

Rapport de Stage

L'intelligence artificielle et son application sur les marchés financiers

CHEN Yipeng



Figure 1: JY Asset

Stage : Quantitative Researcher - Machine Learning (Mars 2022 – Août 2022)



Figure 2: Université de Bordeaux

Master 2 IREF-FQA : Finance Quantitative et Actuariat (Année : 2021 – 2022)

Remerciements :

C'est avec un grand plaisir que je tiens à réserver ces lignes en signe de reconnaissance à tous ceux qui ont contribué de près ou de loin à l'élaboration de ce travail.

Au terme de ce travail, je tiens à remercier vivement l'ensemble du personnel de JY Asset, spécialement le personnel du département Quantitative Trading qui m'a accompagné tout au long de cette expérience professionnelle sans épargner aucun effort pour me guider sur la bonne voie du savoir.

Je tiens à ce titre à remercier M. HOU Yong, en tant que responsable du département Quantitative Trading de JY Asset et mon tuteur professionnel, pour le soutien et l'aide qu'il m'a réservé tout au long de la durée de stage et de la période d'élaboration de ce travail.

Je tiens à témoigner ma reconnaissance envers l'ensemble du corps professoral et administratif du Master IREF pour son dévouement et sa bienveillance, ainsi que pour sa veille sur la qualité de la formation.

Je tiens enfin à remercier tous mes collègues, camarades et amis ainsi que mes proches pour leur soutien et pour leur confiance et à qui je dédie fièrement ce travail.

Table des Matières :

Remerciements :	2
1. Résumé	4
2. Partie I : Introduction	5
2.1 Facteurs de choix du stage	5
2.2 Objectifs et attentes	5
2.3 L'entreprise et Missions du stagiaire Quantitative Researcher	6
2.4 Problématique du stage	6
3. Partie II : L'entreprise du secteur : Gestion d'actifs	7
3.1 Le secteur d'activité	7
3.1. A : La définition de l'Asset management (la Gestion d'actifs)	7
3.1. B : Les principes de l'Asset management	7
3.1. C : La demande des consommateurs dans l'Asset management	8
3.1. D : L'informatique dans l'Asset management	8
3.2 Présentation de l'entreprise	9
3.2. A : Historique et Activité de l'entreprise	9
3.2. B : Force et Faiblesse de l'entreprise	9
4. Partie III : Principales réalisations	10
4.1 Stock Picking + Portfolio Optimization	10
4.1.A : Objectif de la mission	10
4.1.B : Contenu de la mission	11
4.1.C : Outils employés	23
4.1.D : Difficultés rencontrées et Solutions apportées	24
4.2 High Frequency Trading Research	24
4.2.A : Objectif de la mission	24
4.2.B : Contenu de la mission	24
4.2.C : Outils employés	28
4.2.D : Difficulté rencontrée et Solution apportée	28
5. Partie IV : Conclusion	28

1. Résumé

Vers la fin de février 2022, j'ai terminé tous mes cours du master 2 en ingénierie des risques économiques et financiers - parcours finance quantitative et actuariat.

Afin d'approfondir mes connaissances dans le domaine de la finance quantitative et de la programmation, j'ai décidé de trouver un stage lié au métier d'Analyste quantitatif - Quant dans le secteur d'asset management en Chine. J'ai donc fini par me joindre à l'entreprise JY Asset au début du mois de mars 2022 en tant que Stagiaire Quantitative Researcher spécialisé en Machine Learning au sein du département de Quantitative Trading. Pendant mon stage, j'ai pu mieux comprendre le quotidien d'un Quant en faisant des projets professionnels, notamment à travers 2 grands projets utilisant 1 Mesure Robuste de Co-Mouvement : GS¹, 2 Tests Statistiques : Test LS² & Test D'Agostino³, 2 Modèles de Machine Learning : Bernoulli Naïve Bayes⁴ & SVM⁵ et 1 Modèle de Deep Learning : LSTM⁶.

Concernant le premier projet, j'ai finalement réussi à construire un modèle automatique sur RStudio et Python dans le but de sélectionner des actions chinoises qui seront potentiellement rentables dans le futur, déterminer un poids optimal⁷ pour chaque action choisie et évaluer la performance du portefeuille constitué uniquement des actions présélectionnées par le modèle. A propos du deuxième projet, j'ai fini par réaliser une stratégie profitable du trading haute fréquence⁸ à l'aide de Python sur le marché chinois de commodity futures⁹.

¹Gerber Statistic : Nouvelle Statistique introduite par M. Gerber et M. Markowitz dans leur nouvelle thèse «The Gerber statistic : a robust co-movement measure for portfolio optimization».

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3880054

²Lee-Strazicich Unit Root Test ou Test Lee-Strazicich de Racine Unitaire.

³Test de l'Asymétrie d'une Distribution.

⁴Classifieur Bayésien Naïf suivant la loi de Bernoulli.

⁵Support Vector Machine ou Machine à Vecteurs de Support.

⁶Long Short Term Memory ou les réseaux de Longue Mémoire à Court Terme.

⁷Poids Optimal : La Meilleure Proportion du montant de richesse investi dans l'action.

⁸Le Trading Haute Fréquence (THF) est l'une des catégories du trading algorithmique.

⁹Commodity Futures : Contrat à Terme sur des Matières Premières.

2. Partie I : Introduction

2.1 Facteurs de choix du stage

Pendant mes deux ans d'études dans le master IREF, je me suis rendu compte que j'aimais beaucoup la programmation sur Python, les modèles de Machine Learning, l'économétrie et la gestion de portefeuille. Après avoir fait mes propres recherches sur Internet, j'ai enfin trouvé un métier qui combine presque tous les éléments que je préfère. C'est donc le métier de Quant (Analyste quantitatif) qui me convient le plus.

Un Quant ou Analyste quantitatif, est un professionnel de la finance qui développe et adapte des modèles mathématiques pour évaluer la rentabilité des titres financiers, prévoir les mouvements de marché et pour se protéger contre les risques. En particulier, ce sont les missions de Front office Quant¹⁰ et de Research Quant¹¹ qui m'intéressent vraiment. Parallèlement, toute la théorie présentée en gestion de portefeuille, notamment la partie sur le modèle Moyenne-Variance¹², est également très attractive pour moi.

C'est pour cette raison que j'ai décidé de faire mon stage chez JY Asset, dans le département de Quantitative Trading, afin de découvrir ce que font les Front office Quant et Research Quant. Puisque j'ai fait une candidature spontanée, après avoir négocié avec mon tuteur de stage (M. HOU Yong), j'ai pu définir mes propres missions. En dernier lieu, j'ai choisi de prendre les missions d'un Quantitative Researcher spécialisé en Machine Learning comme missions principales vu que toute l'équipe de JY Asset était en train de trouver des modèles de Machine Learning convenables pour sélectionner les titres financiers probablement rentables dans le futur afin de construire une stratégie véritablement profitable de trading.

2.2 Objectifs et attentes

Au cours de ce stage, mon objectif principal est de comprendre l'ensemble du processus de sélection des titres financiers, comment une stratégie du trading algorithmique¹³ est construite dans une entreprise réelle. Et j'espère utiliser mes connaissances pour mieux aider l'entreprise à trouver des nouveaux modèles ou améliorer les modèles déjà existants dans le but de proposer des prédictions plus fiables sur les rentabilités des titres financiers à la fin du stage.

En outre, j'espère également que ce stage me permettra de renforcer mes compétences en Machine Learning et ensuite d'avoir une vision globale et une compréhension précise sur ce que font exactement les Quantitative Researchers dans le domaine de la finance.

¹⁰Front office Quant : il conçoit des modèles directement utilisés par les traders.

¹¹Research Quant : il cherche de nouveaux modèles et de nouvelles méthodologies pour le front office.

¹²Modèle Mean – Variance de Markowitz (1952) : la base de théorie moderne de choix de portefeuille.

¹³Le trading algorithmique est une méthode qui permet d'automatiser, en partie ou totalement, son trading.

2.3 L'entreprise et Missions du stagiaire Quantitative Researcher

JY Asset est une société de gestion d'actifs basée dans la région Asie-Pacifique et agréée en tant que gestionnaire de fonds par AMAC¹⁴. JY a établi des relations solides avec de nombreuses institutions financières de renom.

Au cours de mon stage, l'activité que j'ai analysée était principalement axée sur la prédiction des rentabilités futures de titres financiers (notamment les Actions A chinoises¹⁵).

Ma tâche principale consistait à concevoir et à développer des stratégies algorithmiques basées sur les modèles de Machine Learning les plus populaires¹⁶ afin de sélectionner des titres financiers probablement rentables dans le futur.

2.4 Problématique du stage

Tout mon stage avait pour but de mieux comprendre l'application de l'intelligence artificielle à la finance et de répondre notamment aux questions suivantes :

1. Comment choisir les meilleures actions ?

2. Quelle est la somme idéale à investir ?

3. Quel est le bon moment pour acheter ou vendre ?

¹⁴L'Asset Management Association of China (AMAC) est une association d'autorégulation des sociétés de gestion de fonds en Chine.

¹⁵Les « Actions A chinoises » sont négociées en RMB (RenMinBi : le nom officiel de la monnaie chinoise) sur les places boursières de Shanghai et de Shenzhen par des sociétés constituées en Chine continentale. Elles ne peuvent être achetées que par des investisseurs domestiques chinois et les acheteurs institutionnels étrangers qualifiés. Les « Actions B chinoises » sont cotées en devises (en USD par exemple) sur les bourses de Shanghai et de Shenzhen et sont ouvertes aux investisseurs domestiques comme étrangers.

¹⁶Surtout "Support Vector Machines", "Bernoulli Naïve Bayes" et "Long Short Term Memory".

3. Partie II : L'entreprise du secteur : Gestion d'actifs

3.1 Le secteur d'activité

3.1. A : La définition de l'Asset management (la Gestion d'actifs)

L'asset management est une activité financière qui a pour objectif de créer, gérer, faire fructifier et maximiser les bénéfices de produits ou placements financiers confiés par des entreprises ou des investisseurs particuliers. Aussi appelé gestion d'actifs en français, l'asset management consiste à gérer un portefeuille client et à augmenter sa rentabilité en équilibrant les coûts et les risques afin d'atteindre les objectifs préalablement définis.

Le secteur de la gestion d'actifs, dont l'asset management fait partie, a pour mission principale de faire prospérer le patrimoine d'un client par le biais d'investissements financiers. Basé sur quatre grands principes, qui sont la valeur, le leadership, l'alignement et la garantie, l'asset management concerne la gestion de produits et placements financiers dans le but de faire des investissements rentables.

3.1. B : Les principes de l'Asset management

Via les mouvements des taux d'intérêt et la valeur des positions en portefeuille d'actions, les opportunités varient, tout comme les risques. Afin de faire un bon placement, l'asset management se fonde sur quelques principes clefs :

L'alignement : aligner un portefeuille financier signifie traduire les objectifs en décisions organisationnelles et techniques afin de, par la suite, générer des bénéfices.

La garantie : en asset management, ce principe est fondamental car il concerne différentes étapes (les actifs, leur gestion et le système de gestion) qui conduisent au bon pilotage de la gestion de portefeuilles financiers.

Le leadership : pour créer de la valeur, le leadership est indispensable car il traduit l'engagement de l'entreprise pour mettre en oeuvre une gestion d'actifs efficace et prospère.

La valeur : il s'agit d'une donnée liée à l'actif et qui est déterminante pour estimer les profits que l'investissement permet de réaliser.

3.1. C : La demande des consommateurs dans l'Asset management

Pour la Chine, la construction d'infrastructures et l'immobilier ont précédemment été les principaux moteurs de la croissance économique pendant une dizaine d'années. Mais au cours de ces dernières années, dans le contexte de la maturation du marché financier national, de la politique de "non spéculation immobilière" et de la variété croissante des outils et produits d'investissement financiers, les résidents nationaux accordent de plus en plus d'attention au rôle des titres financiers dans l'allocation d'actifs.

Le fait que la Chine soit devenue le premier pays à maîtriser l'épidémie de covid-19 a donné une grande confiance aux investisseurs domestiques. L'excellente performance du marché des Actions A chinoises a favorisé l'augmentation de l'échelle de gestion de fonds.

En outre, parallèlement au transfert progressif de la richesse des résidents nationaux de l'immobilier vers le marché financier national, l'afflux de capitaux privés provenant d'investisseurs domestiques s'est considérablement accéléré depuis 2020.

3.1. D : L'informatique dans l'Asset management

L'informatique joue désormais un rôle majeur dans l'asset management car les outils numériques permettent une meilleure maîtrise des portefeuilles de clients et un meilleur suivi des risques. L'automatisation de l'analyse de données est donc devenue un atout stratégique majeur pour gérer des actifs.

Limiter l'intervention humaine en automatisant certaines tâches permet, alors, de mieux suivre les données et de dégager du temps à l'humain pour plus de tâches à valeur ajoutée. Les outils numériques offrent plus de visibilité sur le marché et, de ce fait, favorisent un meilleur rendement.

Pour répondre toujours plus vite aux attentes des clients, l'asset management s'est renouvelé en intégrant avec succès l'informatique à ses services.

3.2 Présentation de l'entreprise

3.2. A : Historique et Activité de l'entreprise

En 2016, une équipe d'investissement d'élites a été formée. En 2017, cette équipe a officiellement lancé ses propres stratégies neutres au marché¹⁷. Ensuite, elle a créé son image de marque sous le nom de JY Asset. Entre 2018 et 2021, le nombre de fonds gérés par JY Asset en Chine est passé à 10. Et l'équipe de JY Asset a innové davantage de stratégies de trading. Le total des actifs sous gestion de JY Asset, tant au niveau national qu'international, s'élève désormais à 700 millions de dollars. Aujourd'hui, JY Asset est une société internationale de gestion d'actifs basée dans la région Asie-Pacifique. Elle vise à fournir aux investisseurs des services d'investissement qui maximisent leurs bénéfices à long terme. Elle utilise des technologies récentes pour analyser la volatilité et l'inefficacité des marchés financiers afin de capturer un taux de rendement¹⁸ positif.

3.2. B : Force et Faiblesse de l'entreprise

JY Asset possède des membres assez divers. L'équipe centrale de JY est constituée de talents internationaux dans le domaine financier. Les membres de l'équipe ont des années d'expérience dans le domaine de gestion de fonds. La plupart d'entre eux ont obtenu CFA, FRM¹⁹ et sont familiers avec la philosophie d'investissement sur le marché financier international. Ils ont une forte connectivité avec l'industrie, une large capacité de connaissance et un esprit innovant. L'équipe R&D de JY a plus de 20 ans d'expérience et se spécialise dans le modèle stratégique de trading quantitatif sur le marché de l'Asie-Pacifique. Cependant, l'entreprise a très peu de Data Scientists, ce qui force l'équipe de Quants à acheter ailleurs des données fiables en payant très cher. Sinon, on est obligé d'aider l'équipe de Data de petite taille à collecter des données en sacrifiant une partie du temps de la construction des nouvelles stratégies. Mais ainsi, la qualité des données n'est pas suffisamment vérifiée par des experts dans le domaine de la science des données. Par conséquent, les stratégies de trading basées sur ces données "non fiables" risquent de conduire à une rentabilité positive pendant le Backtesting²⁰ mais une perte d'argent lors de trading réel.

¹⁷ Les stratégies neutres au marché visent à procurer des rendements absolus positifs, peu importe la direction du marché sous-jacent. Les portefeuilles qui en résultent sont habituellement très peu exposés à des formes de risque particulières, notamment au "risque de marché", soit l'effet du mouvement de l'ensemble du marché sur un placement.

¹⁸ Le taux de rendement (ou ROR pour Rate Of Return) est la perte ou le profit d'un placement sur une certaine période, exprimé en pourcentage. ROR mesure le rendement par rapport au coût initial de l'investissement. Un ROR positif signifie que la position a généré un profit, tandis qu'un ROR négatif signifie une perte.

¹⁹ La certification CFA est un moyen d'intégrer le monde de la finance de la manière la plus efficace possible tout en développant les compétences indispensables ; La certification FRM est spécialisée sur la gestion et modélisation des risques.

²⁰ Le Backtesting est un ensemble de procédures statistiques consistant à tester une stratégie financière à partir de données historiques. Ce genre de simulation permet d'affiner un modèle et d'isoler les hypothèses.

4. Partie III : Principales réalisations

Au cours de mon stage de 6 mois, j'ai réalisé deux projets importants. Le premier projet est beaucoup plus long et compliqué que le deuxième. Mais ces deux projets ensemble me permettent de mieux comprendre ce que c'est un vrai Quantitative Researcher en Asset Management.

4.1 Stock Picking + Portfolio Optimization

4.1.A : Objectif de la mission

Le premier projet a pour but : la sélection des Actions A chinoises probablement rentables dans le futur et l'optimisation de portefeuille constitué de ces actions dites "probablement rentables".

Autrement dit, ce projet tente de répondre aux deux premières questions précédemment mentionnées :

1. Comment choisir les meilleures actions ?

2. Quelle est la somme idéale à investir ?

Concrètement, il consiste à utiliser toutes mes connaissances en économétrie, machine learning et gestion de portefeuille afin de fournir une liste d'actions sur laquelle un poids optimal est distribué à chaque Action A chinoise sélectionnée et de montrer les performances de portefeuille final.

4.1.B : Contenu de la mission

Responsabilité :

Je suis censé construire indépendamment un algorithme qui sert à appliquer les tests statistiques et les modèles d'apprentissage automatique déjà existants mais les plus adéquats possibles à la prédiction des rentabilités des actions, et ensuite à former un portefeuille idéal avec un poids optimal indiqué pour chaque action prise dans le portefeuille. Finalement, je dois rédiger un rapport sur ce projet en détaillant sur le principe des tests statistiques et modèles d'apprentissage automatique utilisés, l'efficacité de l'algorithme construit et enfin les indicateurs de performance choisis pour évaluer le portefeuille final.

Présentation de la base de données :

La base de données initiale prétraitée par l'équipe de Data n'a pas de valeurs manquantes. Et elle ne contient que le prix de clôture ajusté des 209 actions différentes sur la période du 04/01/2007 au 08/02/2022.

En ligne, ce sont les dates de chaque jour ouvré à la Bourse entre 01/2007 et 02/2022. En colonne, ce sont les symboles de toutes les 209 actions différentes.

	SH000001	SH000002	SH000016	...	SZ399001	SZ399106	SZ399107
Date				...			
2007/1/4	2715.72	2857.58	1836.14	...	6705.34	553.69	572.48
2007/1/5	2641.33	2778.46	1815.12	...	6706.24	562.23	581.65
2007/1/8	2707.20	2847.60	1856.62	...	6870.50	579.22	599.49
2007/1/9	2807.80	2953.92	1932.28	...	7078.59	594.16	615.05
2007/1/10	2825.58	2972.32	1963.63	...	7346.02	610.42	632.61
...
2022/1/26	3455.67	3621.42	3163.67	...	13780.30	2329.17	2437.13
2022/1/27	3394.25	3557.00	3119.24	...	13398.85	2262.41	2367.20
2022/1/28	3361.44	3522.48	3054.02	...	13328.06	2262.37	2367.16
2022/2/7	3429.58	3594.01	3114.26	...	13456.65	2285.99	2391.90
2022/2/8	3452.63	3618.25	3116.51	...	13325.41	2280.51	2386.15

[3670 rows x 209 columns]

Figure 3: Prix de clôture

Tâches effectuées :

Notations essentielles :

p_t : le cours de l'action à la date t qui est un jour ouvré à la Bourse

r_t : le rendement logarithmique = $\log(p_t) - \log(p_{t-1})$

a. Première sélection basée sur la Stationnarité et l'Asymétrie

²¹{Une propriété importante de la série chronologique est sa stationnarité. Si un processus est stationnaire, cela signifie que ses propriétés statistiques ne varient pas dans le temps, à savoir sa moyenne, sa variance (homoscédasticité) ou encore sa covariance.

Alors pourquoi la stationnarité est-elle si importante ? Parce qu'il est plus aisé de faire des prédictions sur une série s'il est possible de supposer que les futures propriétés statistiques ne seront pas différentes de celles actuellement observées. La plupart des modèles de prévisions de séries chronologiques, d'une manière ou d'une autre, s'appuient sur ces propriétés (moyenne ou variance, par exemple). Ces pronostics seraient faux si la série originale n'était pas stationnaire.}

En partant du principe mentionné au-dessus, j'ai donc décidé de commencer par réaliser le Test LS vu en cours d'économétrie de Madame Lebreton pour trouver des actions ayant des séries temporelles r_t qui suivent un processus stationnaire.

Test de Lee et Strazicich (Test LS) :

Le Test LS est performant en taille et en puissance que si une seule date de rupture est supposée. Par conséquent, j'ai seulement considéré un changement structurel avec modèle "crash"²² ici.

H_0 : Présence de racine unitaire avec 1 changement structurel (Hypothèse nulle)

H_a : Absence de racine unitaire avec 1 changement structurel (Hypothèse alternative)

Critère de sélection :

$sval$: valeur de la statistique du test

$cval$: valeur critique au seuil de risque à 5%

$my\ lag$: le nombre de lag initialement fixé à 5

$op\ lag$: le nombre optimal de lag proposé par la procédure basée sur le BIC²³

Si $my\ lag = op\ lag$ et $sval < cval$:

Alors H_0 sera rejetée et la stationnarité de la série r_t sera acceptée.

Par conséquent, l'action qui possède ce type de série temporelle r_t suivant un processus stationnaire sera retenue.

²¹«MIEUX COMPRENDRE LES MÉTHODES DE PRÉVISION DES SÉRIES CHRONOLOGIQUES»
<https://www.aquiladata.fr/insights/mieux-comprendre-les-methodes-de-prevision-des-series-chronologiques/>

²² DU_t : la variable à ajouter pour modéliser un changement dans le niveau de la partie déterministe de la série temporelle r_t

²³Bayesian Information Criterion : critère d'information bayésien

Après avoir réalisé le Test LS sur l'ensemble des actions dans la base de données, j'ai pu supprimer des actions dites "non stationnaires" dans la liste initiale des 209 actions.

Ici, seulement le résultat du Test LS de la 209ème action est montré :

```
[1] 209
[1] -25.48998
[1] "First possible structural break at position: 1791"
[1] "The location of the first break - lambda_1: 0.5 , with the number of total observations: 3669"
Critical values - Crash model:
      1%      5%     10%
[1,] -4.239 -3.566 -3.211
[1] "Number of lags determined by general-to-specific lag selection: 3"
```

Figure 4: Test LS de la 209ème action

"Number of lags determined by general-to-specific lag selection: 3" $\Rightarrow op\ lag = 3$

Mais $my\ lag = 5 \Rightarrow my\ lag \neq op\ lag$,

donc la 209ème action ne sera pas retenue.

Time difference of 44.77066 mins

Figure 5: Temps passé sur le Test LS

J'ai consacré à peu près 45 minutes à faire le Test LS pour chaque action parmi les 209 actions initiales.

Actions		Actions	
1	SH000001 : OK	204	SZ002083 : OK
2	SH000002 : OK	205	SZ002087 : OK
3	SH000016 : NG	206	SZ002095 : NG
4	SH000300 : NG	207	SZ399001 : NG
5	SH600004 : NG	208	SZ399106 : NG
6	SH600007 : NG	209	SZ399107 : NG

Figure 6: Classification des actions basée sur la stationnarité

OK : Stationnarité Présente

NS : Non Stationnarité

NG : Not Good \Rightarrow Aucun Résultat pour le Test LS

Ensuite, j'ai gardé uniquement des actions avec l'étiquette "OK" :

	new_df		new_df
1	SH000001 : OK	112	SZ002038 : OK
2	SH000002 : OK	113	SZ002039 : OK
3	SH600011 : OK	114	SZ002056 : OK
4	SH600012 : OK	115	SZ002066 : OK
5	SH600015 : OK	116	SZ002083 : OK
6	SH600016 : OK	117	SZ002087 : OK

Figure 7: Actions "stationnaires"

Il reste donc 117 actions dans le portefeuille.

Le Skewness mesure l'asymétrie d'une distribution. Une valeur positive du coefficient d'asymétrie révèle une distribution plus étendue vers les valeurs positives, alors qu'une valeur négative met en évidence une distribution plus étendue vers les valeurs négatives. Pour un investissement, le Skewness permet de relever les extrêmes d'une distribution des rendements : elle permet soit de désigner des cas de pertes extrêmes (Skewness négative élevée), soit des cas de gains extrêmes (Skewness positive élevée).

Lorsque le Skewness est statistiquement nul, la distribution des rendements est symétrique, ce qui veut dire que la probabilité de gains est égale à la probabilité de pertes.

Afin d'exploiter plus efficacement les opportunités de gains potentiels découlant de l'asymétrie de la distribution des rendements, j'ai choisi le Test D'Agostino dans le but d'enlever toutes les actions qui ont une distribution symétrique des rendements.

Test de l'Asymétrie (Test D'Agostino) :

$H_0 : Skewness = 0$ [symétrie]

$H_a : Skewness \neq 0$ [asymétrie]

Critère de sélection :

$p - value$: probabilité de rejeter à TORT l'hypothèse nulle

Si $p - value < 0.05$:

Alors H_0 sera rejetée et l'asymétrie de la distribution des rendements r_t sera acceptée.

Par conséquent, l'action qui possède ce type de distribution des rendements r_t asymétrique sera retenue.

Après avoir réalisé le Test D'Agostino sur l'ensemble des actions dans le portefeuille, j'ai pu enlever toutes les actions dites "symétriques" parmi les 117 actions restantes.

Ici, seulement le résultat du Test D'Agostino de la 117ème action est montré :

```
[1] 117

      D'Agostino skewness test

data: 1rt
skew = -0.48164, z = -11.33582, p-value < 2.2e-16
alternative hypothesis: data have a skewness
```

Figure 8: Test D'Agostino de la 117ème action

$p - value < 2.2e - 16 \Rightarrow p - value < 0.05$

\Rightarrow l'asymétrie de la distribution des rendements r_t est présente dans cette action.

Donc, la 117ème action sera retenue.

Actions		Actions	
1	SH000001 : OK	112	SZ002038 : NG
2	SH000002 : OK	113	SZ002039 : OK
3	SH600011 : OK	114	SZ002056 : OK
4	SH600012 : OK	115	SZ002066 : OK
5	SH600015 : NG	116	SZ002083 : OK
6	SH600016 : NG	117	SZ002087 : OK

Figure 9: Classification des actions basée sur l'asymétrie

OK : \Rightarrow La distribution est asymétrique

NG : Not Good \Rightarrow La distribution est symétrique

Ensuite, j'ai gardé uniquement des actions avec l'étiquette "OK" :

new_ddf		new_ddf	
1	SH000001 : OK	84	SZ002023 : OK
2	SH000002 : OK	85	SZ002039 : OK
3	SH600011 : OK	86	SZ002056 : OK
4	SH600012 : OK	87	SZ002066 : OK
5	SH600020 : OK	88	SZ002083 : OK
6	SH600031 : OK	89	SZ002087 : OK

Figure 10: Actions "asymétriques"

Il reste donc 89 actions dans le portefeuille.

En résumé, les actions qui possèdent les caractéristiques de la stationnarité et de l'asymétrie sont les suivantes :

Actions choisies (stationnarité + asymétrie)			Actions choisies (stationnarité + asymétrie)		
1		SH000001	84		SZ002023
2		SH000002	85		SZ002039
3		SH600011	86		SZ002056
4		SH600012	87		SZ002066
5		SH600020	88		SZ002083
6		SH600031	89		SZ002087

Figure 11: Actions "stationnaires + asymétriques"

Cependant, dans cette liste des actions,
les deux premières actions sont en fait des indices de marché.
Il est nécessaire de les supprimer :

```

code
0  SH600011
1  SH600012
2  SH600020
3  SH600031
4  SH600033
..  ...
82  SZ002039
83  SZ002056
84  SZ002066
85  SZ002083
86  SZ002087

[87 rows x 1 columns]
```

Figure 12: Actions "stationnaires + asymétriques"

Il reste donc 87 actions dans le portefeuille.

Nouvelles bases de données:

Puisque les prix des actions sont déjà utilisés pour faire la première sélection, maintenant c'est mieux d'utiliser des données différentes.

En me basant sur le cours d'analyse financière²⁴, j'ai finalement pris le Tableau de Flux de Trésorerie (entre 2007 et 2021) de chaque entreprise concernée comme les inputs du modèle de Machine Learning.

Remarquons qu'avant nous avions des données journalières, mais les données comptables sont trimestrielles, il nous faut donc calculer les rentabilités (toujours logarithmiques) trimestrielles en utilisant seulement les prix à la fin des mois de mars, juin, septembre et décembre vu que les données comptables sont sorties vers la fin de ces mois.

²⁴L'analyse financière consiste à tirer des renseignements sur la santé d'une entreprise, notamment en matière de solvabilité et de rentabilité, à partir des documents comptables.

Ensuite, après avoir calculé les rentabilités trimestrielles, j'ai fait une classification sur toutes les rentabilités (*return*) :

Si return ≥ 0 : La rentabilité est du type "1" ;

Si return < 0 : La rentabilité est du type "0" ;

Par exemple, pour la première action "SH600011", nous avons :

	Date	Mois	SH600011	销售商品、提供劳务收到的现金
0	2007-03	3	25.74	1.342139e+10
1	2007-06	6	31.67	2.748957e+10
2	2007-09	9	46.72	4.176175e+10
3	2007-12	12	43.99	5.735953e+10
4	2008-03	3	27.17	1.576796e+10
	Date	Mois	SH600011	销售商品、提供劳务收到的现金
55	2020-12	12	20.61	1.836914e+11
56	2021-03	3	21.75	5.314348e+10
57	2021-06	6	21.44	1.042392e+11
58	2021-09	9	27.68	1.622004e+11
59	2021-12	12	33.63	2.270959e+11

Figure 13: Action SH600011

Dans la colonne "SH600011", nous avons des informations sur les prix trimestriels de cette action. Dans la colonne en Chinois, nous avons des informations comptables liées à cette action. Toutes ces informations sont trimestrielles (entre 2007 et 2021, soit une durée de 60 trimestres).

Ici, seulement les 4 premières colonnes sont montrées.

En fait, il y a 62 colonnes au total pour l'action "SH600011".

[60 rows x 62 columns]

Figure 14: Size pour l'action "SH600011"

Les 3 premières colonnes représentent respectivement la date, le mois et le prix trimestriel de l'action. Et les colonnes restantes contiennent toutes les informations comptables dans le Tableau de Flux de Trésorerie de l'entreprise associée.

En supprimant les 3 premières colonnes et en ajoutant 2 nouvelles colonnes “Returns” et “Label” :

	销售商品、提供劳务收到的现金	收到的税费返还	...	Returns	Label
1	1.342139e+10	0.0	...	0.207324	1
2	2.748957e+10	0.0	...	0.388802	1
3	4.176175e+10	0.0	...	-0.060210	0
4	5.735953e+10	0.0	...	-0.481849	0
5	1.576796e+10	0.0	...	-0.186822	0
	销售商品、提供劳务收到的现金	收到的税费返还	...	Returns	Label
55	1.298559e+11	162806198.0	...	-0.160576	0
56	1.836914e+11	250260763.0	...	0.053837	1
57	5.314348e+10	27017375.0	...	-0.014355	0
58	1.042392e+11	123704642.0	...	0.255452	1
59	1.622004e+11	240397028.0	...	0.194708	1

Figure 15: Action SH600011

Nous avons une telle base de données pour chaque action parmi les 87 actions dans le portefeuille.

b. Deuxième sélection basée sur le Classifieur Bayésien Naïf (Loi de Bernoulli)

La classification naïve bayésienne est un type de classification Bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte indépendance (dite naïve) des hypothèses. Elle met en oeuvre un classifieur bayésien naïf, ou classifieur naïf de Bayes, appartenant à la famille des classifieurs linéaires.

La méthode de classification naïve bayésienne est un algorithme d'apprentissage supervisé (supervised machine learning) qui permet de classifier un ensemble d'observations selon des règles déterminées par l'algorithme lui-même. Cet outil de classification doit dans un premier temps être entraîné sur un jeu de données d'apprentissage qui montre la classe attendue en fonction des entrées. Pendant la phase d'apprentissage, l'algorithme élabore ses règles de classification sur ce jeu de données, pour les appliquer dans un second temps à la classification d'un jeu de données de prédiction. Le classificateur bayésien naïf implique que les classes du jeu de données d'apprentissage soient connues et fournies, d'où le caractère supervisé de l'outil.

Ici, il est favorable de choisir la loi de Bernoulli puisque l'on a transformé notre question de prédiction sur les rentabilités en une question de classification binaire sur les rentabilités.

Avant de commencer à entraîner le modèle de Bernoulli Naïve Bayes, il est nécessaire de transformer nos données comptables (quantitatives) qui varient sur une échelle assez large. Finalement, en restant sur la façon binaire, j'ai fait la binarisation²⁵ sur toutes les données comptables en prenant une valeur de seuil égale à 0.

²⁵Toutes les valeurs au-dessus du seuil sont marquées comme 1 et toutes les valeurs qui sont égales ou inférieures au seuil sont marquées comme 0.

En mettant les informations comptables comme les inputs du modèle et les types de rentabilités ("1" ou "0") comme les valeurs à prédire, j'ai utilisé ce modèle de Bernoulli Naïve Bayes afin de faire la classifications des rentabilités futures pour chaque action.

Ensuite, pour évaluer la performance du modèle dans chaque action, j'ai calculé *Accuracy* et *Precision*.

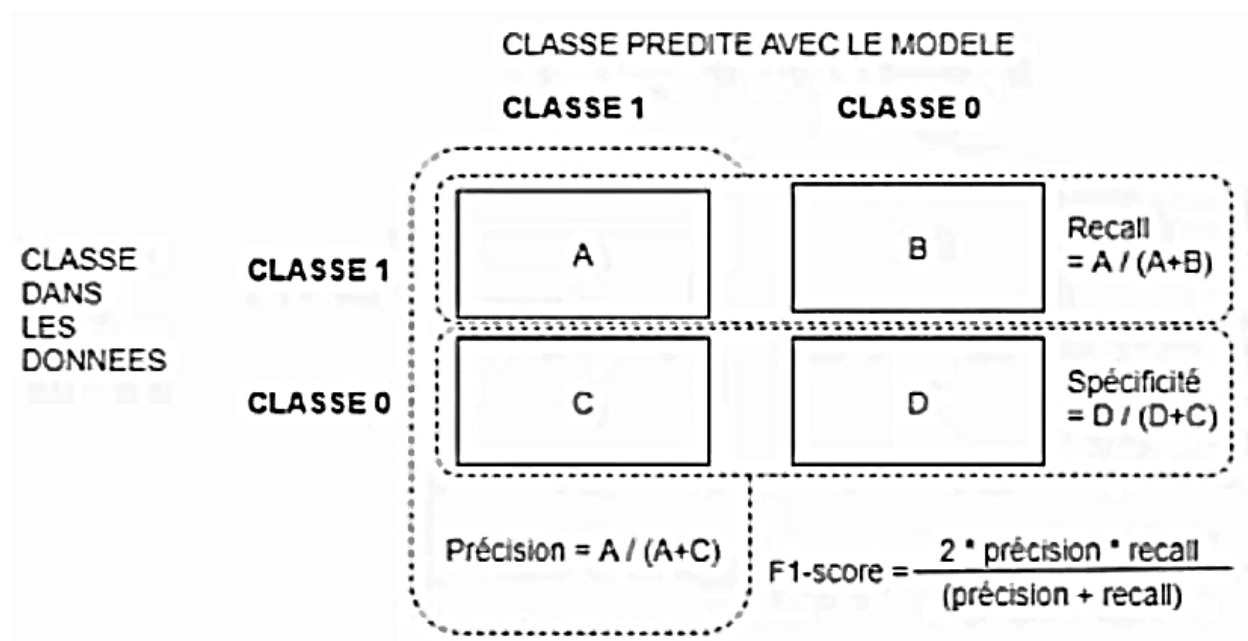


Figure 16: Matrice de confusion

$$Accuracy = (A + D) / (A + B + C + D)$$

Critère de sélection :

Si $Accuracy \geq 0.5$ et $Precision \geq 0.5$:

Alors l'action concernée sera retenue.

Règles de scoring :

Brièvement, dans les actions restantes ayant $Accuracy$ et $Precision \geq 0.5$, j'ai utilisé l'ensemble de données entre 2007 et 2018 pour entraîner le modèle dans le but de prédire les types de rentabilités pour les 12 trimestres suivants (soit 36 mois = 3 ans) {entre 2019 et 2021}.

Pour un trimestre qui va venir prochainement, si la classification de la rentabilité de l'action indique que la rentabilité est du type "1" (rentabilité positive ou nulle), alors cette action va obtenir 1 point.

Pour un trimestre qui va venir prochainement, si la classification de la rentabilité de l'action indique que la rentabilité est du type "0" (rentabilité négative), alors cette action va obtenir 0 point.

Score = la somme de tous les points obtenus pendant les 12 trimestres suivants.

Par exemple, si pendant les 12 trimestres suivants l'action continue à obtenir 1 point sans arrêt, finalement cette action va avoir un score de 12 points au total qui est évidemment le score maximal et le meilleur score.

Après, j'ai utilisé ces règles de scoring afin de calculer un score pour chaque action restante.

Enfin, j'ai ordonné de manière décroissante toutes les actions restantes en fonction de leur *Score*, *Precision* et *Accuracy* :

	code	name	industry	Accuracy	Precision	Score
0	SH600563	法拉电子	元器件	0.92	0.91	11
1	SH600809	山西汾酒	白酒	0.67	0.75	8
2	SH600060	海信视像	家用电器	0.58	0.62	8
3	SH600426	华鲁恒升	农药化肥	0.5	0.62	8
4	SH600067	冠城大通	电气设备	0.5	0.5	8
	code	name	industry	Accuracy	Precision	Score
37	SH600302	标准股份	纺织机械	0.5	0.5	4
38	SZ000957	中通客车	汽车整车	0.58	1.0	3
39	SH600159	大龙地产	区域地产	0.58	1.0	3
40	SH600261	阳光照明	家用电器	0.5	1.0	3
41	SH600883	博闻科技	种植业	0.5	0.67	3

Figure 17: Actions ordonnées

Il reste encore 42 actions.

Or, pour atteindre les niveaux raisonnables de diversification, il faut au moins 25 titres²⁶.

Par conséquent, j'ai simplement pris les 25 premières actions qui sont meilleures que les autres actions en termes de *Score*, *Precision* et *Accuracy* afin de constituer le portefeuille final.

²⁶“Reasonable levels of diversification require as little as 25 securities” : chapitre 1 «Efficient diversification» page 30 (cours de Gestion de portefeuille de madame Boussetta.)

	code	name	industry	Accuracy	Precision	Score
0	SH600563	法拉电子	元器件	0.92	0.91	11
1	SH600809	山西汾酒	白酒	0.67	0.75	8
2	SH600060	海信视像	家用电器	0.58	0.62	8
3	SH600426	华鲁恒升	农药化肥	0.50	0.62	8
4	SH600067	冠城大通	电气设备	0.50	0.50	8
5	SZ002039	黔源电力	水力发电	0.58	0.86	7
6	SH600527	江南高纤	化纤	0.50	0.57	7
7	SH600660	福耀玻璃	汽车配件	0.50	0.57	7
8	SZ002056	横店东磁	元器件	0.67	1.00	6
9	SH600992	贵绳股份	钢加工	0.67	0.83	6
10	SZ000419	通程控股	百货	0.50	0.67	6
11	SH600081	东风科技	汽车配件	0.50	0.50	6
12	SH600123	兰花科创	煤炭开采	0.50	0.50	6
13	SH600305	恒顺醋业	食品	0.75	1.00	5
14	SH600493	凤竹纺织	纺织	0.67	1.00	5
15	SH600470	六国化工	农药化肥	0.67	0.80	5
16	SH600861	北京城乡	百货	0.67	0.80	5
17	SH600210	紫江企业	广告包装	0.58	0.80	5
18	SZ000823	超声电子	元器件	0.50	0.80	5
19	SH600594	益佰制药	中成药	0.50	0.80	5
20	SH600367	红星发展	化工原料	0.67	0.60	5
21	SH600183	生益科技	元器件	0.58	0.60	5
22	SH600467	好当家	渔业	0.58	0.60	5
23	SH600874	创业环保	环境保护	0.58	0.60	5
24	SH600969	郴电国际	水力发电	0.58	0.60	5

Figure 18: Portefeuille Final

c. Optimisation du portefeuille

Après avoir lu la nouvelle thèse de M. Gerber et M. Markowitz : «The Gerber statistic : a robust co-movement measure for portfolio optimization», j'ai décidé d'utiliser "the Gerber Statistic"²⁷ dans le cadre du modèle de "Mean – Variance" afin d'optimiser le portefeuille.

Dans la thèse, les deux auteurs principaux ont utilisé les rentabilités mensuelles de 9 différents titres financiers afin de faire le procédé de backtesting et donc de montrer que la performance du portefeuille optimisé en appliquant "the Gerber Statistic" est vraiment meilleure dans presque toutes les situations.

En suivant cette logique, j'ai donc formé une nouvelle base de données en prenant le prix de clôture ajusté des 25 actions précédemment sélectionnées à la fin de chaque mois entre 2007 et 2021.

²⁷Gerber Statistic => notamment pour remplacer "the Historical Covariance".

	SH600563	SH600809	SH600060	...	SH600467	SH600874	SH600969
Date				...			
2007-01-31	29.56	72.52	13.33	...	24.09	6.48	6.65
2007-02-28	29.29	61.11	15.74	...	26.46	8.03	7.96
2007-03-30	30.08	61.49	19.85	...	26.02	9.60	8.83
2007-04-30	34.26	63.09	29.82	...	29.53	13.71	12.80
2007-05-31	35.56	75.98	23.27	...	33.66	12.84	12.63
...
2021-08-31	345.06	1713.37	107.18	...	22.95	11.30	15.33
2021-09-30	380.07	1619.11	113.90	...	24.70	10.52	14.61
2021-10-29	483.91	1673.00	115.05	...	24.53	10.53	14.76
2021-11-30	434.78	1728.17	124.14	...	25.45	11.48	16.07
2021-12-31	399.72	1465.21	116.52	...	24.37	11.30	14.00

[180 rows x 25 columns]

Figure 19: Prix mensuel de clôture ajusté

En calculant les rentabilités mensuelles :

	SH600563	SH600809	SH600060	...	SH600467	SH600874	SH600969
Date				...			
2007-01-31	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
2007-02-28	-0.009134	-0.157336	0.180795	...	0.098381	0.239198	0.196992
2007-03-30	0.026972	0.006218	0.261118	...	-0.016629	0.195517	0.109296
2007-04-30	0.138963	0.026020	0.502267	...	0.134896	0.428125	0.449604
2007-05-31	0.037945	0.204311	-0.219651	...	0.139858	-0.063457	-0.013281
...
2021-08-31	-0.110762	0.142300	0.050578	...	-0.003474	0.102439	0.108460
2021-09-30	0.101461	-0.055014	0.062698	...	0.076253	-0.069027	-0.046967
2021-10-29	0.273213	0.033284	0.010097	...	-0.006883	0.000951	0.010267
2021-11-30	-0.101527	0.032977	0.079009	...	0.037505	0.090218	0.088753
2021-12-31	-0.080638	-0.152161	-0.061382	...	-0.042436	-0.015679	-0.128811

Figure 20: Rentabilités mensuelles

Maximisation de la rentabilité du portefeuille final sous contrainte de risque :

Notations Essentielles :

$$\max_w R_P \text{ s.c. } \sigma_P = \sigma^*$$

$$0 < w_i < 1 \quad \forall i = 1, \dots, 25$$

$$\sum_{i=1}^{25} w_i = 1$$

Le poids w pour chaque action est supérieur à 0 et inférieur à 1.

La somme des poids w fait 1.

```

MVO based on GS covariance function ...
    meanVariance (25%): ret 0.265, std 0.250, Sharpe 1.058, Sortino 1.832

code      weight
SH600563  17.00%
SH600809  29.00%
SH600060  15.00%
SH600426  13.00%
SH600067   1.00%
SZ002039   1.00%
SH600527   1.00%
SH600660   2.00%
SZ002056   5.00%
SH600992   1.00%
SZ000419   1.00%
SH600081   1.00%
SH600123   1.00%
SH600305   1.00%
SH600493   1.00%
SH600470   1.00%
SH600861   1.00%
SH600210   1.00%
SZ000823   1.00%
SH600594   1.00%
SH600367   1.00%
SH600183   1.00%
SH600467   1.00%
SH600874   1.00%
SH600969   1.00%

```

Figure 21: Performance du Portefeuille + Poids Optimal

Finalement, le portefeuille optimisé par le modèle de “Mean – Variance” avec “the Gerber Statistic” a présenté une rentabilité annuelle de 26.5%, un ratio de Sharpe égal à 1.058 et un ratio de Sortino égal à 1.832 pour un niveau de risque²⁸ donné à 25%.

Le poids optimal de chaque action dans le portefeuille est également proposé.

4.1.C : Outils employés

RStudio : La première sélection basée sur la stationnarité et l’asymétrie est faite sous R.

Python : La deuxième sélection basée sur le modèle de Machine Learning (Bernoulli Naïve Bayes) et l’optimisation du portefeuille sont faites sous Python.

²⁸Le risque du portefeuille est mesuré par l’écart-type annuel ici.

4.1.D : Difficultés rencontrées et Solutions apportées

La première difficulté est de bien connaître les modèles de Machine Learning les plus populaires et de savoir comment les appliquer à la finance.

La deuxième difficulté est de transformer les idées des M. Gerber et M. Markowitz en algorithme sur Python afin d'utiliser leur nouvelle statistique proposée pour optimiser notre propre portefeuille.

Au final, j'ai lu un livre en anglais qui s'appelle «Python for Finance and Algorithmic Trading»²⁹ afin de renforcer ma compétence en programmation sous Python et d'acquérir les bases solides en machine learning. Et je suis également allé étudier la façon de coder plus efficacement sur GitHub³⁰.

4.2 High Frequency Trading Research

4.2.A : Objectif de la mission

Le deuxième projet a pour but de comprendre comment utiliser Machine Learning et Deep Learning dans le marché chinois de commodity futures afin de construire une stratégie réellement profitable du trading haute fréquence.

Autrement dit, ce second projet tente de répondre à la dernière question précédemment mentionnée :

3. Quel est le bon moment pour acheter ou vendre ?

4.2.B : Contenu de la mission

Responsabilité :

Je suis censé construire indépendamment une stratégie de trading haute fréquence basée sur soit le modèle de **SVM** soit le modèle de **LSTM** en prenant uniquement les données sur le contrat à terme de Rebar³¹. Ensuite, il faut choisir le modèle qui arrive à gagner plus d'argent. Finalement, je dois rédiger un rapport sur ce projet en détaillant la différence entre les deux stratégies construites basées sur **SVM** et **LSTM** surtout en termes de la courbe de capital³².

²⁹Auteur : Lucas INGLESE.

³⁰GitHub est un site web et un service de cloud qui aide les développeurs à stocker et à gérer leur code, ainsi qu'à suivre et contrôler les modifications qui lui sont apportées.

³¹Steel reinforcement bars.

³²L'évolution de capital.

Présentation de la base de données :

Les données par minutes de “OHLCV”³³ sur le contrat à terme de Rebar entre le 16/05/2022 et le 19/05/2022 sont utilisées dans ce projet.

Tâches effectuées :

Dans ce projet, j’ai utilisé les prix Open pour calculer les rentabilités logarithmiques (*return*).

Si return $\geq 0 \Rightarrow$ signal d’acheter une unité du contrat à terme de Rebar \Rightarrow “+1” ;

Si return $< 0 \Rightarrow$ signal de vendre une unité du contrat à terme de Rebar \Rightarrow “-1” ;

Pour le modèle de **SVM** :

Les données par minute entre le 21:13 du 16/05/2022 et le 21:38 du 18/05/2022 sont l’ensemble d’entraînement.

Les données par minute entre le 21:39 du 18/05/2022 et le 15:00 du 19/05/2022 sont l’ensemble de test.

Pour le modèle de **LSTM** :

Les données par minute entre le 21:13 du 16/05/2022 et le 21:57 du 18/05/2022 sont l’ensemble d’entraînement.

Les données par minute entre le 21:58 du 18/05/2022 et le 15:00 du 19/05/2022 sont l’ensemble de test.

Première stratégie de trading haute fréquence avec SVM :

Le principe des SVM (Machines à Vecteurs de Support) consiste à ramener un problème de classification à un hyperplan (feature space) dans lequel les données sont séparées en plusieurs classes dont la frontière est la plus éloignée possible des points de données. Le concept de frontière implique que les données soient linéairement séparables. Pour y parvenir, les machines à vecteurs de support font appel à des noyaux, c’est-à-dire des fonctions mathématiques permettant de projeter et séparer les données dans l’espace vectoriel, les “vecteurs de support” étant les données les plus proches de la frontière. C’est la frontière la plus éloignée de tous les points d’entraînement qui est optimale, et qui présente donc la meilleure capacité de généralisation.

Après avoir construit un tel modèle de Machine Learning, j’ai mis une condition pour que le modèle trouvé soit “le meilleur” :

La condition est la suivante :

Total Strat Returns > 5.5 ; *Sharpe Ratio* > 1 ; *Precision* > 0.95 .

Total Strat Returns : Rentabilité Totale de la Stratégie

³³Open : Le premier prix tradé ; High : Le prix tradé le plus élevé ; Low : Le prix tradé le plus bas ; Close : Le dernier prix tradé ; Volume : Le volume total de toutes les transactions.

Finalement, les rentabilités de cette stratégie avec “le meilleur” modèle de SVM ont énormément surpassé celles du marché à partir du 19/05/2022 matin jusqu’à la fin de la journée. Mais après avoir tracé la courbe de capital en considérant les coûts de transaction, cette stratégie est en fait devenue une stratégie non profitable.

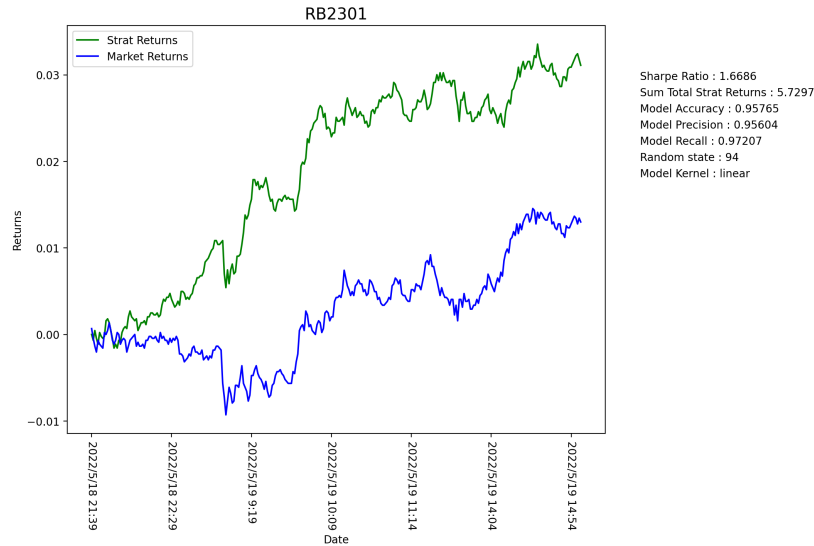


Figure 22: Backtesting - SVM

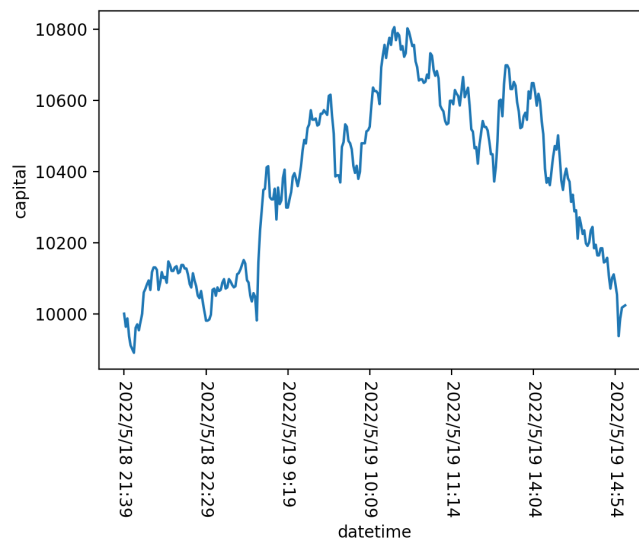


Figure 23: Courbe de Capital - SVM

Deuxième stratégie de trading haute fréquence avec LSTM :

Long Short Term Memory (LSTM) est une architecture de réseau de neurones récurrents (RNN) utilisée dans le domaine de l'apprentissage en profondeur (deep learning). A la différence des réseaux neuronaux à propagation avant, le LSTM a des confections de feed-back. Il peut non seulement traiter des data points uniques, mais également des séquences complètes de données.

Une unité LSTM de base est composée d'une cellule, d'une porte d'entrée, d'une porte de sortie et d'une porte d'oubli. La cellule se souvient des valeurs sur des intervalles de temps arbitraires et les trois portes régulent le flux d'information entrant et sortant de la cellule.

Après avoir construit un tel modèle de Deep Learning, j'ai directement demandé à l'ordinateur de montrer uniquement la courbe de capital sur laquelle la longueur de séquence permet de construire une stratégie profitable.

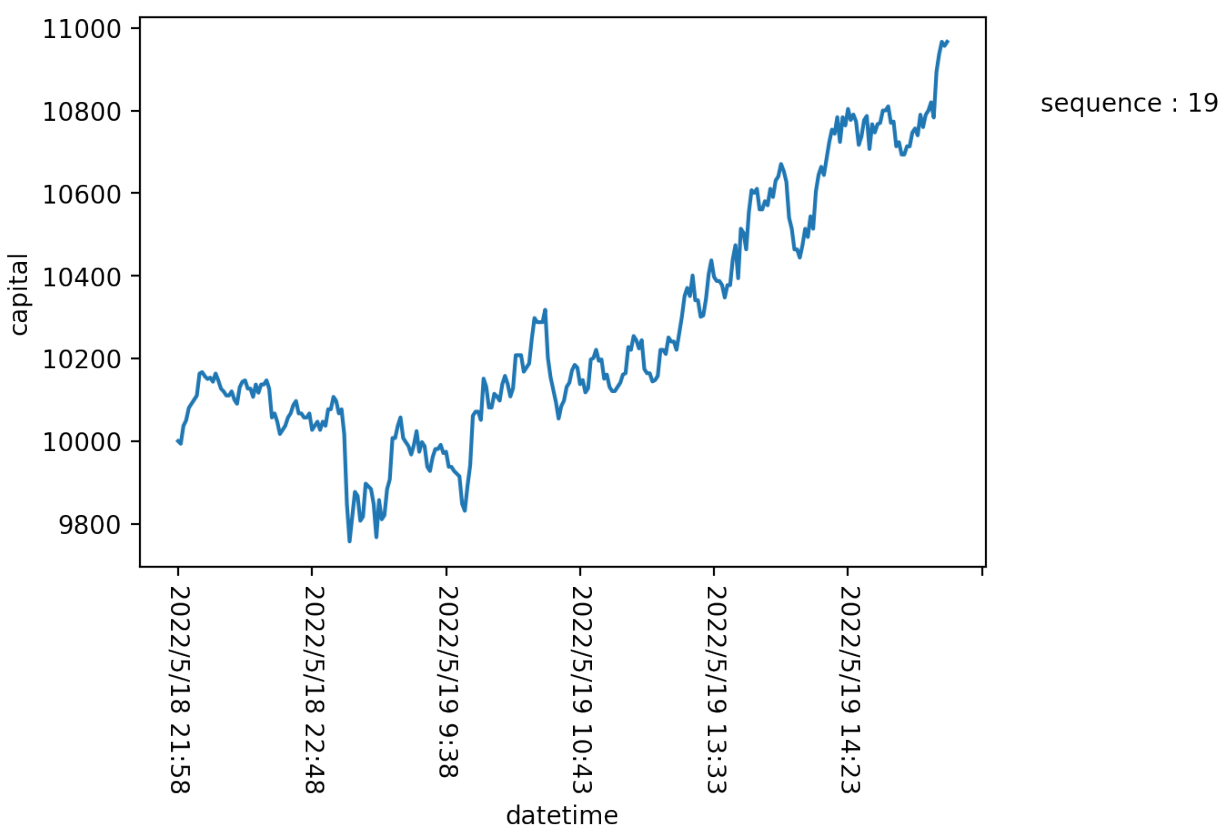


Figure 24: Courbe de Capital - LSTM

D'après le graphique, la meilleure longueur de séquence est de 19.

Et avec cette stratégie basée sur le modèle de LSTM ($sequence = 19$),

nous pourrions gagner à peu près 1000 yuans pendant une journée sur le marché chinois de commodity futures.

4.2.C : Outils employés

Python :

L'ensemble du projet est réalisé sur Python.

En particulier, la partie de Machine Learning (SVM) est faite avec le **Scikit-learn**³⁴.

La partie de Deep Learning (LSTM) est faite avec le **PyTorch**³⁵.

4.2.D : Difficulté rencontrée et Solution apportée

La seule difficulté que j'ai rencontré en faisant ce deuxième projet, c'était la compréhension du modèle de deep learning et l'utilisation de **PyTorch**.

Afin de résoudre ce problème, j'ai finalement lu un livre en anglais qui s'appelle «Deep Learning with PyTorch»³⁶ afin de renforcer ma compétence en deep learning et **PyTorch** simultanément.

5. Partie IV : Conclusion

Enseignements et Apports du stage

Ce stage a été très enrichissant pour moi, car il m'a permis de comprendre en détail surtout la construction des modèles de Machine Learning appliqués à la finance quantitative. Grâce à ce stage, j'ai eu mes propres réponses aux questions centrales liées au trading :

1. **Comment choisir les meilleures actions ?**
2. **Quelle est la somme idéale à investir ?**
3. **Quel est le bon moment pour acheter ou vendre ?**

Au cours de ce stage, j'ai appris plein de nouveaux modèles en Machine Learning sachant qu'il y a un seul cours : **Econometrics of Big Data** qui a présenté la théorie de base sur Machine Learning pendant mon M2 IREF FQA. En ce qui concerne le Deep Learning, j'ai augmenté mon niveau sur ce domaine en me basant sur des livres.

Après avoir effectué les deux projets mentionnés dans la Partie III : Principales réalisations, je pense être plus familier avec la Gestion de Portefeuille, les modèles de Machine Learning et la programmation (notamment en Python).

³⁴ Scikit-learn est la principale bibliothèque d'outils dédiés au machine learning et à la data-science dans l'univers Python.

³⁵ PyTorch est un framework de deep learning relativement récent basé sur Torch.

³⁶ Auteurs : Eli Stevens, Luca Antiga and Thomas Viehmann.

De plus, le moment où j'ai débuté mon stage, toute l'équipe de JY Asset était en train de trouver des modèles de Machine Learning convenables pour sélectionner les titres financiers probablement rentables dans le futur. Finalement, je suis fier de pouvoir leur proposer mes propres solutions pour qu'ils puissent utiliser mes modèles afin de construire une stratégie véritablement profitable de trading.

En même temps, j'ai une compréhension plus profonde de ce que font les Quants (analystes quantitatifs) dans le secteur d'asset management. Le métier de Quant en finance de marché, consiste à développer et implémenter des modèles mathématiques pour évaluer le prix des dérivés, prédire les mouvements de marché. Mais le Quant dans l'asset management consiste plutôt à concevoir des modèles mathématiques complexes utilisables pour trouver des titres financiers qui seront probablement rentables dans le futur à investir et proposer ainsi un meilleur fonds privé aux clients visés par rapport aux autres concurrents.

Utilisation des connaissances et Suggestions du programme

En outre, au cours de ce stage, j'ai utilisé de nombreuses connaissances que j'avais acquises dans le cadre de cours scolaires, comme la programmation Python et R, l'économétrie et la gestion de portefeuille.

J'aimerais également faire une petite suggestion concernant le programme d'études de notre master IREF : J'ai observé qu'il y a de plus en plus d'entreprises qui cherchent un talent en intelligence artificielle dans le domaine de la finance. Donc, je suggère que l'école pourrait offrir un cours spécial pour apprendre l'utilisation de **PyTorch** afin de montrer aux étudiants des 2 parcours (ERDS et FQA) comment construire des modèles de Deep Learning, ce qui serait très utile pour les emplois futurs.

Planification de l'avenir

Selon moi, de façon générale, l'intelligence artificielle est appliquée à la finance dans trois domaines différents : l'optimisation des portefeuilles financiers, la prédiction des prix ou des tendances futures des actifs financiers et l'analyse du sentiment des actualités. Les modèles d'apprentissage automatique sont de plus en plus répandus dans le trading algorithmique et la gestion des investissements.

Par conséquent, à l'avenir, je vais sûrement continuer à développer mes connaissances en intelligence artificielle appliquée à la finance afin de pouvoir devenir un **Research Quant**³⁷ dans le futur.

Plus précisément, après avoir terminé mon stage, je vais concentrer mon attention sur l'application de NLP(Natural Language Processing)³⁸ à la finance. Par exemple, on peut construire un modèle de NLP afin de lire les rapports annuels des grandes entreprises afin de tirer des informations utiles pour la sélection des actions.

³⁷Research Quant : C'est le métier le plus compétitif à décrocher et il n'est pas rare de voir comme prérequis un PhD pour être éligible à ce type de poste. Les équipes de Research Quant ont par ailleurs des effectifs très réduits.

³⁸NLP : Traitement automatique du langage naturel