

# ANALYSE DE SENTIMENTS ET MARCHÉS FINANCIERS

Mémoire d'état de l'art présenté par

Sarah CHATAIGNIER

Pour l'obtention du Master 1 MIAGE

de l'université

Paris 1 Panthéon – Sorbonne

Année Universitaire : **2022-2023**

Date de soutenance : **3 juillet 2023**

Directeur de mémoire : **Camille SALINESI**

Membre du jury : **Camille SALINESI, Irina RICHKOVA, Laeticia VAN BLOEME**

## Remerciements

Je tenais à remercier toutes les personnes qui m'ont aidé à la rédaction de ce mémoire.

Je souhaite tout d'abord remercier sincèrement M. Camille SALINESI, mon tuteur enseignant, Professeur des Universités, enseignant à Paris 1 Panthéon Sorbonne, de m'avoir accordé sa disponibilité, son partage de connaissances et ses précieux conseils qui m'ont permis d'avancer grandement tout au long de la réalisation de ce mémoire.

Je remercie également toute l'équipe pédagogique de l'université Paris 1 Panthéon Sorbonne, pour leurs encadrements, les connaissances et les techniques que j'ai pu assimiler pendant ma formation.

Pour finir, je remercie les membres de mon équipe au sein du Crédit Agricole CIB où j'ai pu effectuer mon alternance cette année et plus particulièrement Madame Laure MADE pour son encadrement et sa bienveillance durant mon alternance, ainsi que son soutien pour la rédaction de ce mémoire.

# Table des matières

Remerciements	2
1.Introduction	4
A. Motivation et impact	4
B. Définition de la problématique	7
C. Explication de la structure du mémoire	7
2. Les différents acteurs du marché financier	8
3. L'analyse de sentiments	9
A. Définition	9
B. Les différents algorithmes existants en analyse de sentiment	10
4.Méthodologie de recherche et littérature	13
5.Etat de l'art : synthèse	16
A. Résumés des articles	16
B. Tableau de synthèse	21
6.Etat de l'art » : l'analyse	24
A.Analyse	24
Efficacité des algorithmes	24
Importance des valences	25
Horizon temporel des prédictions	26
Conditions particulières de non-performance	26
Intérêt de ces prédictions	27
B.Perspective de recherche	28
7. Conclusion/Discussion	29
8. Glossaire	31
9.Table des figures	32
10. Références	32

# 1.Introduction

## A. Motivation et impact

Effectuant mon alternance au *Crédit Agricole Corporate Investment and Bank (CACIB)*, qui est la banque de financement et d'investissement du groupe Crédit Agricole, j'ai été très intéressée par les projets de mon équipe. En effet, nous y développons une application utilisée par les Traders et les Vendeurs en salle de marché. Venant d'une formation informatique, je ne disposais seulement que de peu de connaissances sur le monde financier. Ainsi, lors du choix du sujet pour le mémoire, l'occasion s'est présentée pour approfondir le sujet et en apprendre davantage.

De plus, ayant un certain attrait pour la Data, après avoir effectué récemment un stage de Data Analyst, je souhaitais établir un lien dans le contenu de mon mémoire avec l'analyse de données.

Pour me documenter sur le sujet, j'ai découvert de nombreuses publications scientifiques sur l'utilisation de l'intelligence artificielle pour prédire l'évolution du marché financier.

Le tout premier article sur le sujet, paru en 1975, de J. Felsen [1] traite des techniques d'apprentissage de reconnaissance de modèles appliqués à la prédiction des marchés financiers. La publication de référence sur le thème est parue en 2011 dans la revue scientifique *Expert Systems with Applications* [2] et évoque l'utilisation des réseaux de neurone dans la prédiction des indices boursiers. Une des dernières publications qui existe sur le thème du Machine Learning et du trading est de de *Han et al* date de 2023 [12].

De nos jours, les entreprises utilisent des plateformes techniques de prévision des cours de marché par des intelligences artificielles. Une des plus connues est *High Frequency Trading*, qui peut être utilisée pour prédire les mouvements futurs des prix.

Désirant me renseigner sur le sujet auprès d'un trader, j'ai eu la chance de pouvoir en interroger un dans la salle de marché du Crédit Agricole CIB, grâce à l'intervention de mon maître d'apprentissage Laure Madé.

Le trader avec lequel j'ai échangé et qui travaille chez CACIB depuis 2006, m'a indiqué que dans cette banque les traders n'utilisent aucun algorithme d'intelligence artificielle. Il m'a cependant expliqué que des initiatives de recherche sont en cours dans le pôle Intelligence Artificielle de la banque notamment concernant la prédiction des *trades* et également des recherches et développement de *Natural Langage Processing* (NLP).

De plus, il m'a dit avoir connaissance des *algorithmes de passage d'ordre* que la banque américaine Goldman Sachs utilise. Ces algorithmes font référence au *High Frequency Trading*.

Désirant en savoir plus sur le pôle intelligence artificielle, j'ai eu l'opportunité de pouvoir aussi interroger dans le cadre de ce mémoire Nicolas Damay, qui est à la tête de ce service. Il m'a indiqué que son équipe est actuellement en train de travailler sur une quinzaine de sujets : 7 de NLP, 5 de prédiction et le reste étant de l'optimisation.

Ce pôle a été créé en 2021. Sa mission est d'intégrer des outils IA au sein de la banque pour faciliter certains traitements ou prédire des résultats spécifiques.

A titre d'exemple de NLP, il m'a cité le projet déjà réalisé concernant l'analyse de rapport annuel à l'aide de *Topic classification* (Machine Learning) qui classe le contenu du rapport selon 25 topics différents.

Concernant un sujet de prédiction, il m'a évoqué le projet concernant le trading secondaire permettant à l'aide de Machine Learning et de classification de prédire si le client veut plutôt acheter ou vendre. Il a conclu en me disant que l'utilisation de cette IA divisera le taux d'erreur habituel par 2, ce qui sera une énorme avancée.

Le Trading et l'intelligence artificielle étant des sujets très vastes pour un thème de mémoire, mon directeur de mémoire Mr Salinesi m'a conseillé le thème de l'analyse de sentiments et du trading, qui rejoignait parfaitement mon souhait d'allier les marchés financiers à la data.

Le marché financier est un sujet particulièrement intéressant puisqu'il impacte considérablement l'activité économique d'un pays. Ses enjeux financiers sont majeurs et peuvent entraîner des répercussions très fortes.

Par conséquent, il faut être prudent quant à l'utilisation de l'IA en trading car les pertes financières peuvent s'avérer très grandes. Des exemples d'échecs existent, nous pouvons

évoquer le groupe américain Knight Capital qui a perdu près de 440 millions de dollars en 2012 en raison d'un problème informatique de son outil de trading automatisé.

De plus, la dimension réglementaire dans l'utilisation de l'IA est capitale : en France, l'Autorité des marchés financiers (AMF) donne des recommandations précises en rapport avec son utilisation.

Ainsi, les algorithmes d'analyse de sentiments sont des outils qui vont nous permettre de comprendre les influences des fluctuations des marchés financiers et par conséquent prévenir ces différents impacts, anticiper et éviter l'effondrement du marché.

L'objectif de mon état de l'art est de montrer les différents algorithmes d'analyse de sentiments liés au trading et de les analyser.

Les états de l'art déjà existants sur ce sujet évoquent seulement certains algorithmes d'IA liés au trading. Mon camarade Alexandre Chekroun en 2021 [11] évoquait un aspect de ce sujet qui était l'utilisation des réseaux de neurones convolutifs.

Ce qui distingue mon état de l'art des précédents est la réunion de la grande majorité des algorithmes d'IA liés à l'analyse de sentiment et à la prédiction boursière en les comparant, en les analysant et en établissant des conclusions.

## B. Définition de la problématique

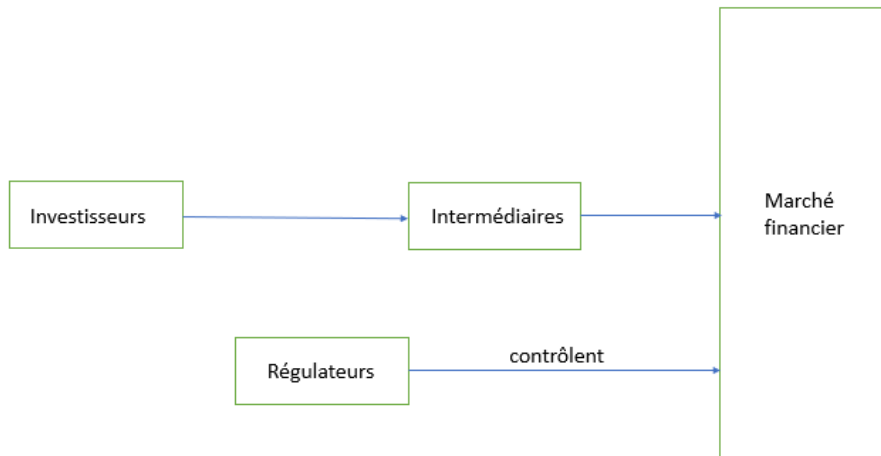
Aussi serait-il intéressant de nous demander :

Comment les algorithmes d'analyse de sentiments montrent-ils que les sentiments des acteurs du marché financier influent sur l'évolution du cours des actions ?

## C. Explication de la structure du mémoire

Afin de répondre à cette problématique, ce mémoire est structuré en plusieurs parties : dans un premier temps nous présenterons les différents acteurs du marché financier, puis nous expliquerons ce qu'est l'analyse de sentiments et les différents algorithmes existants dans ce domaine. Ensuite, nous présenterons les différents articles puis nous comparerons les différents algorithmes et les analyseront, ce qui nous permettra de répondre à la question initiale.

## 2. Les différents acteurs du marché financier



*Figure 1 : Schéma des acteurs du marché financier*

Dans cet état de l'art, nous évoquons deux acteurs importants du marché financier : les investisseurs et les traders. Il peut être intéressant de définir l'ensemble des acteurs de ce secteur.

Le marché financier est la réunion de deux marchés distincts : le marché primaire où les actions sont créées, et le marché secondaire où ces actions sont vendues et achetées. C'est ce marché secondaire que l'on appelle le marché boursier.

Le marché financier est composé de différents acteurs : les institutions financières, les investisseurs, les émetteurs, les intermédiaires, les régulateurs.

- **Les institutions financières** ont des clients pour lesquels elles investissent dans le marché. Ce sont notamment des banques, des fonds de placements ou les assurances.
- **Les investisseurs** sont des particuliers qui effectuent des placements afin d'en tirer des profits.



- **Les émetteurs** lèvent des capitaux propres ou actions et des titres et créent ainsi de la dette.
- **Les intermédiaires** sont des entreprises ou personnes qui font l'intermédiaire entre les investisseurs et les marchés financiers. On les appelle aussi les courtiers. En effet, les courtiers aident les investisseurs à acheter et vendre des titres.
- **Les traders** achètent et vendent des titres pour leur propre compte ou pour le compte de leur entreprise
- **Les régulateurs** contrôlent la validité des pratiques commerciales sur le marché boursier. Par exemple, c'est l'Autorité des marchés financiers ou AMF qui est le régulateur du marché français.

### 3. L'analyse de sentiments

#### A. Définition

L'analyse de sentiment consiste à étudier les émotions, opinions, attitudes d'individus au sujet d'évènements, de produits, d'organisations et de services. Cela passe par l'extraction de sentiments, leur classification ou la synthèse d'opinions.

Les défis de l'analyse de sentiments résident dans la difficulté de traduire les émotions en expressions mathématiques susceptibles d'être informatisées. De plus, il est nécessaire d'extraire ces émotions depuis des textes ou documents. Pour cela, l'analyse de sentiments utilise les méthodes du Traitement Automatique des Langues TAL ou Natural Language Processing (NLP) qui combine linguistique, informatique et intelligence artificielle.

## B. Les différents algorithmes existants en analyse de sentiment

Afin de mieux comprendre notre étude sur les algorithmes existants en matière d'analyse de sentiments et de leur efficacité, il peut être intéressant de les définir.

- Stratégie d'analyse de mots avec le dictionnaire Harvard-IV

Dictionnaire fréquemment utilisé en analyse financière pour sa classification émotionnelle des mots (négatif ou positif). [3]

- Technique de régression linéaire

Méthode statistique qui consiste à établir une relation linéaire entre deux variables. Le plus souvent, on cherche à montrer qu'une variable  $y$  se comporte comme une fonction affine d'une autre variable  $x$  (en recherchant la droite passant par le maximum de points). [3,5,10]

- Algorithme de BERT (Bidirectional Encoder Representations from Transformers)

Méthode de Machine Learning développée par Google qui vise à modéliser un langage et à prédire la probabilité de la suite d'une phrase. [6]

- Arbre syntaxique (parse tree)

Arbre qui représente la structure syntaxique d'une chaîne de caractère. [7]

- Naive Bayesian (génératif)

Algorithme d'Apprentissage Supervisé qui consiste à classier un texte en utilisant l'indépendance de ses éléments. [7]

- Support Vector Machine (SVM)

L'algorithme SVM est utilisé pour séparer différentes classes d'exemple, en trouvant l'hyperplan c'est-à-dire la meilleure ligne ou courbe qui sépare ces classes. [7]

- ARIMA

Méthode de prédiction des séries temporelles. Sa particularité est qu'il supporte les valeurs extrêmes. [10]

- Long short-term memory (LSTM)

LSTM est un type particulier de réseaux de neurones. Sa particularité est de mémoriser et d'oublier des informations sur de longues durées. [8,10]

- Réseau de neurone

Assemblage en parallèle de processeurs interconnectés qui permet à la machine d'apprendre. [8,10]

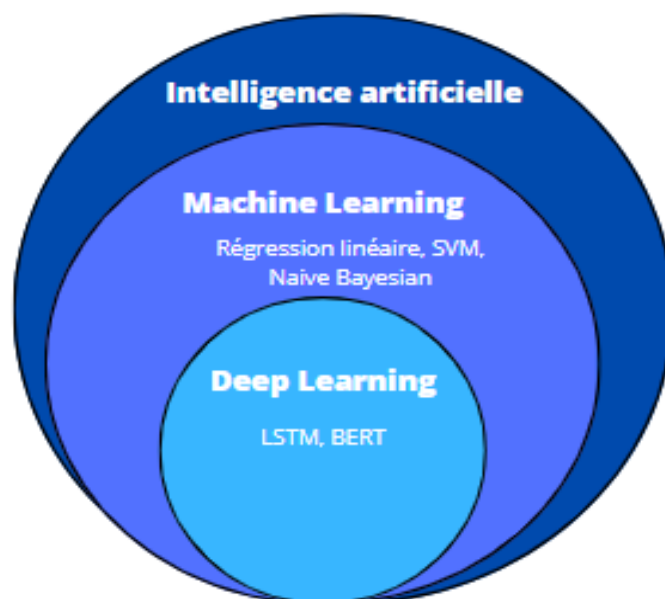
- Apprentissage supervisé :

Domaine du Machine Learning qui utilise des Datasets étiquetés (labeled) pour entraîner un algorithme à classier des données ou à prédire des résultats. [7]

Enfin, nous pouvons classifier les différents algorithmes présentés en Deep Learning ou Machine Learning.

Le Machine Learning consiste à apprendre à la machine à reconnaître un pattern pour ensuite prendre une décision de manière autonome. Si l'algorithme s'est trompé, l'humain va rectifier et la machine apprendra de son erreur.

D'autre part, le Deep Learning est une sous-catégorie du Machine Learning qui utilise des réseaux de neurone à plusieurs couches afin d'améliorer les prédictions. Les techniques Deep Learning ne nécessitent pas d'intervention humaine.



*Figure 2 : Classification des algorithmes de cet état de l'art en Machine Learning ou en Deep Learning*

## 4.Méthodologie de recherche et littérature

Afin de trouver des articles de recherche pour documenter mon mémoire sur le thème de l'analyse de sentiment et du marché financier, j'ai consulté plusieurs moteurs de recherche : Google Scholar, domino.univ-paris1.fr et researchgate.org. Afin d'affiner ma recherche, j'ai utilisé les mots clés suivants "trading emotion analysis" "sentiment analysis" "stock market sentiment" "algorithm sentiment trading" "NLP finance" "BERT sentiment analysis trading", "Artificial intelligence sentiment analysis trading". J'ai également essayé de sélectionner les articles les plus récents à l'aide du filtre des dates, afin de n'étudier que des articles récents.

J'ai écarté de mon choix les publications en ligne, les articles de revues grand public ou publiés par des entreprises, les mémoires de Master et de thèse, les blogs, et les livres, même écrits par des auteurs reconnus.

De plus, j'ai écarté les articles qui ne comportaient pas d'expérimentation car cette dernière est nécessaire pour mon état de l'art.

Afin de confirmer mon choix sur la pertinence des articles sources à étudier, je n'ai pas sélectionné les articles qui n'étaient pas publiés dans une revue de conférence classée A ou B selon le ranking CORE.

Sur 10 articles que j'ai finalement sélectionnés, j'en ai conservé 8.

Par exemple, un article que j'ai écarté parmi la liste de départ était le suivant : *Dong, Hang, et al. "How are social and mass media different in relation to the stock market? A study on topic coverage and predictive value." Information & Management 59.2 (2022): 103588.*

En effet, je l'ai retiré car il n'évoque pas d'algorithme d'analyse de sentiment mais explique seulement la relation des réseaux sociaux et des médias avec la bourse.

Voici la liste finale des articles retenus :

- [3] Sul, Hongkee, Alan R. Dennis, and Lingyao Ivy Yuan. "**Trading on twitter: The financial information content of emotion in social media.**" 2014 47th Hawaii International Conference on System Sciences. IEEE, 2014.
- [4] Risius, Marten, Fabian Akolk, and Roman Beck. "**Differential emotions and the stock market-the case of company-specific trading.**" (ECIS 2015 completed research papers).
- [5] Liu, Qing, et al. "**Synergy between stock prices and investor sentiment in social media.**" *Borsa Istanbul Review* (2022).
- [6] Sousa, Matheus Gomes, et al. "**BERT for stock market sentiment analysis.**" *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2019
- [7] Yadav, Ritu, Ashwani Kumar, and A. Vinay Kumar. "**Event Based Sentiment Analysis on Futures Trading.**" *The Journal of Prediction Markets* 13.1 (2019): 57-80.
- [8] Wang, Cuiyuan, et al. "**Learning to trade on sentiment.**" *Journal of Economics and Finance* 46.2 (2022): 308-323.
- [9] Fenton-O'Creevy, Mark, et al. "**Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders.**" *Journal of Organizational Behavior* 32.8 (2011): 1044-1061.
- [10] Mehta, Yash, Atharva Malhar, and Radha Shankarmani. "**Stock price prediction using machine learning and sentiment analysis.**" 2021 2nd International Conference for Emerging Technology (INCET). IEEE, 2021.

<b>Article</b>	<b>Date de parution</b>	<b>Type</b>
[3]	2014	Conférence
[4]	2015	Papier de recherche
[5]	2022	Revue scientifique
[6]	2019	Conférence
[7]	2019	Journal scientifique
[8]	2022	Journal scientifique
[9]	2011	Journal scientifique
[10]	2021	Conférence

*Figure 3 : Tableau des articles et de leur type*

## 5. Etat de l'art : synthèse

### A. Résumés des articles

**[3] Sul, Hongkee, Alan R. Dennis, and Lingyao Ivy Yuan. "Trading on twitter: The financial information content of emotion in social media." 2014 47th Hawaii International Conference on System Sciences. IEEE, 2014.**

Cet article propose d'étudier un Dataset composé des données financières pour le cours de clôture de chaque action du S&P 500, qui ont été obtenues auprès de Compustat, du Centre de recherche sur les prix des titres (CRSP) et du site Web de Kenneth French. La période d'échantillonnage va de mars 2011 à février 2012. De plus, il est composé de données Twitter : 2 503 385 tweets publics d'investisseurs contenant le symbole \$ correspondant avec un symbole boursier S&P500.

L'indice S&P 500, pour Standard & Poor's 500 est un indice boursier construit à partir de 500 grandes entreprises cotées sur les bourses américaines.

Twitter est un réseau social intéressant pour analyser les sentiments, sur lequel les utilisateurs publient de courts messages de 140 caractères. Les abondantes données échangées sur Twitter chaque minute offrent aux chercheurs, notamment en informatique, l'occasion d'observer les comportements individuels et même de les prédire à travers des outils d'analyse.

Pour l'analyse de sentiments des tweets des investisseurs, une stratégie d'analyse de mots avec le dictionnaire Harvard-IV a été effectuée, puis une technique de régression linéaire. Les hypothèses sont faites sur les impacts du jour même, du lendemain, et dans 10 jours. Une analyse de la valence émotionnelle a été faite : positive, négative ou neutre.

Cet article démontre que les investisseurs avec beaucoup d'abonnés sur Twitter ont un impact sur la bourse du jour même. Les investisseurs avec peu d'abonnés ont un impact sur la bourse dans le futur.



**[4] Risius, Marten, Fabian Akolk, and Roman Beck. "Differential emotions and the stock market-the case of company-specific trading." (ECIS 2015 completed research papers).**

L'objet de cet article est d'analyser un Dataset composé des données financières qui représentent les mouvements du cours des actions NYSE spécifiques à l'entreprise obtenus auprès de Yahoo!Finance. De plus, ce Dataset est composé de données Twitter : 5,5 millions de messages Twitter sur 33 sociétés du S&P100 qui ont été collectés sur une période de trois mois. Le S&P 100 est un indice boursier. Il est composé des 100 plus grandes sociétés cotées sur les bourses américaines de l'indice « S&P 500 ».

Pour l'analyse de sentiments, l'outil de sentiment "SentiStrength 2" a été utilisé ainsi que les descriptions du modèle émotionnel d'Ekkekakis.

7 émotions différentes sont analysées :

- les émotions positives : affection, bonheur, satisfaction
- les émotions négatives : peur, colère, dépression et mépris

Les hypothèses émises sont à propos des différents sentiments (parmi les 7 émotions) des investisseurs par rapport au cours des actions d'une entreprise.

Cet article démontre que :

- 1) Les émotions différenciées sont plus fortement associées aux variations du cours des actions spécifiques à la société que le sentiment moyen indifférencié.
- 2) Les émotions négatives ont généralement un pouvoir explicatif plus élevé.
- 3) La force des émotions se référant à des événements spécifiques (dépression et bonheur) expliquent les mouvements de prix.

**[5] Liu, Qing, et al. "Synergy between stock prices and investor sentiment in social media." *Borsa Istanbul Review* (2022).**

Contrairement à l'article précédent, le parti pris de cet article est de montrer qu'il existe une synergie entre les sentiments des investisseurs et l'évolution du marché boursier. Le Dataset étudié provient de messages des investisseurs sur la plus grande plateforme de médias sociaux de Chine : DongFang Fortune Sack Bai.

Pour l'analyse de sentiments, les techniques de : NLP, robot d'indexation, IC (indice de confiance des investisseurs), d'apprentissage supervisé ainsi que de régression linéaire (coefficient de corrélation entre -1 et 1) ont été utilisées. Les émotions positives, négatives et neutres sont analysées (la valence émotionnelle).

Les hypothèses émises sont à propos de la synergie entre les sentiments des investisseurs et les rendements boursiers.

Les auteurs de l'article prennent pour expérimentation l'analyse de 2 villes chinoises sur l'effet du COVID 19 sur la synergie entre les sentiments et rendements boursiers. On observe dans

le cas de Wuhan une plus grande synergie pendant le confinement tandis que le cas de Shanghai est opposé (synergie plus grande avant le confinement).

Enfin, l'article démontre que le phénomène de synergie ne s'applique pas sur des intervalles locaux.

**[6] Sousa, Matheus Gomes, et al. "BERT for stock market sentiment analysis." 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2019.**

L'originalité de cet article est la méthode utilisée dans l'analyse de sentiment : l'algorithme de Bert.

Le Dataset utilisé est constitué d'informations financières issues des sites web: CNBC, Forbes, Investopedia, New York Times, Washington Post (et d'autres) et l'indice financier Dow Jones. Les articles d'actualités boursières sont étiquetés comme positifs, neutres ou négatifs.

L'hypothèse émise est que BERT est plus efficace que d'autres algorithmes d'analyse de sentiment. Cet article démontre que BERT a des performances supérieures que les réseaux de neurones convolutifs (TextCNN).

**[7] Yadav, Ritu, Ashwani Kumar, and A. Vinay Kumar. "Event Based Sentiment Analysis on Futures Trading." The Journal of Prediction Markets 13.1 (2019): 57-80.**

L'objectif poursuivi dans cet article est de démontrer qu'il est plus efficace d'analyser les sentiments liés à des événements plutôt qu'aux actualités pour prévoir l'évolution du marché.

Le Dataset étudié est le service d'information en continu de Bloomberg. La raison consiste dans le fait que l'on est capable d'identifier un événement sur un service d'information en continu. En effet, les nouvelles relatives à un même événement sont proches dans le temps et suivent des ajouts et des répétitions.

Les données financières extraites viennent du National Stock Exchange NSE of India. L'objectif est de proposer une nouvelle méthode pour collecter des nouvelles, les regrouper en un événement puis associer l'événement avec la tendance du marché.

Pour l'analyse de sentiment, un arbre syntaxique (*parse tree*) est utilisé pour extraire les différentes clauses d'une phrase (de la nouvelle) pour en retirer les événements correspondants (une clause= un événement). Pour chaque clause, on crée un vecteur événement (event tuple) qui contient le verbe, le sujet et l'objet de la clause. A chaque vecteur événement est associé un sentiment (positif, négatif ou neutre) selon deux méthodes : Naive Bayesian (génératif) et Support Vector Machine SVM (discriminatif).

Les résultats de l'article sont les suivants. L'objet de la clause n'a pas d'influence sur les résultats : le vecteur événement utilisé ne contient que le verbe et le sujet de la clause. Cela montre que le sujet et le verbe sont des facteurs clés de l'événement.

Le modèle d'analyse de sentiment basé sur les événements permet de prédire le comportement du marché (profit de 11.06% par mois contrairement à 0% de profit avec le modèle basé sur les nouvelles).

Le modèle basé sur les évènements est plus performant que celui basé sur les nouvelles pour classer les sentiments positifs et négatifs.

Cependant, il est moins performant que le modèle d'analyse de sentiment basé sur les nouvelles en ce qui concerne les sentiments neutres

**[8] Wang, Cuiyuan, et al. "Learning to trade on sentiment." *Journal of Economics and Finance* 46.2 (2022): 308-323.**

Cet article se propose de démontrer l'efficacité des méthodes de *deep learning* pour évaluer l'influence de l'analyse de sentiments sur le marché boursier.

Le Dataset étudié est constitué d'articles de presse financiers du site Seeking Alpha, un fournisseur de services de contenu en ligne pour les marchés financiers, ainsi que l'indice financier Dow Jones. Seeking Alpha est une plateforme de réseaux sociaux d'investissement populaire avec des millions d'utilisateurs enregistrés, donc intéressant à analyser.

Pour analyser les sentiments dans ces articles, LSTM (long short-term memory), réseau de neurones particulier a été utilisé. Les sentiments positifs et négatifs sont analysés.

Une hypothèse est émise : les modèles de deep learning comme le réseau de neurones à mémoire à court et à long terme sont plus performants que les méthodes de machine learning (régression linéaire) pour montrer l'influence des sentiments des investisseurs sur le marché.

Cet article démontre qu'en utilisant ce réseau de neurones, on peut créer une stratégie de trading qui prend en compte l'influence des sentiments des investisseurs pour investir sur le marché. Cette stratégie permet un rendement intéressant, ce qui démontre l'efficacité du réseau de neurones. L'hypothèse de départ est validée.

**[9] Fenton-O'Creivy, Mark, et al. "Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders." *Journal of Organizational Behavior* 32.8 (2011) : 1044-1061.**

Contrairement à tous les articles précédents qui se concentrent sur les sentiments des investisseurs, cet article parle d'un autre acteur du marché financier : le trader. Le dataset est composé d'interviews de 30 à 90 minutes de 118 traders professionnels et 10 senior managers en banque d'investissement.

L'algorithme utilisé est le programme NVIVO, qui identifie des déclarations communes aux interviews pour créer des catégories. Il sépare l'échantillon selon le niveau d'expérience (années de travail) et d'expertise (salaire) et compare les réponses des traders aux interviews selon ce qu'ils sont : non expérimentés, peu payés, très expérimentés mais peu payés et enfin très bien rémunérés.

Par rapport aux émotions, l'article fait la distinction entre humeur ("mood") et émotions. Parmi les émotions (ou humeurs) citées par les traders lors des interviews, on retrouve : l'euphorie, la prudence, l'arrogance, le relâchement, le stress, la confiance, l'empathie.

Cet article démontre que le contrôle des émotions joue un rôle central dans la prise de décision des traders. Les traders utilisent différentes stratégies de contrôle des émotions suivant leurs performances.

Les traders très performants ont la volonté de supprimer les sentiments négatifs pour maintenir une objectivité et poursuivre des objectifs à long-terme. Cependant, les traders peu expérimentés adoptent une stratégie basée sur l'évitement (quitter le bureau) ou sur le travail intellectuel de modulation de leur réponse émotionnelle.

**[10] Mehta, Yash, Atharva Malhar, and Radha Shankarmani. "Stock price prediction using machine learning and sentiment analysis." 2021 2nd International Conference for Emerging Technology (INCET). IEEE, 2021.**

Partant du constant qu'il existe de nombreuses méthodes pour démontrer l'impact de l'analyse de sentiments sur le marché boursier, cet article se propose de trouver la plus efficace d'entre elles.

Le dataset étudié est composé de données financières : différentes actions de l'API Yahoo Finance. De plus, le dataset est composé de données Twitter : des tweets sont collectés en utilisant l'API de Twitter appelée Tweepy, qui permet d'analyser les sentiments de tweets.

Dans cet article sont analysés et comparés 3 algorithmes différents : ARIMA, LSTM, et la régression linéaire.

Le modèle ARIMA donne la meilleure précision pour chaque action.

## B. Tableau de synthèse

Le tableau suivant synthétise les articles de cet état de l'art, en différentes colonnes afin de répondre à notre problématique sur l'influence des sentiments sur la bourse.

Dans ce tableau, nous indiquons tout d'abord la date de parution de l'article, ses Datasets, le type d'émotion analysée ainsi que les algorithmes et techniques utilisés. Enfin, nous expliquons les conclusions et démonstrations de l'article en question.

Le Dataset est divisé en deux catégories : une correspondant aux données financières sur lesquelles les auteurs des articles se basent pour étudier les variations des actions, et une autre pour les données des réseaux sociaux, ou de presse, où sont extraits les sentiments, afin de les analyser avec les différents algorithmes.

Article	Date	Dataset : données financières	Dataset : données réseau sociaux	Algorithme ou technique utilisés	Type d'émotion	Ce qui est démontré
[3]	2014	Cours de clôture de chaque action du S&P 500	Twitter : 2 503 385 de tweets publics d'investisseurs	-Dictionnaire Harvard-IV  -Régression linéaire	Valence émotionnelle : positive, négative (ou neutre)	Investisseur ayant beaucoup d'abonnés ☒ impact la bourse du jour même.  Investisseur ayant peu d'abonnés ☒ impact la bourse dans le futur
[4]	2015	Actions NYSE obtenus auprès de Yahoo!Finance	Twitter : 5,5 millions de tweets sur 33 sociétés du S&P 100 qui ont été collectés sur une période de trois mois.	-SentiStrength 2  - Modèle émotionnel d'Ekkekakis	7 émotions différentes : - positives : affection, bonheur, satisfaction  - négatives : peur, colère, dépression et mépris	- Emotions différenciées : plus fortement associées aux variations des actions  -Emotions négatives : pouvoir explicatif plus élevé  -Evénements spécifiques (dépression et bonheur) expliquent les mouvements de prix.
[5]	2022		Messages d'investisseurs sur le réseau social Dong Fang Fortune Sack Bai	-NLP -robot d'indexation -méthode de régression linéaire	Positive, négative, neutre	Synergie entre les sentiments et rendements boursiers sauf dans certains cas

Article	Date	Dataset : données financières	Dataset : données réseau sociaux	Algorithme ou technique utilisés	Type d'émotion	Ce qui est démontré
[6]	2019	Indice financier Dow Jones	Les sites web: CNBC, Forbes, Investopedia, New York Times, Washington Post	Algorithme de BERT	Positive, négative, neutre	Meilleures performances de BERT par rapport à CNN
[7]	2019	Net Buying Pressure of National Stock Exchange of India	Le service d'information en continu de Bloomberg.	-Arbre syntaxique (parse tree) -Naive Bayesian (génératif) et -Support Vector Machine SVM (discriminatif).	Positive, négative, neutre	Le modèle basé sur événements (par rapport à celui basé sur les nouvelles) : -meilleure classification - meilleure prédiction - moins bonne performance pour les sentiments neutres
[8]	2022	Indice financier Dow Jones	Articles de presse financiers du site Seeking Alpha	LSTM	Positive, négative	Les sentiments permettent de construire une stratégie de trading qui affiche des rendements constamment favorables sur une longue période.
[9]	2011		Interviews de traders et de senior managers	Programme NVIVO	- L'euphorie, la prudence, l'arrogance, le relâchement, le stress, la confiance, l'empathie.	Le contrôle des émotions joue un rôle central dans la prise de décision des traders.
[10]	<u>2021</u>	Différentes actions de Yahoo ! Finance	Twitter : tweets de l'API de Twitter : Tweepy, permettant d'analyser les sentiments de tweets	-Tweepy -ARIMA, LSTM, et la régression linéaire	L'analyse de sentiments plus évoquée que les "émotions"	Le modèle ARIMA donne la meilleure précision pour chaque action.

A travers ce tableau, nous pouvons constater que les articles s'étalent sur 11 ans : de 2011 à 2022 : on peut y étudier une certaine évolution notamment sur les techniques et algorithmes d'analyse de sentiments utilisés.

En effet, on remarque une nette différence entre les articles de 2014 et 2015 qui utilisent la régression linéaire et l'outil Sentistrength et les articles entre 2019 et 2022 qui utilisent tous l'intelligence artificielle sous toutes ses formes : LSTM, NLP, BERT, ARIMA. De plus, on constate que la régression linéaire est une technique fortement utilisée : elle apparait dans 3 articles [3,5,10]

Donc, on peut dire que l'utilisation de l'intelligence artificielle, notamment au service de l'analyse de sentiment, a émergé ces 5 dernières années

Concernant les Datasets, nous pouvons voir que le réseau social Twitter revient 3 fois : il est intéressant à analyser pour extraire les sentiments de ses utilisateurs. On remarque d'ailleurs que dans l'article [10], l'API de Twitter de Tweepy est utilisée pour analyser directement les sentiments

Tous ces articles montrent que nous pouvons prédire l'évolution du marché grâce aux algorithmes d'analyse de sentiments utilisés sur les investisseurs, sauf l'article [9] qui montre que les émotions des traders jouent un rôle sur leur performance et par conséquent peuvent influencer sur la bourse.

## 6. Etat de l'art » : l'analyse

### A. Analyse

Rappel de la problématique :

Comment les algorithmes d'analyse de sentiments montrent-ils que les sentiments des acteurs du marché financier influent sur l'évolution du cours des actions ?

Afin de synthétiser et d'analyser les articles qui permettent de répondre à notre problématique, nous pouvons diviser notre réflexion en plusieurs parties : l'efficacité des algorithmes, l'importance des valences, l'horizon temporel des prédictions, la pertinence des prédictions et enfin l'intérêt de ces prédictions par rapport aux autres outils déjà existants.

### Efficacité des algorithmes

Nous avons 2 types d'algorithmes dans notre étude :

- les algorithmes de classification des sentiments
- les algorithmes de prédiction

Afin de classer une information comme étant positive ou négative, les vecteurs de support SVM sont plus performants que le classificateur naïf bayésien. C'est cette étape de classification qui va ensuite permettre d'effectuer la prédiction sur le cours des actions.

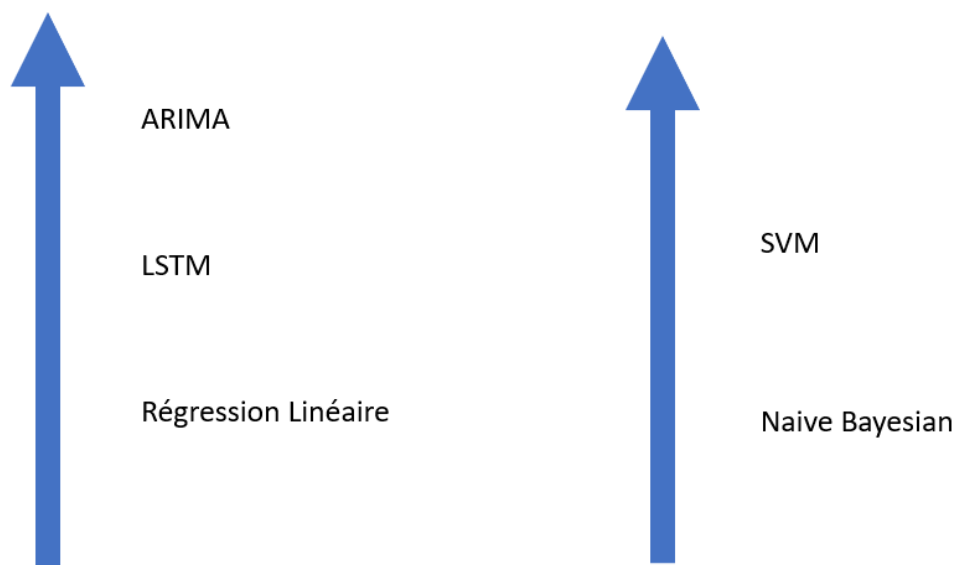
Concernant les algorithmes de prédiction qui ne fonctionnent pas bien : nous pouvons évoquer la régression linéaire [3] qui ne fonctionne pas pour des valeurs extrêmes et les réseaux de neurone convolutifs [6] ne sont pas très efficaces non plus.

Le modèle ARIMA [10] est efficace quant à la prédiction des fluctuations du marché boursier. IL est plus performant que les réseaux de neurones LSTM, lui-même meilleur que la régression linéaire. [10]



Enfin, nous avons obtenu que l'algorithme BERT est plus efficace que les réseaux de neurones convolutifs dans la prédiction financière basée sur l'analyse de sentiments.

En résumé, sur le schéma ci-dessous, à droite nous pouvons voir l'efficacité croissante des algorithmes de classification des sentiments et à gauche l'efficacité des algorithmes de prédiction.



*Figure 4 : Performance des algorithmes*

## Importance des valences

Le type d'émotion analysé est important. En effet, Risuis et al [4] expliquent que les différentes émotions ont un effet différent sur le cours de la bourse. Ils obtiennent plusieurs hypothèses concluantes :

- Les valences positives n'ont pas d'influence
- Plus la valence est négative, plus le cours de l'entreprise est bas
- La force d'affection n'est pas liée aux actions
- La force de satisfaction n'est pas liée aux cours des actions

- Plus la dépression est forte, plus le cours de l'action est bas
- Le sentiment de mépris n'est pas lié à la bourse

## Horizon temporel des prédictions

D'une part, Wang et al. [8] ont créé deux stratégies de trading basé sur les prédictions issues de l'analyse de sentiments : l'une à courte terme et l'autre à long terme. Ils ont remarqué qu'il y a une forte différence de profits entre ces deux stratégies. En effet, les profits de la stratégie à long terme sont toujours plus importants que celle à court terme, excepté la première année.

D'autre part, l'article [3] montre une relation linéaire entre les sentiments contenus dans des tweets et des comportements anormaux du marché. Les auteurs utilisent une régression linéaire pour démontrer que les tweets non *retweetés* d'utilisateurs ayant peu de *followers* avait un impact sur les futurs profits, 10 ou 15 jours plus tard. Cependant, les tweets *retweetés* d'utilisateurs avec beaucoup d'abonnés n'ont pas d'impact sur les futurs profits. L'article développe ainsi une stratégie de trading basée sur les sentiments de tweets non *retweetés* d'utilisateurs ayant peu d'abonnés. Cette stratégie repose sur une période de rétention (temps entre l'achat et la vente) de 10 ou 20 jours et permet d'important résultats positifs.

## Conditions particulières de non-performance

Les auteurs de l'article [5] prennent pour expérimentation l'analyse de 2 villes chinoises sur l'effet du COVID 19 sur la synergie entre les sentiments et rendements boursiers. On observe dans le cas de Wuhan une plus grande synergie pendant le confinement tandis que le cas de Shangai est opposé (synergie plus grande avant le confinement).

L'article démontre que le phénomène de synergie ne s'applique pas sur des intervalles locaux : dans ce cas, c'est une condition particulière de non-performance des algorithmes de sentiments utilisés.

## Intérêt de ces prédictions

Yadav et al. [7] sont parvenus à un modèle de prédiction basé sur l'analyse de sentiments contenus dans les nouvelles. Ce modèle génère un profit de 11.06 % par mois, ce qui est considérable.

Wang et al [8] ont utilisé l'analyse de sentiment pour construire une stratégie de trading neutre au marché. Cette stratégie est plus performante que le rendement du marché. En effet, avec un investissement initial de 1\$ en 2013, leur stratégie permet d'atteindre 2.28\$ à la fin de 2017 tandis que le rendement du marché atteint 2.07 \$.

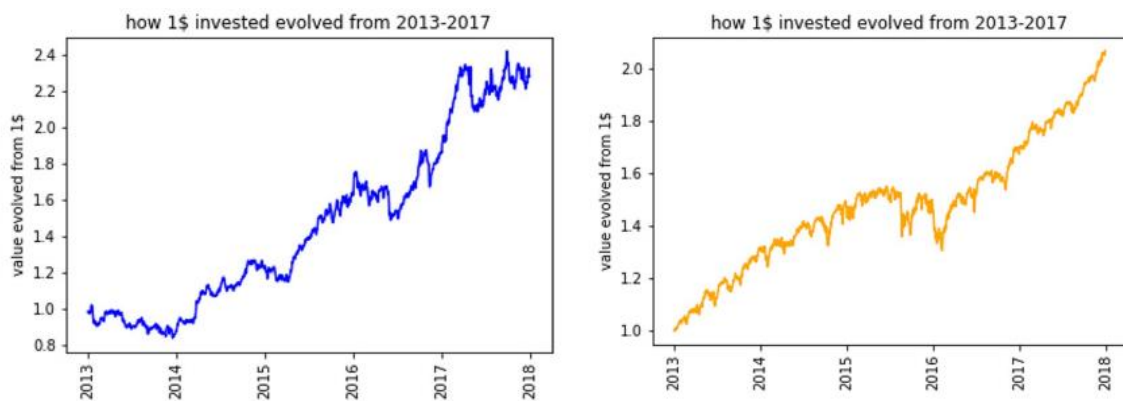


Figure 5 : Evolution du profit avec la stratégie de trading (à gauche) et le rendement du marché (à droite) [7]

Il est donc plus intéressant d'utiliser des prédictions réalisées grâce à l'analyse de sentiments plutôt que de se baser sur les tendances boursières ou la saisonnalité.

## B.Perspective de recherche

Les acteurs du marché financier n'ont pas encore intégré de nombreuses applications de l'intelligence artificielle. Il y a beaucoup de développement possible dans ce domaine. Une section Intelligence Artificielle spécialisée sur l'analyse de sentiment a récemment été créée au Crédit Agricole CIB. Cela témoigne de l'attractivité qu'exerce ce domaine.

Sousa et al. [6] ont tenté de prédire le cours du DJI ou *Dow Jones Industrial Average*, qui regroupe plusieurs sociétés tels que Microsoft ou Intel. Cependant, le nombre important d'entreprises présentes ont rendu difficile la prédiction. Ainsi, les auteurs aimeraient à l'avenir choisir une société et récolter les informations financières qui la concernent pour prédire son évolution avec plus de précision.

Yadav et al. [7] ont été limités dans leur extraction de sentiments par l'inexactitude de leurs analyseurs syntaxique. Une de leurs directions de recherche est d'entraîner et de modéliser des analyseurs syntaxiques pour extraire les sentiments des titres de nouvelles.

Les auteurs de l'article [8] prévoient de collecter des articles d'informations financières depuis plus de plateformes. Ils cherchent à construire des stratégies d'investissement prenant en compte plus de facteurs notamment macro-économiques comme le taux d'inflation et le taux de chômage.

Fenton-O'Creevy, Mark, et al. [9] ont plusieurs directions pour leurs futures recherches :

- L'analyse de l'authenticité du discours des traders lors des interviews (cachent-ils leurs émotions ?)
- L'hypothèse selon laquelle les traders les plus performants ont une plus forte critique personnelle
- L'hypothèse selon laquelle l'efficacité de l'intuition des traders est réduite par une stratégie défensive de contrôle des émotions
- L'interaction entre l'intuition et l'analyse dans le travail des traders

On peut observer une tendance dans les perspectives de recherche. Les auteurs cherchent le plus souvent à améliorer leurs algorithmes en réduisant les défauts de fonctionnement. Ils

peuvent également améliorer leurs travaux en ajoutant des facteurs externes ou en explorant des questions laissées de côté.

## 7. Conclusion/Discussion

Dans cet état de l'art, nous nous sommes demandé comment les algorithmes d'analyse de sentiments montrent-ils que les sentiments des acteurs du marché financier influent sur l'évolution du cours des actions. Afin d'y répondre, nous avons étudié des articles de recherches existants sur ce sujet. Cela nous a permis de comparer les différents algorithmes d'analyse de sentiments appliqués aux émotions des investisseurs.

Nous avons déterminé quels algorithmes étaient efficaces dans la prédiction du cours des actions ainsi que le rôle joué par les valences. En effet, plus la valence est négative plus cela impact la bourse négativement. Cependant, cela n'est pas vérifié dans l'autre sens. Concernant l'horizon temporel, les prédictions à long terme fonctionnent davantage.

Nous avons vu également qu'il existe certaines conditions pour lesquelles l'analyse de sentiment ne fonctionne pas toujours.

De plus, nous avons expliqué l'intérêt de ces prédictions en termes de profit : nous avons constaté un meilleur rendement avec l'utilisation d'un algorithme de sentiment plutôt qu'en suivant les tendances de marché.

Du point de vue scientifique, nous avons examiné des perspectives de recherche, notamment l'élimination des erreurs produites par les algorithmes. Du point de vue industriel, les résultats ne paraissent pas pouvoir être exploités dans une banque sans risque. En effet, les stratégies de prédiction du marché utilisant l'analyse de sentiment sont encore au stade de la recherche.

Lors de la conception du mémoire, j'ai pu découvrir les deux grands domaines de l'intelligence artificielle à savoir le Machine Learning et le Deep Learning.

De plus, cela m'a permis de mieux comprendre certains aspects de la finance. J'ai eu l'opportunité de rencontrer des traders et discuter avec eux sur ce sujet ainsi que de discuter avec la personne à la tête du pôle IA de CACIB.

Les idées que j'ai trouvé intéressantes dans cet état de l'art sont le fonctionnement de l'outil Sentistrength, mais également l'apprentissage supervisé lorsque l'algorithme est capable de prendre une décision à partir des essais passés.

De plus, les différents types de Dataset étudiés dans les articles sont pertinents notamment Twitter et le service d'information Bloomberg. La difficulté a été de comprendre certains algorithmes d'analyse de sentiments complexes comme LSTM.

L'an prochain, j'aimerais effectuer une expérience sur ce sujet en se basant sur l'article 2. En effet, je souhaite récupérer des données financières d'investisseurs sur l'application Telegram. J'utiliserai ensuite le logiciel, SentiStrength (utilisé dans l'article 2) afin d'extraire le sentiment contenu dans ces données. Puis, je superposerai le cours des actions concernées par ces sentiments afin de voir si les opinions des investisseurs influent sur le marché boursier.

## 8. Glossaire

<b>Mot</b>	<b>Définition</b>
Trades	L'achat ou la vente de biens et de services.
Natural Langage Processing	NLP ou Traitement Automatique du Langage : branche de l'Intelligence Artificielle qui vise à permettre à la machine de comprendre et de reconnaître le langage humain.
Emotion différenciée	Emotion plus précise que « positive » ou « négative ».
NYSE	New York Stock Exchange, bourse de New-York
CNN	Réseau de neurone convolutif
Neutre au marché	Eviter d'être exposé par rapport au risque

## 9. Table des figures

Figure 1 : Schéma des acteurs du marché financier .....	8
Figure 2 : Classification des algorithmes de cet état de l'art en Machine Learning .....	12
Figure 3 : Tableau des articles et de leur type .....	15
Figure 4 : Performance des algorithmes .....	25
Figure 5 : Evolution du profit avec la stratégie de trading (à gauche) et le rendement du marché (à droite) [7] .....	27



## 10. Références

- [1] J. Felsen, "Learning Pattern Recognition Techniques Applied to Stock Market Forecasting," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-5, no. 6, pp. 583-594, Nov. 1975, doi: 10.1
- [2] Guresen, Erkam, Gulgun Kayakutlu, and Tugrul U. Daim. "Using artificial neural network models in stock market index prediction." *Expert Systems with Applications* 38.8 (2011): 10389-10397/109/TSMC.1975.4309399.
- [3] Sul, Hongkee, Alan R. Dennis, and Lingyao Ivy Yuan. "Trading on twitter: The financial information content of emotion in social media." *2014 47th Hawaii International Conference on System Sciences*. IEEE, 2014.
- [4] Risius, Marten, Fabian Akolk, and Roman Beck. "Differential emotions and the stock market-the case of company-specific trading." (ECIS 2015 completed research papers).
- [5] Liu, Qing, et al. "Synergy between stock prices and investor sentiment in social media." *Borsa Istanbul Review* (2022).
- [6] Sousa, Matheus Gomes, et al. "BERT for stock market sentiment analysis." *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2019
- [7] Yadav, Ritu, Ashwani Kumar, and A. Vinay Kumar. "Event Based Sentiment Analysis on Futures Trading." *The Journal of Prediction Markets* 13.1 (2019): 57-80.
- [8] Wang, Cuiyuan, et al. "Learning to trade on sentiment." *Journal of Economics and Finance* 46.2 (2022): 308-323.
- [9] Fenton-O'Creevy, Mark, et al. "Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders." *Journal of Organizational Behavior* 32.8 (2011): 1044-1061.

[10] Mehta, Yash, Atharva Malhar, and Radha Shankarmani. "Stock price prediction using machine learning and sentiment analysis." 2021 2nd International Conference for Emerging Technology (INCET). IEEE, 2021.

[11] Alexandre Chekroun, "L'utilisation des réseaux de neurones convolutifs pour fournir un opinion mining à des fins de prédiction de cours de marché " 2021, Mémoire état de l'art, Paris 1 Panthéon Sorbonne.

[12] Han, Yechan, Jaeyun Kim, and David Enke. "A machine learning trading system for the stock market based on N-period Min-Max labeling using XGBoost." *Expert Systems with Applications* 211 (2023): 118581