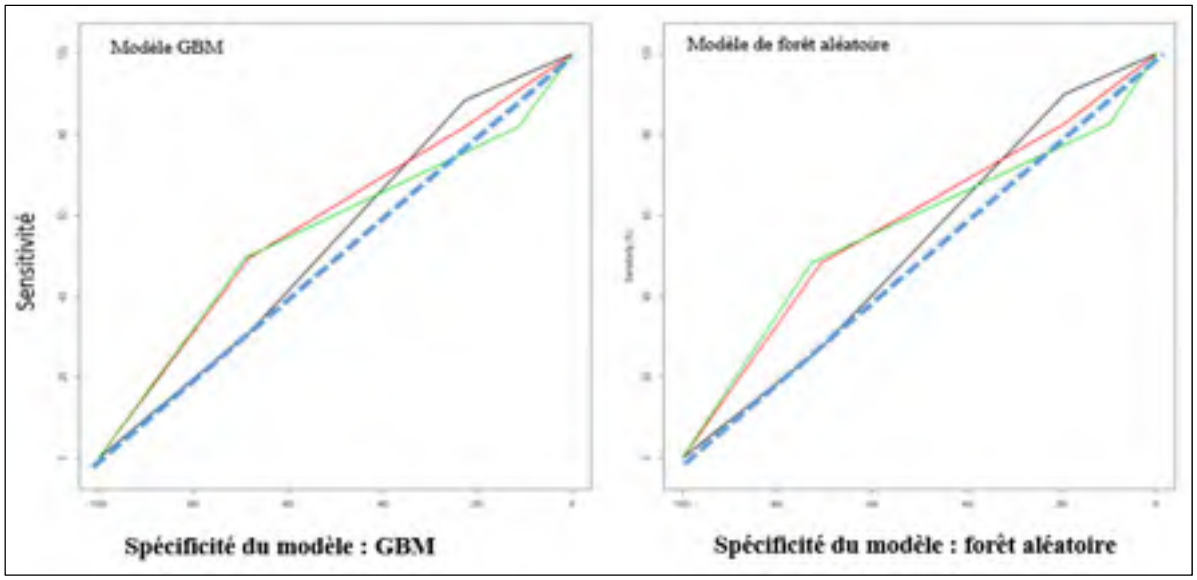


Figure 6.4 La courbe ROC pour le modèle GBM et forêt aléatoire pour des données agrégées à 5 minutes

c) Résultats pour des données agrégées à 15 Minutes

Ici nous avons utilisé un paramètre d'agrégation de 15 minutes tout en gardant les mêmes paramètres que les tests précédents. L'objectif étant de voir comment la performance changera suite à ce changement. Le sommaire des résultats obtenu est présenté dans le tableau en bas.

Tableau 6.5 Résultats de la performance des modèles pour des données d'évaluation fréquence = 15 minutes, Contrat : E-mini S&P 500

Modèle	**% bonnes classifications Tendance = 'Baisse'**		**% bonnes classifications Tendance = 'Neutre'**		**% bonnes classifications Tendance = 'Hausse'**		**Taux de succès global**	**AUC ROC**
	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée		
KNN	31%	50%	28%	60%	61%	55%	44%±[4.3%]	55%
Régression logistique	12%	52%	6%	50%	90%	55%	48%±[5.3%]	53%
SVM	13%	52%	20%	55%	89%	59%	48%±[4.3%]	55%

Modèle	% bonnes classifications Tendance = 'Baisse'		% bonnes classifications Tendance = 'Neutre'		% bonnes classifications Tendance = 'Hausse'		Taux de succès global	AUC ROC
	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée		
Réseau de neurones	37%	55%	20%	54%	62%	56%	44%±[4.3%]	54%
GBM	21%	52%	27%	59%	76%	57%	47%±[4.3%]	53%
Forêt aléatoire	20%	51%	14%	53%	80%	57%	45%±[4.2%]	54%

Interprétation des résultats :

On voit une légère amélioration des résultats par rapport aux résultats des données agrégées à 5 minutes. Toutefois, ils restent inférieurs à ceux obtenus en utilisant des données non agrégées. On constate aussi que la variance du taux de succès (en rouge, nous avons l'écart type) de tous les modèles a augmenté significativement, presque le double pour certains modèles.

Tel que le montre les résultats des trois tests précédents pour le contrat E-mini sur l'indice S&P 500, l'agrégation des données n'améliore pas la performance de nos modèles. Par contre, elle la dégrade. En plus la variance du taux de succès s'élargit plus en fonction du paramètre d'agrégation.

Plus que ce dernier est élevé, plus que la variance tend à s'élargir davantage, ce qui augmente l'incertitude de nos modèles et réduit leurs robustesses. Ceci est dû à la diminution de la taille des données utilisées pour entrainer et tester nos modèles.

Pour le reste de ce chapitre nous allons utiliser seulement des données non agrégées.

6.1.2 Contrat à terme sur le pétrole (CLG8)

Tableau 6.6 Données et critères utilisés pour le contrat à terme CLG8

	Données d'apprentissage	Données d'évaluation
Période couverte	Du : 30 Juin. 2016 Au : 05 Aout. 2016	Du : 06 Aout. 2016 Au : 22 Aout. 2016
Granularité des données brutes	1 minute	1 minutes
Nombre de points (données brutes)	27,759	10,629
Taille d'un contrat	1 000 barils 42 000 gallons	1 000 barils 42 000 gallons
Taille du tique	0,010 $ par baril 10 $ par contrat	0,010 $ par once 10 $ par contrat
Valeur d'un point	1000 $	1000 $
Temps d'exposition	120 minutes	120 minutes
Seuil de profitabilité	0.2 points (200$)	0.2 points (200$)

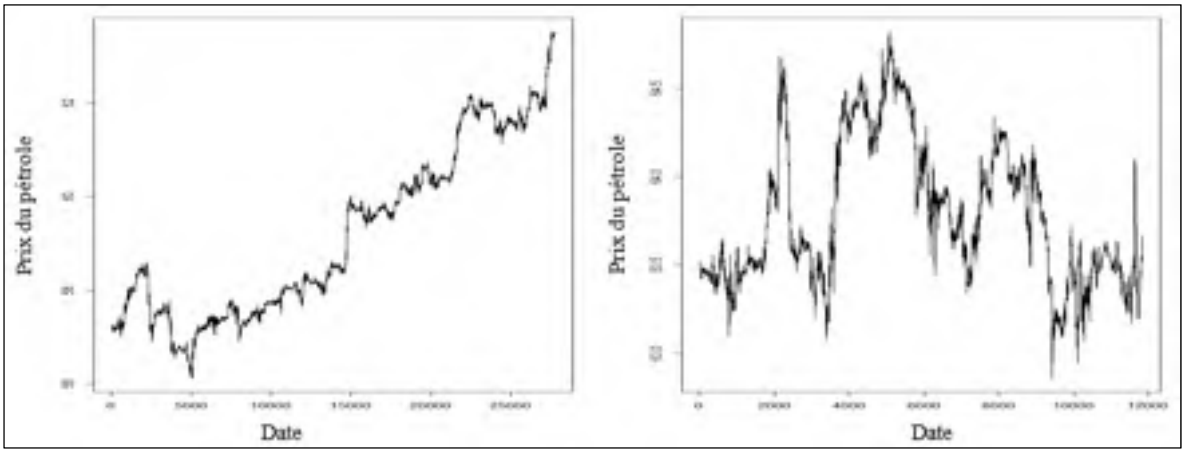


Figure 6.5 Comparaison entre les données d'entrainement et les données d'évaluation - Contrat : CLG8

Tableau 6.7 Distribution de la tendance pour les données du contrat sur le pétrole CLG8 fréquence = 1 minute

Ensemble de données	Ensemble d'apprentissage			Ensemble d'évaluation		
Tendance	Baisse	Neutre	Hausse	Baisse	Neutre	Hausse
Distribution en %	13%	68%	19%	24%	51%	25%

On peut constater sur la figure 6.5 et le tableau 6.7 comment la distribution de la tendance pour les données d'entrainement diffère de celle de l'ensemble des données d'évaluation. Même si le graphe sur les données d'entrainement montre une tendance haussière nette. Dans la plupart du temps la variation du prix durant le temps d'exposition (120 minutes) n'a pas été suffisante pour la considérer comme haussière. Toutefois, sur l'ensemble d'évaluation on voit plus de volatilité, ce qui génère plus de tendances haussières ou baissières.

Tableau 6.8 Résultats de la performance des modèles pour des données d'évaluation fréquence = 1 minute, Contrat : CLG8 (pétrole)

Modèle	**% bonnes classifications Tendance = 'Baisse'**		**% bonnes classifications Tendance = 'Neutre'**		**% bonnes classifications Tendance ='Hausse'**		**Taux de succès global**	**AUC ROC**
	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée		
GBM	27%	58%	57%	65%	49%	60%	48%	48%
Forêt aléatoire	57%	65%	56%	64%	28%	54%	53%	**54%**
KNN	13%	50%	73%	59%	35%	59%	49%	53%
Régression logistique	7%	51%	98%	55%	2%	50%	69%	48%
SVM	0%	50%	100%	50%	0%	50%	51%	50%
Réseau de neurones	13%	50%	81%	60%	32%	61%	53%	**54%**

Interprétation des résultats :

- le tableau en haut montre encore une fois, que le modèle de la forêt aléatoire est celui qui donne les meilleurs résultats. Pour les trois classes, le taux de succès par classe est supérieur à la fréquence empirique des données d'évaluation. L'aire sous la courbe ROC est de 54%. Ce qui veut dire que nos prédictions sont relativement meilleures qu'une classification aléatoire.

- à l'opposé du modèle GBM qui ne prédit pas bien la tendance haussière. La forêt aléatoire prédit correctement, un peu plus de 57% des instances baissières. Soit l'équivalent de de 65% pour des données équilibrées. Cependant, ce dernier modèle à de la misère à prédire correctement la tendance haussière, mais il reste quand même, meilleur qu'une classification aléatoire. Les deux modèles ont pratiquement le même taux de succès pour la classe neutre.

6.1.3 Contrat à terme sur l'or (GCZ7)

Tableau 6.9 Résultats de la performance des modèles pour des données d'évaluation - fréquence = 1 minute, Contrat : GCZ7 (l'or)

	Données d'apprentissage	Données d'évaluation
Période couverte	De : 24 Aout. 2017 Au : 06 Nov. 2017	De : 06 Nov. 2017 À : 29 Nov. 2017
Granularité des données brutes	1 minute	1 minutes
Nombre de points (données brutes)	69,917	23,723
Taille d'un contrat	1 00 onces	100 onces
Taille d'une tique	0,10 $ par once 10 $ par contrat	0,10 $ par once 10 $ par contrat
Valeur d'un point	100 $	100 $
Temps d'exposition	120 minutes	120 minutes
Seuil de profitabilité	2 points (200$)	2 points (200$)

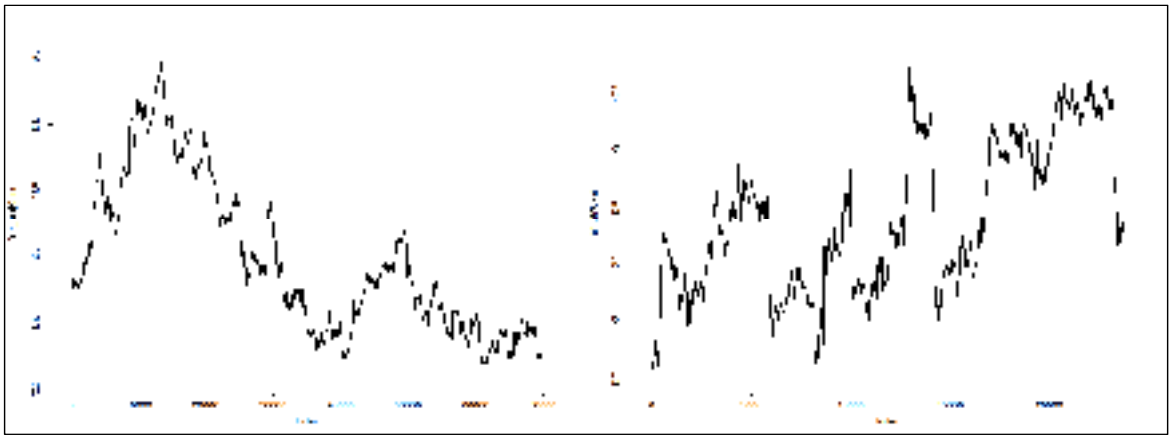


Figure 6.7 Comparaison entre les données d'entrainement et les données d'évaluation - Contrat : GCZ7

Tableau 6.10 Distribution de la tendance pour les données du contrat GCZ7 fréquence = 1 minute

Ensemble de données	**Ensemble d'apprentissage**			**Ensemble d'évaluation**		
Tendance	Baisse	Neutre	Hausse	Baisse	Neutre	Hausse
Distribution en %	26%	48%	25%	16%	62%	22%

Tableau 6.11 Résultats de la performance des modèles pour des données d'évaluation - fréquence = 1 minute, Contrat : GCZ7 (Or)

Modèle	**% bonnes classifications Tendance = 'Baisse'**		**% bonnes classifications Tendance = 'Neutre'**		**% classés correctement Tendance = 'Hausse'**		**Taux de succès global**	**AUC ROC**
	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée		
GBM	35%	55%	50%	60%	40%	55%	49%	52%
Forêt aléatoire calibrée	31%	56%	66%	59%	28%	54%	52%	**53%**

Modèle	% bonnes classifications Tendance = 'Baisse'		% bonnes classifications Tendance = 'Neutre'		% classés correctement Tendance = 'Hausse'		Taux de succès global	AUC ROC
	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée	Non Équilibrée	Équilibrée		
Régression logistique	0%	50%	99.%	53%	1%	49%	69%	50%
SVM	2%	51%	99%	54%	4%	51%	63%	48%
Réseau de neurones	18%	54%	88%	61%	15%	54%	61%	50%
KNN	26%	53%	65%	57%	23%	53%	50%	51%

6.1.4 Conclusion sur les résultats de la performance statistique

- pour les trois contrats utilisés pour évaluer nos modèles, les méthodes d'ensemble étaient les meilleurs en termes de performance prédictive.

- le modèle 'forêt aléatoire' surpasse les autres modèles. Notre métrique de sélection était l'aire sous la courbe ROC. Les résultats de ce modèle ont toujours été meilleurs qu'une classification aléatoire.

- il est important de mentionner qu'en terme de rentabilité les modèles qui classifient mieux la classe baissières et la classe haussière génèreront plus de profits, ce qui n'est pas le cas pour la classe neutre. Car la décision à prendre dans le cas d'une tendance neutre serait de ne rien faire. Cependant, pour les autres classes, l'action à prendre sera d'entrer dans une position. Plus que nos modèles sont précis pour les deux classes, plus que la rentabilité générée sera plus élevée.

Dans la section suivante nous allons évaluer la performance financière (rentabilité) de notre modèle, en utilisant les mêmes données d'évaluation utilisées dans l'évaluation statistique.

6.2 Simulation et évaluation de la rentabilité

Dans ce paragraphe nous allons exposer les résultats de nos tests rétroactifs pour mesurer la capacité de nos modèles à générer des profits. Les points suivants ont été considérés dans la conception de notre système de négociation :

- Comme mentionné dans le paragraphe précédent, on va utiliser seulement les données non agrégées (données à la minute) car elles ont une performance prédictive meilleure que les données agrégées.

- Pour prendre en considération la microstructure du marché, et considérer le fait que les ordres envoyés ne seront pas exécutés tout le temps aux prix désirés. Nous allons considérer un facteur '*s'comme* un pourcentage de surcharge à appliquer aux prix de fermeture utilisés. Alors, le prix d'entrée et de sortie seront calculés en fonction du prix de fermeture au moment où le signal d'entrée est obtenu, plus ou moins une surcharge (slippage) dépendamment de la position prise. La logique du calcul est décrite dans les deux équations en bas :

$$P_{entrée} = \begin{cases} P_i * (1+s), & si\ la\ position\ est\ longue \\ P_i * (1-s), & si\ la\ position\ est\ courte \end{cases} \quad (6.1)$$

$$P_{sortie} = \begin{cases} P_{i+t} * (1-s), & si\ la\ position\ est\ longue \\ P_{i+t} * (1+s), & si\ la\ position\ est\ courte \end{cases} \quad (6.2)$$

  À chaque fois quand une position est prise, deux ordres sont envoyés. Un pour fermer la position si elle est gagnante (TP), avec un prix égal au prix d'entrée plus le nombre de points cible espéré (Pi + r) si la position est longue et (Pi - r) si la position est courte.

Le deuxième ordre (SL) est pour limiter la perte, en fermant la position si le seuil de perte est atteint.

$$TP = \begin{cases} P_i + r, & si\ la\ position\ est\ longue \\ P_i - r, & si\ la\ position\ est\ courte \end{cases} \quad (6.3)$$

$$SL = \begin{cases} P_i - L\ , & si\ la\ position\ est\ longue \\ P_i + L, & si\ la\ position\ est\ courte \end{cases} \quad (6.4)$$

Où :

Pi est le meilleur prix sur le marché à l'instant i.

r est le rendement attendu en points sur le prix du sous-jacent.

L est la perte maximale permise en points.

- La marge de maintenance est calculée en fonction de la valeur du contrat. Nous avons considéré qu'au moins 5% à 10% de la valeur du contrat devra être disponible pour pouvoir entrer sur une nouvelle position. Autrement, nous devons emprunter de l'argent. Le pourcentage utilisé est un peu plus que la marge à maintenir en pratique, mais nous voulons s'assurer que la stratégie ne va pas ruiner notre portefeuille. La majorité du temps une partie du capital ne sera pas utilisée, mais elle va servir de garantie. Exemple : marge de maintenance pour E-mini = 6500$, Pour un capital initial de 10,000$, il reste 3500$ qui ne peut pas être utilisé.

- Le nombre de contrat pour chaque position est déterminé en fonction de l'équité disponible et la marge de maintenance nécessaire. On le calcule en prenant la partie entière de la division de l'équité disponible par la marge de maintenance nécessaire pour le contrat à négocier.

- Nous fermons toutes les positions le même jour. Soit, quand le temps d'exposition est terminé, ou quand le seuil de profitabilité ou la perte maximale sont atteints.

Notre système de négociation calcule la différence entre le temps qui reste avant la fermeture du marché pour le contrat transigé et le temps d'exposition. S'il reste moins que trente minutes, toutes les prédictions générées seront neutres. C'est-à-dire qu'aucune décision ne doit être prise.

6.2.1 Contrat à terme sur l'indice boursier Standard & Poor's 500

Les paramètres suivants ont été utilisés dans notre test d'évaluation de ce contrat :

- contrat : ESM6
- temps d'exposition : 120 minutes
- rendement cible : 4 points (200$)
- maximum de perte accepté par transaction : 2 points (100$)
- coût par transaction : 2.5$
- facteur de surcharge (slippage) = 0.025%

Le choix du rendement cible et de la perte maximale sont faits de telle sorte que la perte maximale soit égale à la moitié du rendement cible. Nous avons :

$$E[R] = P(Gain).TP - P(perte).SL \quad (6.5)$$

Où :
E[R] est l'espérance mathématique du gain sans inclure les frais des transactions,
TP est le rendement attendu en points (positif),
SL est la perte maximale permise en points (négative).

$$\text{Si : } \mathrm{TP} = 2\,\mathrm{SL} \rightarrow \mathrm{E[R]} = \mathrm{SL}(2\,\mathrm{P(Gain)} - \mathrm{P(perte)}\,);\ \mathrm{SL} > 0 \quad (6.6)$$

Il suffit d'avoir $\mathrm{P(Gain)} = \frac{1}{3}$ Pour avoir un rendement nul. Toute probabilité supérieure à ce seuil aura pour effet de générer un rendement positif, sans considérer les frais de transactions.

Il suffit d'intégrer ces frais avec le facteur de surcharge pour trouver un seuil plus élevé, qui prend en considération ces deux facteurs. Une chose que nous avons appliqué dans nos tests de rentabilité.

On utilise les mêmes paramètres du système pour tous les modèles.

Le tableau en bas montre le détail des transactions effectuées et les résultats obtenus.

Tableau 6.12 Résultats de l'évaluation de la performance financière pour des données d'évaluation - Contrat : ESM6

Capital initial	10,000$
Nombre de transactions	195
Nombre de transactions longues	131 – (67%)
Nombre de transactions courtes	64 – (33%)
Gain moyen par transaction	0.5 points (25$)
Gain moyen pour les transactions gagnantes	3.5 points (175$)
Perte moyenne pour les transactions perdantes	-1.9 points (-95$)
Profit maximum réalisé par position	4.25 points (212.5$)
Perte maximale réalisée	-2.25 points (-112.5$)
Profit ou perte total réalisé(e)	**2662.5$**
Retour sur investissement	**26.62%**

Figure 6.8 Équité cumulative en utilisant le modèle sélectionné versus le prix du contrat ESM6

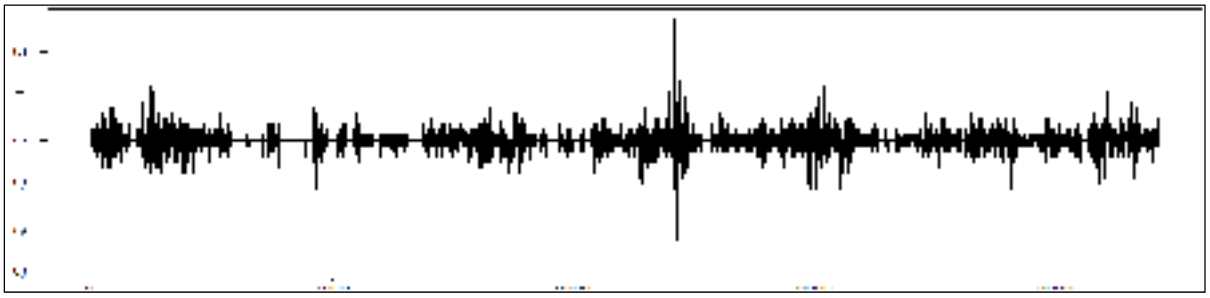


Figure 6.9 Rendement du modèle par point de données - contrat : ESM6

6.2.2 Contrat à terme sur le pétrole (CLG8)

Les paramètres suivants ont été utilisé dans notre test d'évaluation de ce contrat :

- contrat : CLG8
- temps d'exposition : 120 minutes
- rendement cible : 0.2 points (200$)
- maximum de perte accepté par transaction : 0.1 points (100$)
- coût par transaction : 2.5$
- facteur de surcharge = 1%

Le tableau en bas montre le détail des transactions effectuées et les résultats obtenus.

Tableau 6.13 Résultats de l'évaluation de la performance financière pour des donnée d'évaluation - Contrat : CLG8

Capital initial	10,000$
Nombre de transactions	223
Gain moyen par transaction gagnante	0.165 points (165$)
Perte moyenne par transaction perdante	-0.0903 points (-90.03$)
Profit ou perte moyenne par transaction	0.014051 (+14.05$)
Profit maximum réalisé	0.22 points (220$)
Perte maximale réalisée	-0.12 points (-120$)
Profit ou perte réalisé(e)	**3133.5$**
Retour sur investissement	**31.33%**

Figure 6.10 Équité cumulative en utilisant le modèle sélectionné versus le prix du contrat CLG8

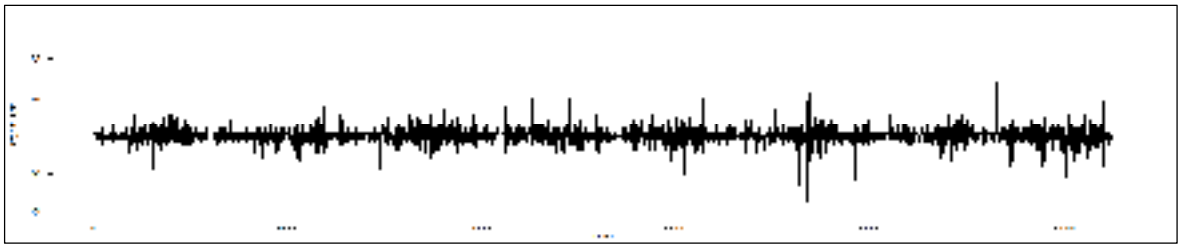


Figure 6.11 Rendement du modèle par point de données pour le contrat CLG8

6.2.3 Contrat à terme sur l'or (GCZ7)

Les paramètres suivants ont été utilisé dans notre test d'évaluation de ce contrat :

- contrat : GCZ7
- temps d'exposition : 120 minutes
- rendement cible : 2.5 points (250$)
- maximum de perte accepté par transaction : 5 points (500$)
- coût par transaction : 2.5$
- facteur de surcharge = 0.1%

Le tableau en bas montre les résultats obtenus.

Tableau 6.14 Résultats de l'évaluation de la performance financière pour des données d'évaluation - Contrat : GCZ7

Capital initial	10,000$
Nombre de transactions	193
Gain moyen par transaction gagnante	1.41 points (141$)
Perte moyenne par transaction perdante	-1.72 points (-172$)
Profit ou perte moyenne par transaction	0.05 points (5$)
Profit maximum réalisé	4 points (400$)
Perte maximale réalisée	-5.8 points (-580$)
Profit ou perte réalisé(e)	**+1840.88 $**
Retour sur investissement	**18.41%**

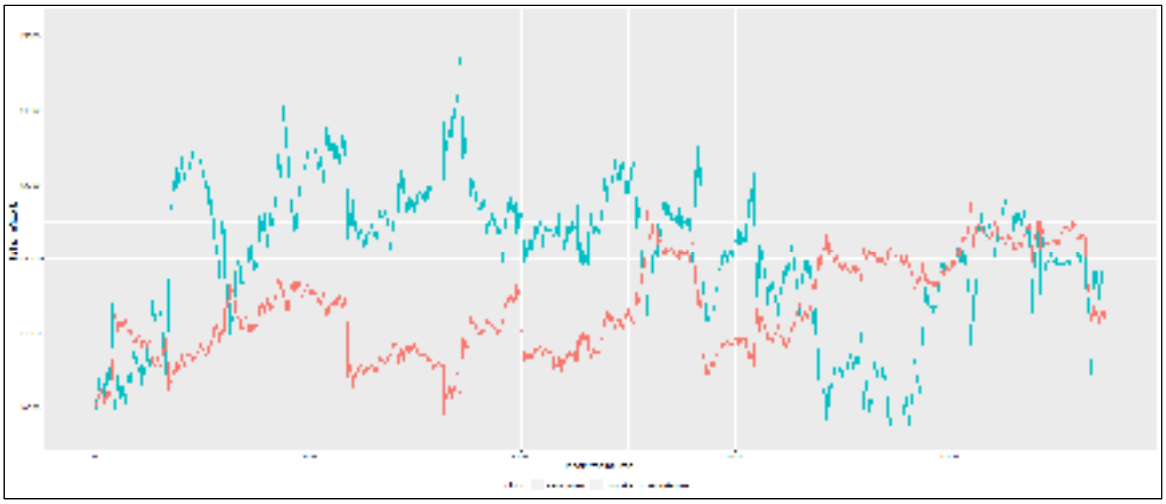


Figure 6.12 Équité cumulative en utilisant le modèle sélectionné versus le prix du contrat GCZ7

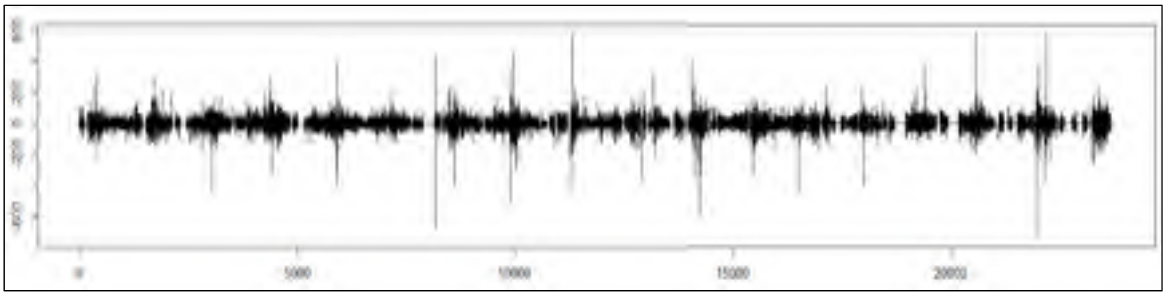


Figure 6.13 Rendement à la minute de notre modèle pour le contrat GCZ7

6.2.4 Sommaire des résultats

Il est important de noter qu'en réalité, il est peu envisageable de prendre une position sur un contrat à terme pour une longue durée. Déjà le contrat a une durée de maturité. La comparaison de nos résultats avec une gestion passive est faite seulement pour avoir une idée sur l'évolution du prix du contrat. Si on assume que la trajectoire de ce prix est purement aléatoire, il est donc quasi-impossible de prédire les prochains mouvements. Ce qui veut dire, que le fait de générer des profits est très peu probable en utilisant des stratégies bien définies.

Le fait de générer de tel profits peut expliquer que la trajectoire du prix de ces contrats fourni assez d'information pour prédire les prochains mouvements.

Il est aussi à noter que les périodes d'évaluation utilisées sont très courtes (de quelques jours à quelques semaines, ne dépassant pas plus de huit semaines) et que tous les contrats utilisés dans n'arrivent pas à échéance durant la période d'évaluation.

Selon les résultats des tests expérimentaux, on voit que le retour sur investissement est positif et supérieur à celui qui serait obtenu si on garde le contrat durant la période d'évaluation, pour les trois contrats utilisés.

Pour le contrat E-mini S&P 500, le modèle a généré plus de 26% de rendement sur une période de 8 jours. La courbe de l'équité accumulée montre une croissance, presque continue et qui dépasse de loin celle d'une gestion passive. Sur la courbe de l'équité accumulée pour le contrat sur le pétrole, on voit plus de volatilité. Toutefois, le retour sur investissement a été positif et supérieur à celui d'une gestion passive dans la majorité du temps. À la fin de la période d'évaluation, le modèle a généré un peu plus de 31% de retour sur investissement sur une période d'évaluation de 13 jours de négociation. En moyenne 17 transactions par jour ont été effectuées et ont généré un peu plus de 14$ par transaction.

Sur la période du 06 au 29 novembre 2017, notre modèle a généré un peu plus de 18% de rendement pour le contrat sur l'or. Pendant une courte période à la fin, la courbe de l'équité accumulée de notre modèle se situait au-dessous de celle qui serait obtenue si nous avons gardé le contrat sans faire aucune transaction. Toutefois, à la fin notre courbe termine un peu au-dessus de l'autre courbe. Ce qui veut dire que notre modèle a généré plus de rentabilité que celle qui serait obtenu si on opte pour une gestion passive.

Il est important de mentionner qu'il est nécessaire de définir de façon judicieuse la politique de gestion de risques. Selon nos simulations, sans les seuils de perte prédéfinies, on se retrouve avec des pertes colossales.

Dans certains cas, juste quelques pertes ont été très couteuses même si le pourcentage des transactions gagnantes était très élevé. Il est même possible qu'une seule perte ruine notre capital si on n'utilise pas de seuils pour limiter nos pertes. Il y a plusieurs techniques pour déterminer ses seuils. Dans ce travail, ceci a été déterminé en se basant sur l'optimisation du gain / perte historique.

CONCLUSION

À travers ce projet nous avons pu répondre aux questions de recherche que nous avons discutées au début du chapitre 1. Cela a été accompli en réalisant une revue de littérature et une analyse expérimentale centrées sur les différents aspects du problème de prévision des séries temporelles financières.

Sur la base de notre analyse expérimentale nous avons abouti à des conclusions qui répondent à nos questions. Ces conclusions ne seront pas répétées dans cette section. Le lecteur est référé à l'analyse expérimentale présentée au chapitre 6. Six modèles d'apprentissage automatique différents ont été conçus et évalués en utilisant trois contrats à terme avec des données d'une granularité d'une minute. Nous avons testé plusieurs scénarios et nous avons abouti à la conclusion que les données granulaires ont un pouvoir prédictif meilleur que les données agrégées.

En analysant les distributions des données nous avons constaté que les classes de la tendance ne sont pas distribuées de façon uniforme. Ceci ajoute une grande complexité à notre tâche de prédiction et réduit la capacité de nos modèles à prédire correctement de nouvelles données. Nous avons proposé trois techniques pour faire face à ce problème. Nous avons pu observer que le fait de calibrer nos données permet d'améliorer le pouvoir prédictif de nos modèles et réduit le risque de sur-apprentissage. Nous avons trouvé aussi que les méthodes d'ensemble résistent d'avantage aux distributions non uniformes des classes de la tendance.

Comme mentionné précédemment dans ce mémoire, l'hypothèse sur l'efficience du marché (voir section 1.1) stipule qu'il est impossible de prévoir la direction du prix d'une série financière. Cependant, plusieurs anomalies et études statistiques que nous avons cité, contredisent cette hypothèse. Les tests d'évaluation rétroactifs que nous avons effectués montrent aussi le contraire. Nous avons démontré mathématiquement que le nombre de bonnes

prédictions obtenu en utilisant nos deux meilleurs modèles de prévision est quasi-impossible à obtenir si on suppose que la trajectoire du prix d'un actif boursier est purement aléatoire.

En termes de rentabilité, sur les trois contrats utilisés, le modèle de prévision que nous avons sélectionné a généré des résultats positifs, qui surpassent une stratégie simple qui consiste à détenir le contrat tout au long de la période d'évaluation. Bien entendu, nous sommes conscients qu'une si simple stratégie est très peu utilisée dans les transactions concernant les contrats à termes. Cependant, nous avons considéré cette stratégie pour uniformiser la comparaison des rendements avec nos stratégies basées sur des techniques d'apprentissage automatique.

Lorsqu'on investit dans le marché boursier, un premier objectif doit être de ne pas perdre de l'argent, ce qu'aucun des tests que nous avons faits n'a produit, comme indiqué précédemment. Le deuxième objectif, qui suit de près, consiste à maximiser les bénéfices de l'investissement et, même si tous les tests d'évaluation ont généré des résultats positifs, il reste encore du travail à faire quant à la performance de la méthode. Les creux sur la courbe de l'équité du contrat sur le pétrole et sur l'or, ne sont pas un bon signe, contrairement au contrat sur l'indice boursier S&P, où l'équité n'affiche pas de grand écart entre pics creux. L'une des recommandations à ce stade, consiste à utiliser l'échantillonnage par roulement tout en calibrant la distribution des classes, tel que décrit dans la section 5.1.3. Ceci aura pour effet de rendre l'apprentissage de notre modèle dynamique et non biaisé, ce qui pourra améliorer d'avantage la qualité de nos prédictions.

ANNEXE I

INDICATEURS TECHNIQUE

Le tableau en bas contient la liste des indicateurs techniques en français et en anglais, utilisés dans ce projet.

Tableau A I.1 : Liste des indicateurs techniques utilisés dans ce projet

Symbole	Terme en anglais	Terme en français
ATR	Average True range	La plage réelle moyenne
RSI	Relative Strength Index	L'indice de la force relative
MMS	Simple moving average	Moyenne mobile simple
MME	Exponential moving average	Moyenne Mobile Exponentielle
VWAP	Volume Weighted Average Price	prix moyen pondéré en volume
MACD	Moving Average Convergence Divergence	Moyenne mobile de convergenge et de divergence
SAR	parabolic stop and reverse	Arrêt et retour parabolique
ADX	Average Directional Index (ADX)	L'indicateur directionnel moyen
BB	Bollingers bandes	Bandes de Bollingers
MFI	Money Flow Index	Indice de flux monétaire
CMO	Chande Momentum Oscillator	Oscillateur Momentum Chande
SMI	Stochastic Momentum Index	L'indice de momentum stochastique
CLV	Close Location Value	L'emplacement de la valeur de fermeture

Définitions des indicateurs techniques :

1. Taux de variation du prix (taux de rendement simple) :

C'est le taux de variation du prix d'actif financier entre un temps t et un temps t + n, où n = 1, 2, 3, etc.

$$R_{k;t} = \frac{P_{t+k}}{P_t} - 1 \qquad \text{(A I.1)}$$

2. Moyenne mobile simple (Simple Moving Average) :

C'est le prix moyen d'un titre à un moment donné. Lors du calcul d'une moyenne mobile, on doit spécifier le temps nécessaire pour calculer le prix moyen (par exemple, 25 jours).

$$MMS(n) = \frac{1}{n} \sum_{t=0}^{n} P_t \qquad \text{(A I.2)}$$

où Pt est le prix de fermeture à l'instant t

3. La Moyenne Mobile Exponentielle (Exponential Moving Average) :

Il s'agit d'une moyenne mobile où on accorde plus d'importance aux prix les plus récents ce qui la rend plus réactive.

$$MME = \alpha(P_t + (1-\alpha) * P_{t-1} + (1-\alpha)^2 P_{t-2} + (1-\alpha)^3 P_{t-3+} \dots) \qquad \text{(A I.3)}$$

4. La plage réelle moyenne (Average True Range ("ATR")) :

C'est une mesure de la volatilité. Il mesure la volatilité du marché en décomposant toute la plage du prix d'un actif pour cette période trading.

on commence par calculer la plage réelle comme suit :

$$TR = Max[(P_{haut} - P_n), abs(P_{haut} - P_{n-1}), abs(P_{bas} - P_{n-1}) \qquad \text{(A I.4)}$$

où P_n est le prix de fermeture à l'instant n.

Le ATR au moment t est calculé en utilisant la formule :

$$ATR_t = \frac{ATR_{t-1}\,(n-1) + TR_t}{n} \qquad \text{(A I.5)}$$

La première valeur du ATR est calculée en utilisant la moyenne arithmétique :

$$ATR = \frac{1}{n}\sum_{i=0}^{n} TR_i \qquad \text{(A I.6)}$$

5. L'oscillateur stochastique (stochastic oscillator) :

C'est un indicateur de momentum comparant un cours de de fermeture d'un titre à une fourchette de prix sur une certaine période. Il est utilisé pour déterminé si un titre est sur-achetés ou sur-vendus, en utilisant une plage de valeurs délimitées par 0 à 100.

L'oscillateur stochastique se calcule en utilisant la formule suivante :

$$\%K_t = \left(\frac{P_t - L14}{H14 - L14}\right) x\ 100 \qquad \text{(A I.7)}$$

Où :

Pt : est le dernier prix de fermeture;

L14 : le prix le plus bas des 14 derniers points;

H14 : le prix le plus haut des 14 derniers points;

%K : la valeur de l'oscillateur stochastique au moment t.

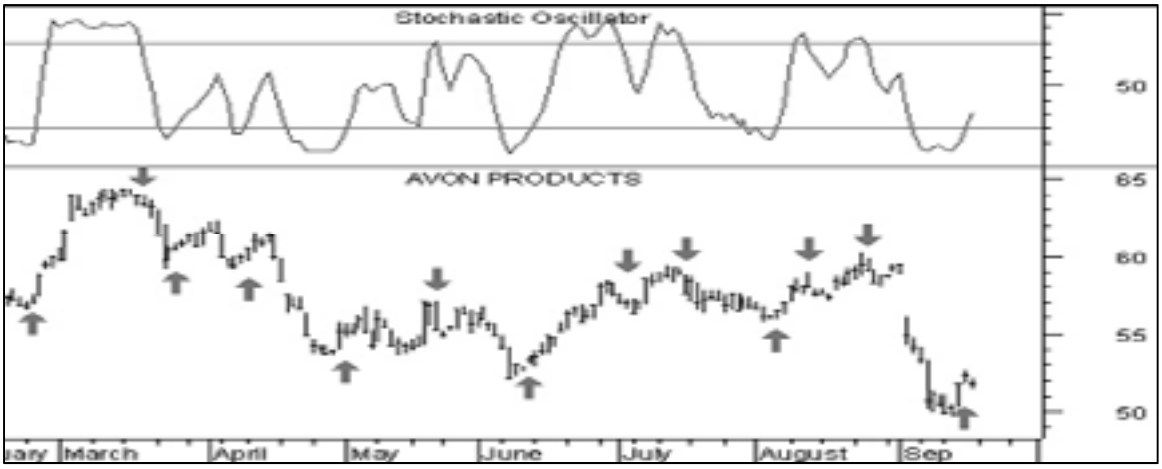


Figure A I.1 Exemple d'un oscillateur stochastique
Tirée de Metastock.com

L'indicateur de Aaron (Aaron Indicator) : cet indicateur est utilisé pour identifier les changements de tendance du prix d'un actif et la force de cette tendance. Il mesure le temps entre les hauts et les bas sur une période donnée. L'idée est que les fortes tendances à la hausse connaîtront régulièrement de nouveaux sommets et que les fortes tendances à la baisse connaîtront régulièrement de nouveaux plus bas. L'indicateur signale quand cela se produit et quand ce n'est pas le cas.

Figure A I.2 Exemple d'un graph boursier avec l'indicateur Aaron
Tirée de Metastock.com

Le prix moyen pondéré en volume / Volume Weighted Average Price (VWAP) : Il correspond à la valeur en dollars de toutes les périodes de négociation divisée par le volume total des transactions pour la journée en cours Le calcul commence à l'ouverture du négoce et se termine à la fermeture. Les périodes et les données intraday étant utilisées dans le calcul, elles ne sont valables que pour le jour de bourse en cours.

$$\text{VWAP} = \frac{\sum Prix\ .Volume}{\sum volume} \qquad \text{(A I.8)}$$

6. Les bandes de Bollingers / Bollingers bandes:

Cet indicateur est composé de trois courbes. La courbe au milieu qui est une moyenne mobile des prix sur n périodes (en général 20). La bande supérieure est de "d" écart types (généralement d = 2) au-dessus de la courbe au milieu et la bande inférieure est de "d" écarts types au-dessous.

Mathématiquement :

Soit P une série de prix de fermeture et σn est l'écart type de cette série.

La bande au milieu (MM(n)) : $\text{MM(n)} = \frac{1}{n} \sum_{t=0}^{n} P_t$

La bande supérieure Bupper (d, n) : $\text{Bupper (d, n)} = \text{MM(n)} + d\ .\ \sigma n$

La bande inferieure Blower (d, n) : $\text{Blower(d, n)} = \text{MM(n)} - d\ .\ \sigma n$

L'écart entre les bandes indique le niveau de volatilité qui existe.

7. Moyenne mobile convergenge divergence / Moving Average Convergence Divergence (MACD) :

Il s'agit simplement de la différence entre deux moyennes mobiles exponentielles de périodes différentes. En général, on utilise des périodes de 12 et 26 jours. On emploie aussi une courbe

de tendance (moyenne mobile exponentielle de la MACD) pour obtenir le signal de la MACD suite à un croisement des deux.

- Un signal d'achat est obtenu quand la MACD traverse à la hausse la ligne de signal.
- Quand on a croisement inverse à la baisse, il s'agit probablement d'un signal de vente

Figure A I.3 Exemple d'utilisation du MACD sur l'action MICROSOFT.
Tirée de Yahoofinance

8. Arrêt et retour parabolique / parabolic stop and reverse (SAR) :

Il est utilisé pour déterminer la direction du prix d'un actif et détecter les changements dans la tendance du prix. L'indicateur utilise une méthode d'arrêt et de retour appelée "SAR" ou "arrêt et retour" pour identifier les points de sortie et d'entrée appropriés. Le SAR parabolique apparaît sur un graphique sous la forme d'une série de points, situés au-dessus ou au-dessous du prix d'un actif, en fonction de l'évolution du prix. Un point est placé en dessous du prix lorsqu'il a tendance à la hausse, et au-dessus du prix lorsqu'il est à la baisse.

Figure A I.4 Exemple d'un SAR parabolique.
Tirée de Metastock.com

9. L'indicateur directionnel moyen / Average Directional Index (ADX) :

Il s'agit d'un indicateur de momentum, ou de la force de la tendance. Deux mouvements directionnels sont générés du système, positifs et négatifs, en comparant la différence entre deux minimums consécutifs et la différence entre leurs maximums respectifs.

10. L'indice de la force relative / Relative Strength Index (RSI)

C'est un oscillateur qui mesure la vitesse à laquelle les mouvements de prix changent et l'ampleur de ces fluctuations sur une période déterminée. De façon générale, quand l'indice dépasse un seuil de 70 points, ceci est un signal que l'actif est suracheté et c'est probablement le moment idéal pour le vendre. Si cette valeur est en bas de 30, ceci est un signal alarmant que le titre est survendu. En effet, le RSI peut aider à savoir si une tendance se précisera ou subira un renversement. Il s'adapte aux fluctuations en fonction des nouveaux cours intégrés dans son calcul.

Figure A I.5 Exemple d'utilisation du RSI sur l'action MICROSOFT.
Généré à partir de Yahoofinance

11. L'indice de flux monétaire / Money Flow Index (MFI) :

C'est un oscillateur qui utilise le prix et le volume pour mesurer la pression des achats et des ventes. Il s'agit en fait d'une version du RSI pondéré en fonction du volume. . Le flux monétaire est positif lorsque le prix typique augmente (pression d'achat) et négatif lorsque le prix typique diminue (pression de vente).

Figure A I.6 Exemple d'un indicateur MFI.
Tiré de Metastock.com

12. Close location value (CLV) :

C'est une mesure utilisée dans l'analyse technique pour déterminer où se situe le prix de l'actif à la fermeture par rapport au prix le plus bas. Le CLV est compris entre +1 et -1, où une valeur de +1 signifie que la fermeture est égale à la valeur haute et une valeur de -1 que la clôture est égale au point bas du jour.

ANNEXE II

ESTIMATION DES PARAMÈTRES

1. **Nombre de K proches voisins**

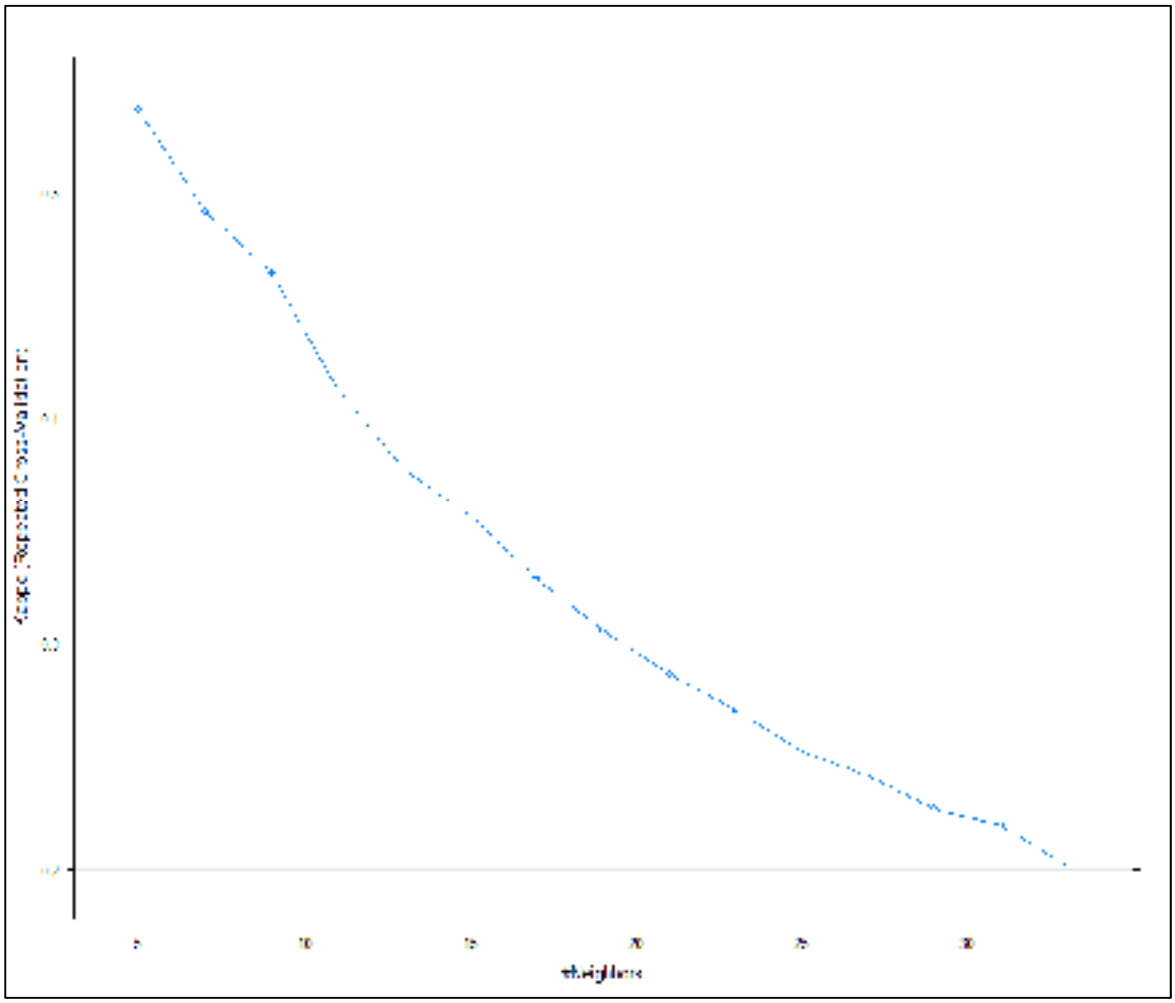


Figure A II.1 La statistique Kappa en fonction du nombre de K proches voisins

On voit que le meilleur score de la statistique Kappa s'obtient avec K = 5 comme nombre de voisins les plus proches.

2. Nombre d'arbres pour la forêt aléatoire

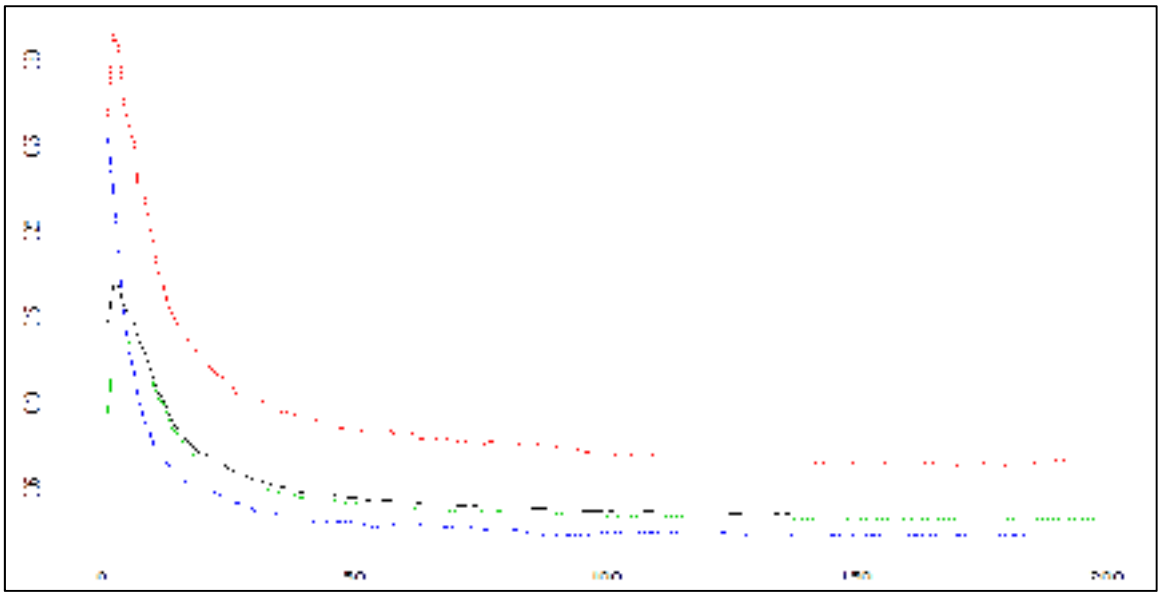


Figure A II.2 Erreurs des forêts aléatoires en fonction du nombre d'arbres

Chaque courbe représente une classe. La courbe en vert représente l'erreur totale (toute classe combinée)

3. Performance prédictive en fonction du nombre de variables sélectionnées

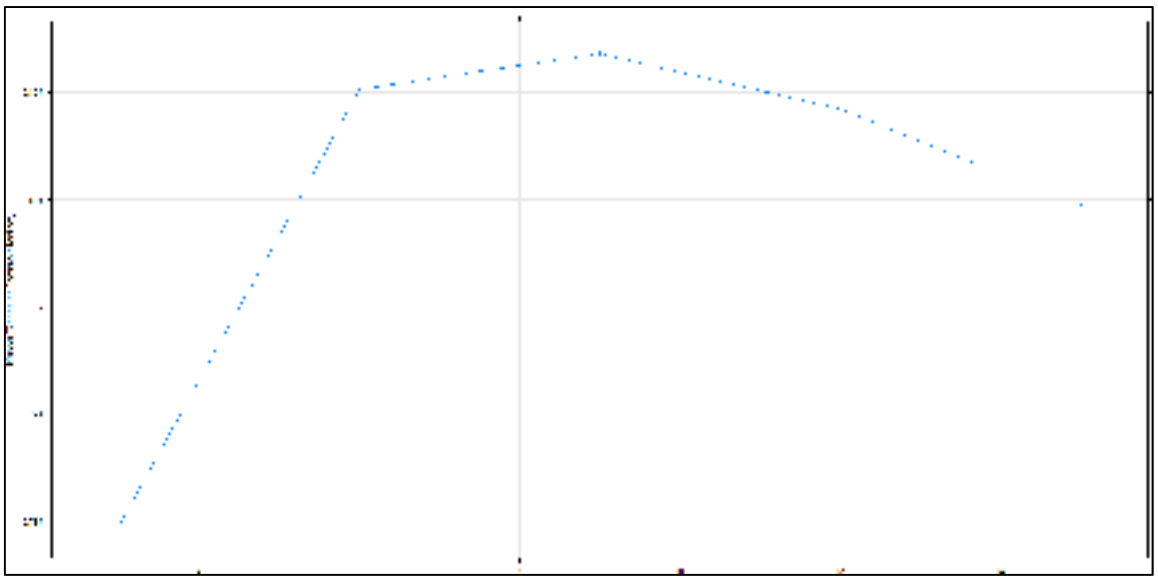


Figure A II.3 Performance de la statistique Kappa par nombre d'indicateurs utilisés

Sur ce graphe nous avons la performance de notre modèle en fonction du nombre d'attributs utilisés (indicateurs techniques). La performance maximale du modèle est atteinte quand on utilise les 9 indicateurs techniques les plus importants.

4. Estimation des paramètres du modèle GBM

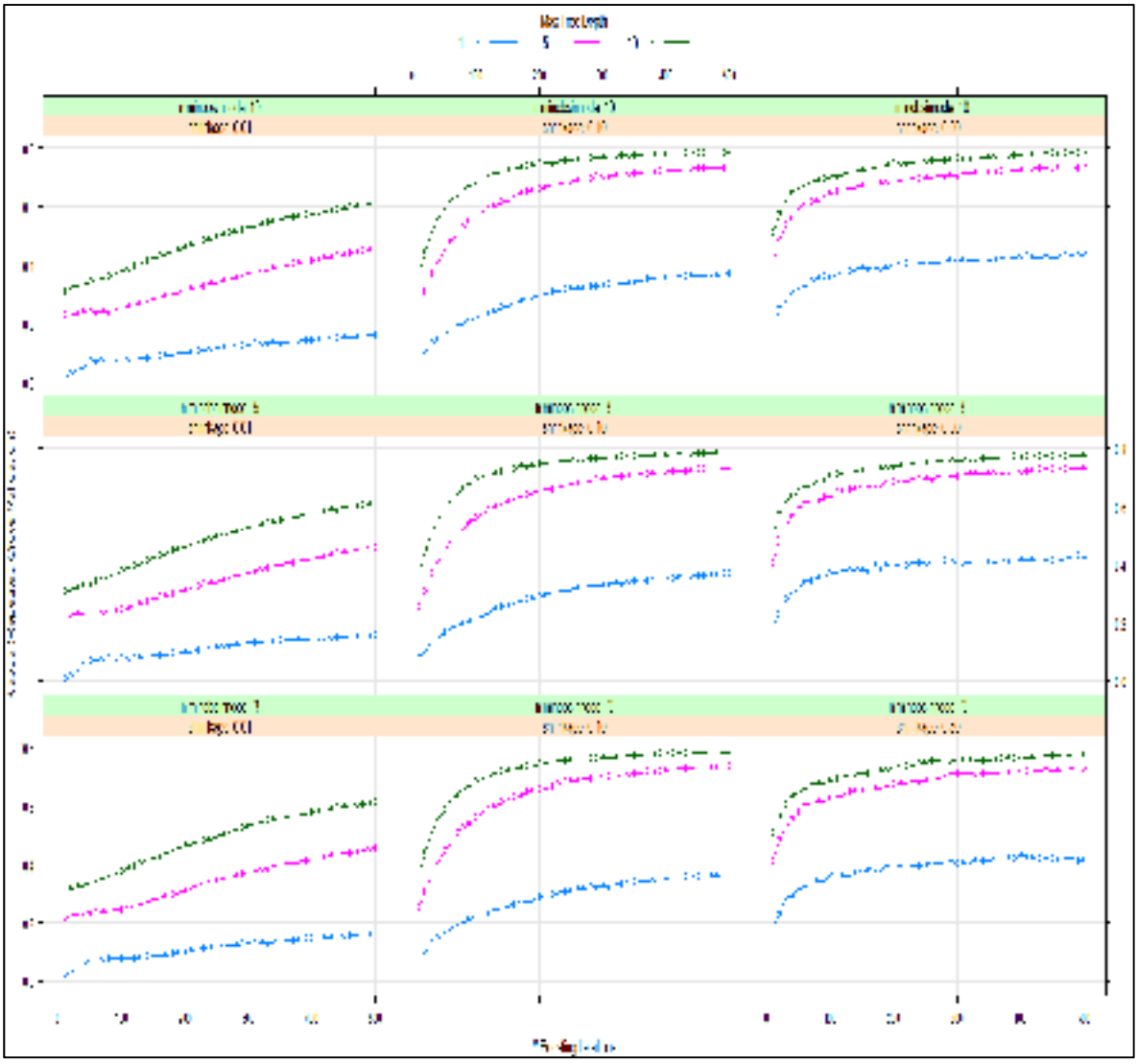


Figure A II.4 La statistique Kappa en fonction de la sélection de paramètres pour le model GBM

Ici on fait une recherche des meilleurs paramètres pour le modèle GBM en variant la valeur de chaque paramètre et observant comment la performance du modèle change.

5. Sélection du nombre de couches pour le réseau de neurones

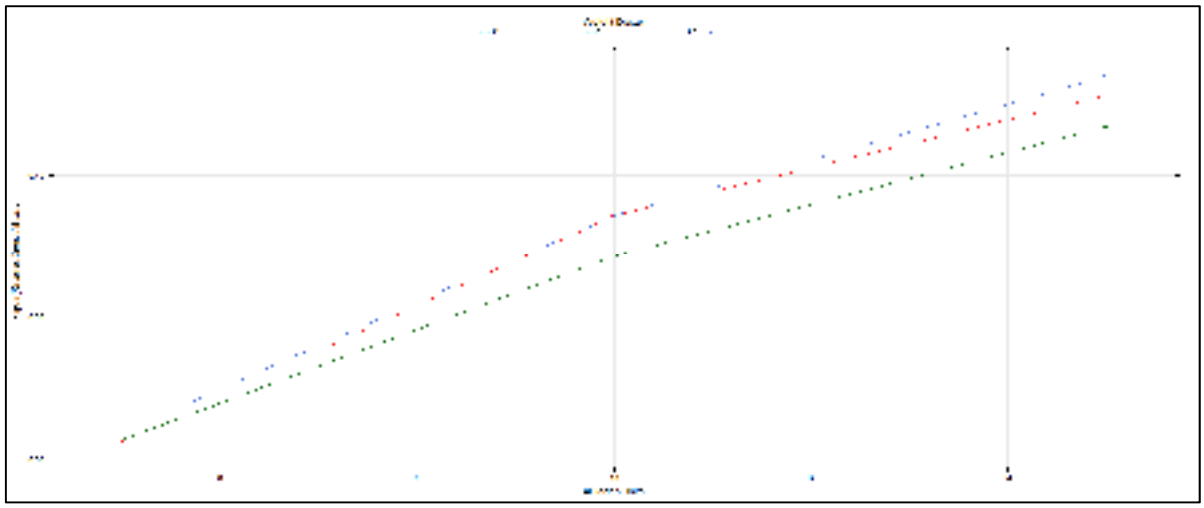


Figure A II.5 La statistique Kappa en fonction du nombre de couches pour un réseau de neurones

Ici nous avons la performance du modèle réseau de neurones en fonction du nombre de couches dans le réseau.

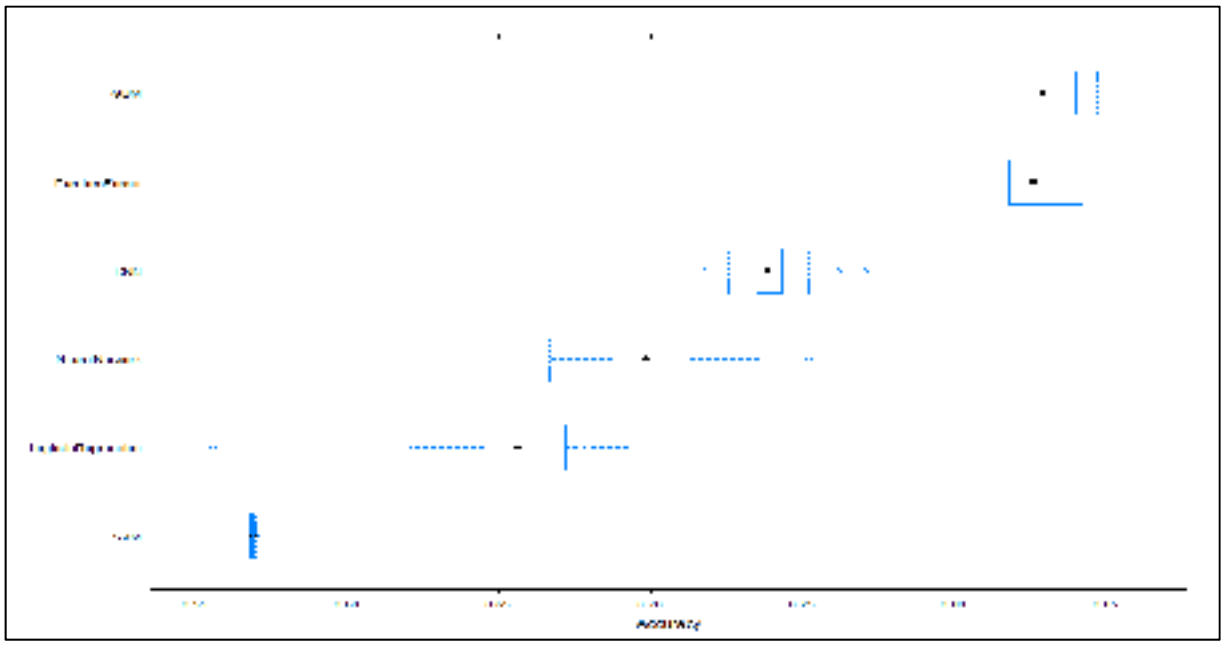


Figure A II.6 La performance des modèles (Taux de succés ou Accuracy) pour les données d'entrainement

ANNEXE III

IMPORTANCE ET SÉLECTION DES ATTRIBUTS

1. **Pouvoir prédictif des attributs utilisés**

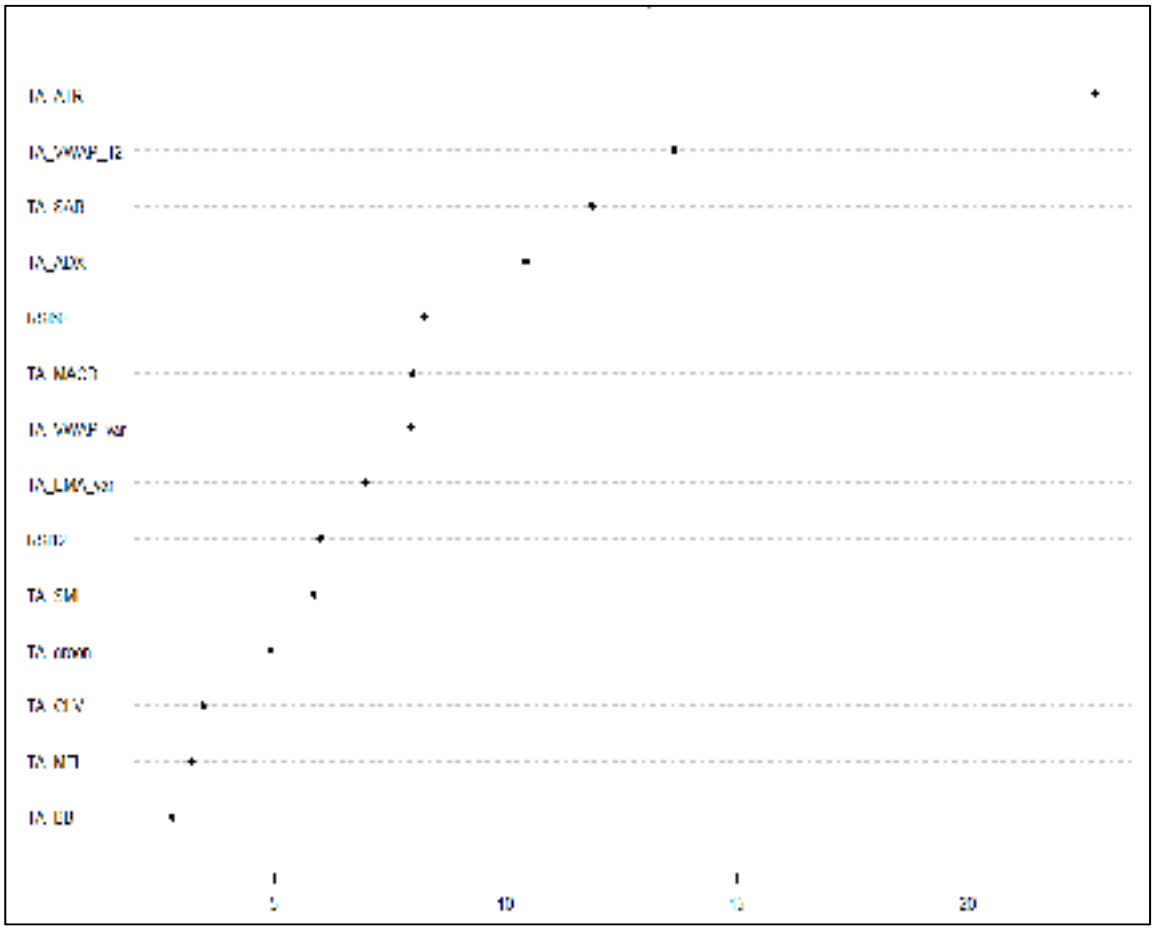


Figure A III.1 Importance des variables en utilisant un modèle de forêts aléatoire

Sur ce graphe, nous avons l'importance de chaque attribut (indicateur technique) en utilisant une forêt aléatoire comme modèle prédictif. Ce modèle utilise une méthode de sélection par élimination itérative. À chaque fois, une variable est exclue, et la moyenne de l'accuracy est observé. À la fin on calcule la moyenne de la baisse en accuracy par variable et l'importance de chaque variable est mesurée en fonction de cette mesure.

2. Distribution de la tendance par variable explicative utilisée

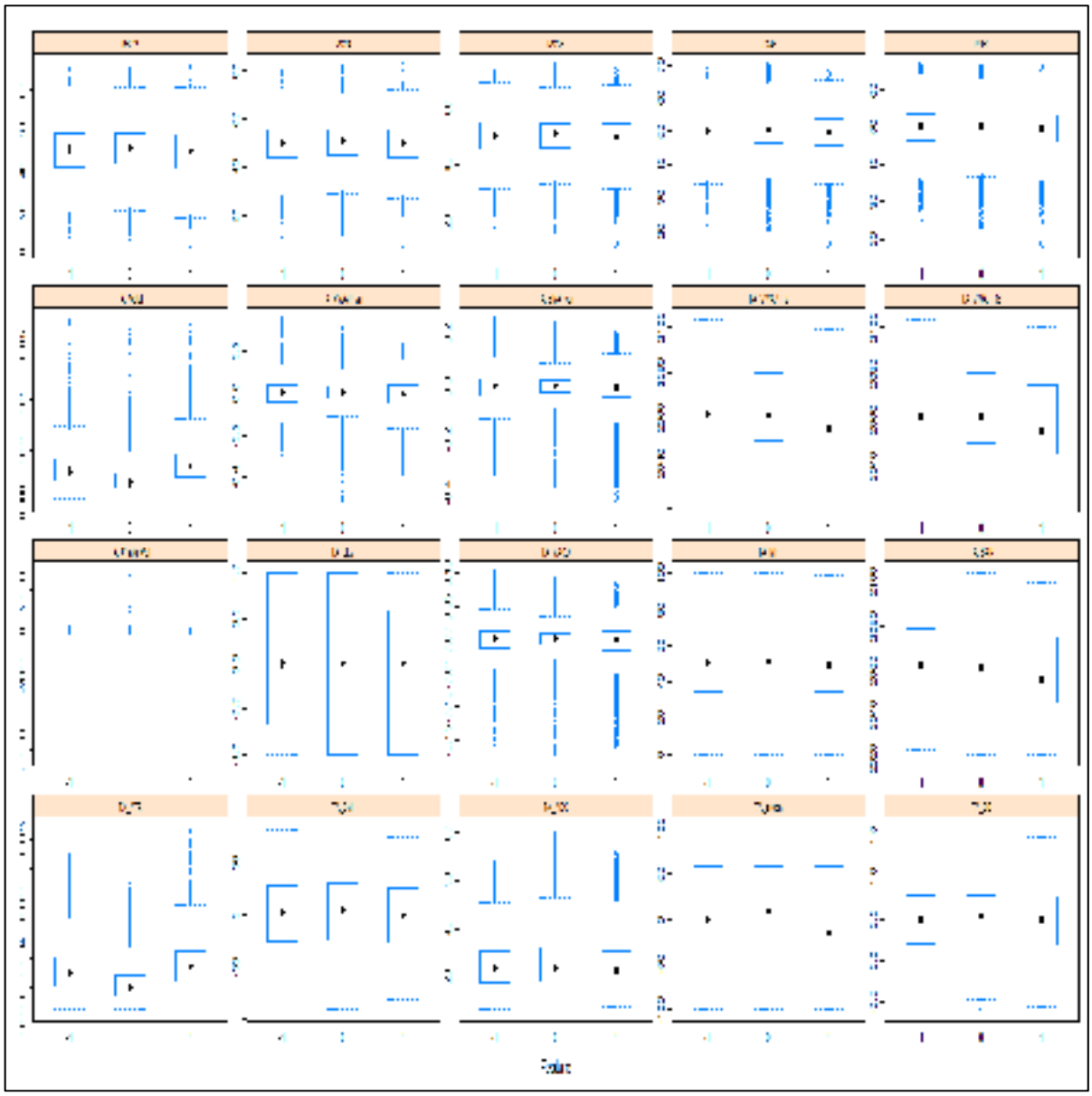


Figure A III.2 Distribution de la tendance par indicateur

Ici nous avons la variation de chaque indicateur par classe de la tendance. L'existence de différence entre les classes est un signe fort que l'indicateur est un bon prédicteur de notre variable cible (la tendance)

3. Sélection des indicateurs les plus importants

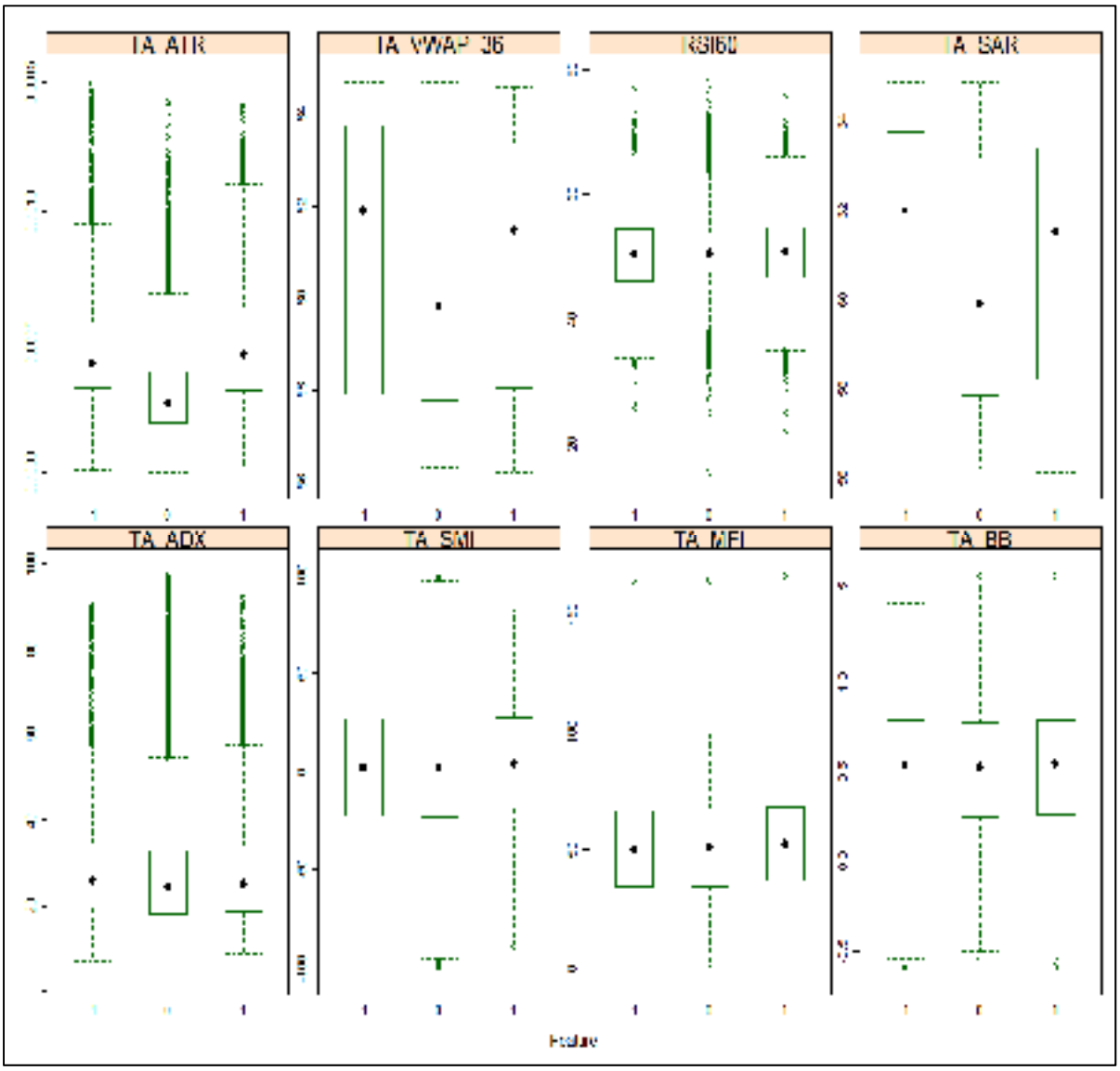


Figure A III.3 Distribution de la tendance pour les 8 indicateurs les plus importants

4. Densités des indicateurs les plus importants

C'est la même chose que le graph précédant, mais ici nous avons gardé seulement les 8 indicateurs les plus importants.

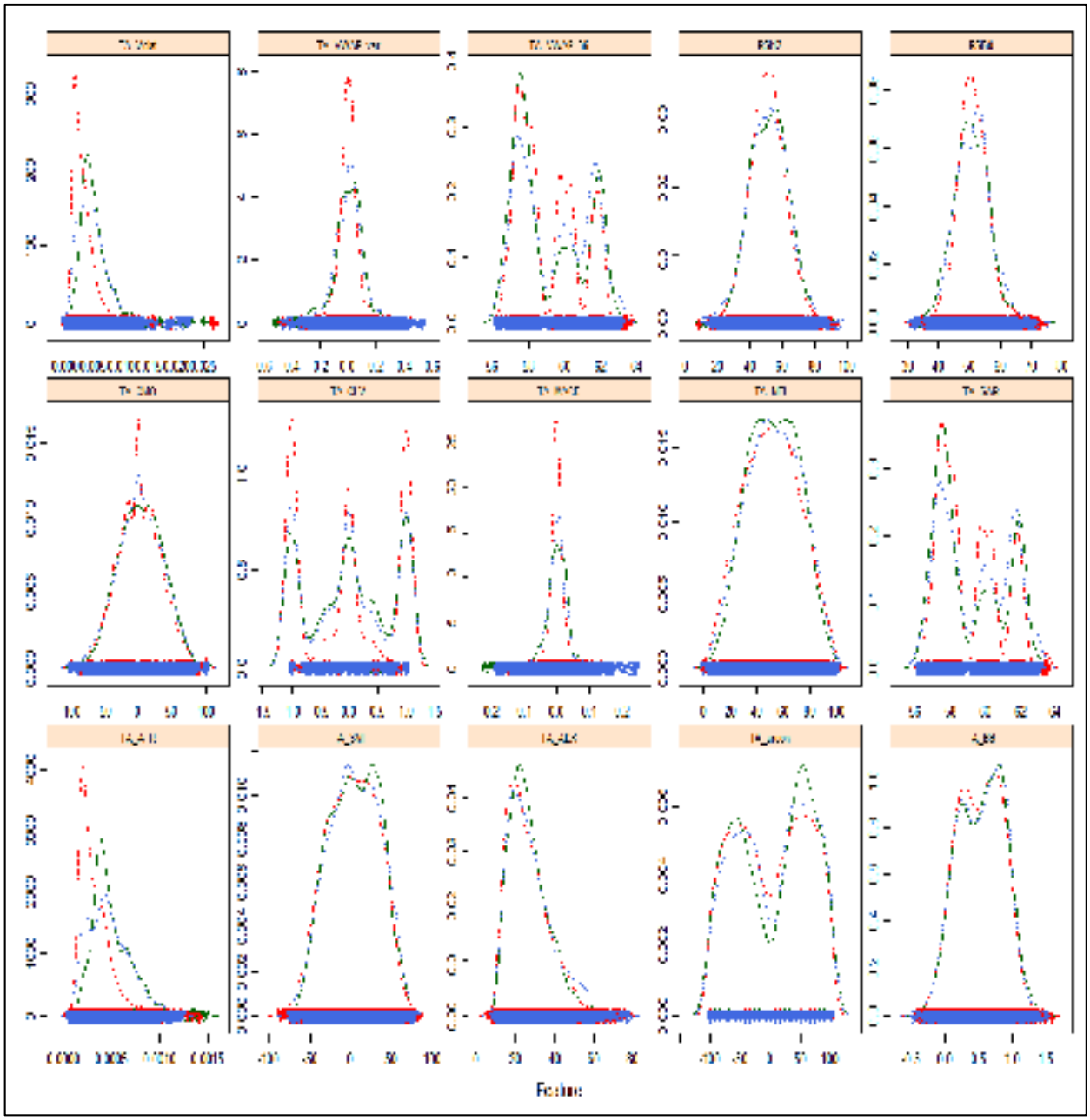


Figure A III.4 Densités des indicateurs technique par classe de la tendance

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

Abarbanell, J. S., & Bushee, B. J. (1997). Fundamental analysis, future earnings, and stock prices. *Journal of accounting research, 35*(1), 1-24.

Ahn, J., Jo, I., & Lee, J. (2002). The use of apple pomace in rice straw based diets of Korean native goats (Capra hircus). *Asian-Australasian journal of animal sciences, 15*(11), 1599-1605.

Bachelier, L. (1900). Théorie de la spéculation. Dans *Annales scientifiques de l'École normale supérieure* (Vol. 17, pp. 21-86).

Bollinger, G. (1981). Book Review: Regression Diagnostics: Identifying Influential Data and Sources of Collinearity: Sage Publications Sage CA: Los Angeles, CA.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of finance, 47*(5), 1731-1764.

Bruce L. Bowerman, R. T. O. C. (1993). *Forecasting and Time Series: An Applied Approach* South-Western College Pub; 3 edition (January 7, 1993).

Bühlmann, P. (2012). Bagging, boosting and ensemble methods. Dans *Handbook of Computational Statistics* (pp. 985-1022). Springer.

Cengiz, H., Bilen, Ö., Büyüklü, A. H., & Damgacı, G. (2017). Stock market anomalies: the day of the week effects, evidence from Borsa Istanbul. *Journal of Global Entrepreneurship Research, 7*(1), 4.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research, 16*, 321-357.

Cheung, J. K., Chung, R., & Kim, J. B. (1997). The profitability of trading strategies based on book value and earnings in Hong Kong: Market inefficiency vs. risk premia. *Journal of International Financial Management & Accounting, 8*(3), 204-233.

Chung, H. Y., & Kim, J.-B. (2001). A structured financial statement analysis and the direct prediction of stock prices in Korea. *Asia-Pacific Financial Markets, 8*(2), 87-117.

Dunsmuir, W., & Robinson, P. (1981). Estimation of time series models in the presence of missing data. *Journal of the American Statistical Association, 76*(375), 560-568.

Engelbrecht, A. P. (2007). *Computational intelligence: an introduction*. John Wiley & Sons.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Fama, E. F. (1995). Random walks in stock market prices. *Financial analysts journal, 51*(1), 75-80.

Fernandez-Rodrıguez, F., Gonzalez-Martel, C., & Sosvilla-Rivero, S. (2000). On the profitability of technical trading rules based on artificial neural networks:: Evidence from the Madrid stock market. *Economics letters, 69*(1), 89-94.

Frazzini, A., Kabiller, D., & Pedersen, L. H. (2013). *Buffett's alpha*. National Bureau of Economic Research.

Gençay, R., Dacorogna, M., Muller, U. A., Pictet, O., & Olsen, R. (2001). *An introduction to high-frequency finance*. Elsevier.

Glantz, M., & Kissell, R. L. (2013). *Multi-asset risk modeling: techniques for a global economy in an electronic and algorithmic trading era*. Academic Press.

Grothmann, R. (2003). *Multi Agent Market Modeling Based on Neutral Networks* (University of Bremen, Germany).

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton university press Princeton, NJ.

Haykin, S., & Network, N. (2004). A comprehensive foundation. *Neural networks, 2*(2004), 41.

Hellström, T. (1998). *A random walk through the stock market* (Univ.).

Hong, H., Kaplan, R. S., & Mandelker, G. (1978). Pooling vs. purchase: The effects of accounting for mergers on stock prices. *Accounting Review*, 31-47.

Jarrett, J. E., & Schilling, J. (2008). Daily variation and predicting stock market returns for the Frankfurter Börse (stock market). *Journal of Business Economics and Management, 9*(3), 189-198.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of finance, 45*(3), 881-898.

Kayaçetin, V., & Lekpek, S. (2016). Turn-of-the-month effect: New evidence from an emerging stock market. *Finance Research Letters, 18*, 142-157.

Khun, M. (2019). The caret Package. Repéré le 09/07/2019 à https://topepo.github.io/caret/

Koza, J. R., Bennett, F. H., Andre, D., & Keane, M. A. (1996). Automated design of both the topology and sizing of analog electrical circuits using genetic programming. Dans *Artificial Intelligence in Design'96* (pp. 151-170). Springer.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics, 54*(1-3), 159-178.

Lo, A. W., & MacKinlay, A. C. (2002). *A non-random walk down Wall Street*. Princeton University Press.

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of finance, 55*(4), 1705-1765.

Lui, Y.-H., & Mole, D. (1998). The use of fundamental and technical analyses by foreign exchange dealers: Hong Kong evidence. *Journal of international Money and Finance, 17*(3), 535-545.

McLeod, A. (1978). On the distribution of residual autocorrelations in Box–Jenkins models. *Journal of the Royal Statistical Society: Series B (Methodological), 40*(3), 296-302.

Miles, R. P. (2004). *Warren Buffett wealth: principles and practical methods used by the world's greatest investor*. John Wiley & Sons.

Mushtaq, R. (2011). Augmented Dickey Fuller Test. *SSRN Electronic Journal*.

Neely, C., & Weller, P. (2001). *Intraday technical trading in the foreign exchange market. Federal Reserve Bank of St*. Louis Working Paper 99-016B.

Perkins, N. J., & Schisterman, E. F. (2006). The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American journal of epidemiology, 163*(7), 670-675.

Piccoli, P., Chaudhury, M., & Souza, A. (2017). How do stocks react to extreme market events? Evidence from Brazil. *Research in International Business and Finance, 42*, 275-284.

Raina, R., Shen, Y., Mccallum, A., & Ng, A. Y. (2004). Classification with hybrid generative/discriminative models. Dans *Advances in neural information processing systems* (pp. 545-552).

Ray, B., Chen, S., & Jarrett, J. (1997). Identifying permanent and temporary components in daily and monthly Japanese stock prices. *Financial Engineering and the Japanese Markets, 4*(3), 233-256.

Roberts, H. (1967). StatisticalversusClinicalPredictionofthe Stock Market. *UnpublishedPaperPresentedtotheSeminarontheAnalysis ofSecurityPrices, UniversityofChicago*.

Rogalski, R. J., & Tinic, S. M. (1986). The January size effect: anomaly or risk mismeasurement? *Financial analysts journal, 42*(6), 63-70.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science, 290*(5500), 2323-2326.

Salzberg, S. L. (1994). C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. *Machine Learning, 16*(3), 235-240.

Samuelson, P. A. (2016). Proof that properly anticipated prices fluctuate randomly. Dans *The World Scientific Handbook of Futures Markets* (pp. 25-38). World Scientific.

Takens, F. (1981). Detecting strange attractors in turbulence. Dans *Dynamical systems and turbulence, Warwick 1980* (pp. 366-381). Springer.

Taylor, M. P., & Allen, H. (1992). The use of technical analysis in the foreign exchange market. *Journal of international Money and Finance, 11*(3), 304-314.

Touzet, C. (1992). *les réseaux de neurones artificiels, introduction au connexionnisme*.

TSAY, R. S. (2005). *Analysis of Financial Times Series, Second Edition*. Wiley-Interscience.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Zhou, C. (1996). *Stock market fluctuations and the term structure*. Citeseer.