



Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: XXXX-XXXX



Natural Language Processing Stock Prediction Model Inclusion Innovation

Poh Soon JosephNg^{1,*}, Cheng Kian Wong², Koo Yuen Phan³, Jianhua Sun⁴

¹ Institute of Computer Science and Digital Innovation, UCSI University, Kuala Lumpur 56000, Malaysia

² Faculty of Data Science and Information Technology, INTI International University, Nilai 71800, Negeri Sembilan, Malaysia

³ Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Malaysia

⁴ School of Public Fundamentals, Jiangsu Medical College, Jiangsu 224000 China

ARTICLE INFO

ABSTRACT

In the burgeoning era of technology, Artificial Technology plays a pivotal role across various sectors, including the financial market for a responsive institution. With the implementation of AI tools, the financial market is expected to function more efficiently while simultaneously reducing costs and time. The financial industry, grappling with biases in stock analysis and limited stock prediction tools, seeks an integrated solution merging technical analysis with current information through advancements like Natural Language Processing (NLP) to enhance the accuracy and efficiency of stock trading, considering investors' preferences and time constraints. In the current manual processes, investors often spend substantial time reading articles and processing information before making decisions. This approach is inefficient, consuming excessive time and energy, thereby reducing the precious time that should be saved for personal relationships. Moreover, suboptimal decision-making could be made due to frequently gathered of inaccurate information. This research aims to discover the impact of Natural Language Processing integration with the stock prediction model on the financial market and evaluates the acceptance of the public towards the employment of NLP tools in their investment process for inclusive innovation. The evaluation will examine 4 different perspectives which are the factors that drive them to invest in the stock market, assessing the model's effectiveness, and the user experience respectively. This study utilized a mixed-method approach, which consists of quantitative and qualitative surveys. The respondents evidence the results and are being analyzed using SmartPLS, a statistical tool. Most respondents indicated a willingness to utilise NLP provided it effectively helps them achieve their financial goals and fosters positive experiences. With the implementation of NLP, respondents anticipated that NLP would significantly reduce the time for investment research and analysis. Consequently, investors could achieve higher profits while dedicating more time to their families. Besides, by identifying the trends through the textual data, NLP is expected to enhance the accuracy of stock prediction results. Thus, the potential opportunities can be uncovered with reduced downside risks.

Keywords:

Artificial Intelligence; natural language processing; big data; responsive institution; stock prediction model; influence factors; inclusive innovation

* Corresponding author.

E-mail address: joseph.ng@ucsiuniversity.edu.my

1. Introduction

The financial industry has been applying AI to replace humans from repetitive work using automation. The conversation on ensuring financial engagement at the bottom level of the pyramid in society increasingly emphasizes the importance of digital financial inclusion [18]. The shift from direct investment toward digital domains should be strictly aligned with the transformation objective to maximize effectiveness and efficiency, thus leading to heightened profitability, elevated service standards, and product quality [12]. In historical times, investors tend to ascertain the trend or pattern of the stock market based on their experience [30]. Investment banks utilized forecasting models to predict stock price fluctuations. Based on that, investment banks can minimize risk while earning greater profit margins. Some current models are useful in predicting the stock market as shown in Table 1.

Table 1

Various models that are created to predict stock prices

Model	Description
CEEMDAN-LSTM Yu, L., Yan, Y., Xu, J., Liao, Y., & Ma, F. (2021). Forecasting stock index price using the CEEMDAN-LSTM model. <i>The North American Journal of Economics and Finance</i> , 57, 101421. https://doi.org/10.1016/j.najef.2021.101421	Mixture model that Long Short-term Memory (LSTM) combines Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). Applied to forecast index price of Standard & Poor's 500 index (S&P500) and China Securities 300 Index (CSI300). This model uses decomposed data to obtain prediction sequences and subsequently reconstructs to generate the outcome. Decomposition time series data has been adopted in this model as well as neural network training, and stability evaluations, to improve the accuracy of stock price forecasting.
Fine-tuned Support Vector Regression (SVR) Dash, R. K., Nguyen, T. N., Cengiz, K., & Sharma, A. (2021). Fine-tuned support vector regression model for stock predictions. <i>Neural Computing and Applications</i> , 35(32), 23295–23309. https://doi.org/10.1007/s00521-021-05842-w	Fined-tuned to become more accurate and faster. Use the grid search technique to look for the best kernel function and optimize the parameters by training the dataset. Specifically built to analyze different performance parameters of the stock market such as daily and monthly return, cumulative return, volatility nature, and risk associated.
Deep Neural Network (DNN) Yu, P., & Yan, X. (2019). Stock price prediction based on deep neural networks. <i>Neural Computing and Applications</i> , 32(6), 1609–1628. https://doi.org/10.1007/s00521-019-04212-x	Mixture model of deep RNN and LSTM. Effective and expandable especially when dealing with time series problems. Excellent in handwriting recognition, concise natural language translation, and audio frequency data analysis.

These models are effective in predicting stock prices. Using these models, the investment banks can earn with the "buy low sell high" method as the top and bottom lines are somehow known. Besides, the models can help to reduce the time consumption and deviations during the analysis of data.

However, these models do not include market sentiments in forecasting. Sentiments are categorized as an affective state in terms of psychology [1]. Investors are still required to read and follow up with the most recent news that might impact stock price fluctuations because the models are not able to summarize the events that trigger the fluctuations. In addition, these models are using past data as input to train the models. This will lead to a certain level of inaccuracy due to incomprehensive analysis.

Therefore, this paper proposes the application of Natural language processing (NLP). Natural language processing (NLP) is a subordinate of Artificial Intelligence (AI) that enables computers to have mutual understanding as humans in text and speech. It comprises a set of computational techniques, that help the computers to do analysis automatically and representation of human language [6,13]. NLP is efficient and widely applied in natural language translation, retrieving, and extracting information, summarizing sentences, and so on [6]. There are 2 main components in NLP: computational linguistics and theoretical linguistics. Computational linguistics has been using natural language as input to create algorithms; whilst theoretical linguistics enhances language performance and grammar [6].

1.1 Background Studies

The indicator to identify the performance of the stock market is known as the stock market index [24]. The stock market is dynamic, volatile, and nonlinear and responds sensitively to various factors. In Malaysia, the stock market is also unstable as Malaysia is an emerging country. According to Siang and Rayappan [4], the emerging stocks market reflects quicker than developed countries when there is an alteration in the view of economic, social, and political. Investment banking serves to assist companies, governments, and even other business entities to fulfill financial needs via selling securities. RHB Investment Bank provides various financial trading services and stock trading. In investment banks, time series analysis and econometric models play a crucial role in predicting trends. An example of a time series model is ARIMA (Auto Regressive Integrated Moving Average) while the econometric model is Vector Autoregression (VAR) which combines multiple time series variables in the analysis.

1.2 Problem Statement

Investors might not have time to monitor the movement of stock prices or respond to any unexpected issue that happens. Therefore, time constraints stopped them from learning more and engaging in more stock trading. Based on the existing problems in the stock industry, first, we should apply an indicator that contributes effectively to stock prediction, to improve the efficiency and accuracy of stock trading. In addition, we will strengthen the intelligence of the application, in which we are going to examine the expectations and preferences of the investors. This helps to set a clearer and achievable goal (target price) according to their personality and acceptance.

Prediction of stock prices is very challenging as it depends on different parameters which makes it complicated. However, the data being applied to the prediction model is usually notable data such as high, low, open, close, volumes of shares traded, and so on. For example, Ding and Qin [29] suggested a deep network model that is effective in predicting the stocks in terms of opening price, and lowest, and highest price using past data and technical parameter data. The weakness of the prediction model is the historical data or patterns that are analyzed. In forecasting, despite only historical data needing to be considered, it also requires consideration with the current situation, to improve the accuracy of forecasting. The proposed solution is a tool that can help the application capture the keywords from the information, which is Natural Language Processing (NLP). NLP is well-known in syntax and semantics. Linguistics of NLP are separated into 2 main categories which are computational linguistics and theoretical linguistics. The algorithms are designed to use natural language as its input. In terms of syntax analysis, it is selected with its 2 main functions: determining structure and regularizing the syntax structure. With these, the application can determine the keywords of the information faster. Thus, leading to a more accurate result.

1.3 Research Hypothesis

Here the research hypothesis in this study,

H1: A better understanding of the factors has a positive correlation with investing in stocks.

Based on preliminary literature, we hypothesize that different factors would result in different investing preferences.

H2: An effective forecasting model will increase the interest of investors in applying.

Building upon existing research and the capabilities of NLP, we propose that the use of this model could significantly enhance the effectiveness of profitability in terms of higher efficiency to capture the market trend.

H3: Investors agreed the forecasting model improved their investment experience.

Considering the potential for automation and efficiency offered by NLP, we hypothesize that the implementation could significantly increase the satisfaction of investors while making investments. High satisfaction is mainly due to timesaving and accuracy.

H4: Enhanced effectiveness will lead to higher investor acceptance of the forecasting model.

This hypothesis posits that the high effectiveness provided by the integration of NLP in investment workflows will correlate directly with increased acceptance among investors. Short information collective time is preferred by the investors to pre-empt the initiatives.

H5: Positive investors' experience will genuinely increase the acceptance level of the NLP forecasting model.

The hypothesis here is that a positive investment return will result in a positive investor experience by applying the NLP forecasting model, which in turn will foster greater acceptance of such technology. It is anticipated that the ability of NLP could maximize the profit of investors. The relationship between each hypothesis is shown in Figure 1.

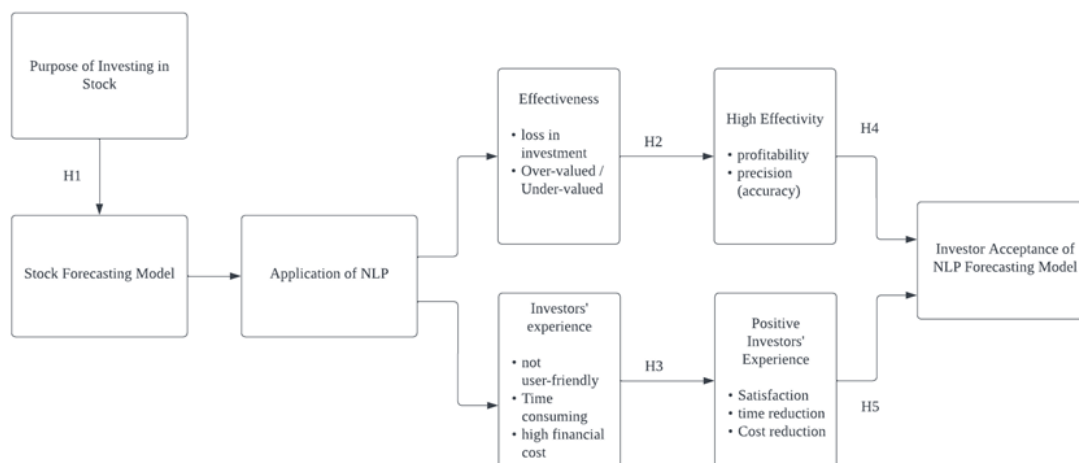


Fig. 1. Research framework diagram

2. Literature Review

The stock market is an assembly point that allows investors to trade shares from all over the world. Like other commodities, it has an open market to trade. However, the stock market is not a single global financial market, instead, the stock market exists in each country. It has the characteristics of being dynamic, variable, and not sequential [26].

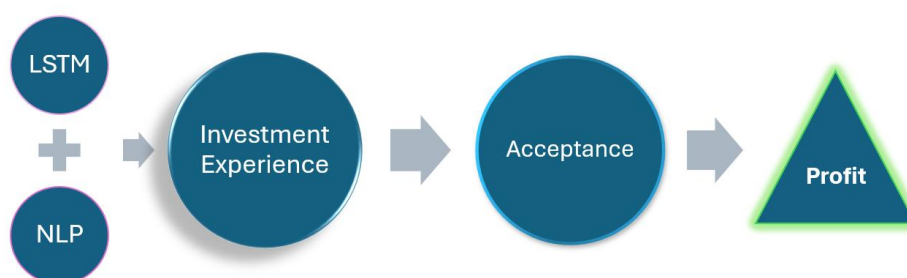


Fig. 2. Overview flowchart

The stock market is a platform where shares of public-listed companies are traded between buyers and sellers. Stock market plays a crucial role as it expedites the formation of capital, and investment, and contributes to higher economic growth. Across the world, there are various major financial centres. For instance, in the United States, the New York Stock Exchange (NYSE) is a globally famous exchange centre. Other than that, there are Nasdaq, London Stock Exchange (LSE), Hong Kong Stock Exchange (HKEX), Shanghai Stock Exchange (SSEC), as well as Bursa Malaysia. The investors' bias on the stock's market performance nowadays is due to its significant role in the economy [24].

Malaysia, as one of the emerging countries, the stock market performance is relatively more dynamic and inconsistent due to its nature of interdependency, therefore the level of volatility is higher than in developed countries [24]. However, when there is a risk, it is always accompanied by a reward. Therefore, the availability of investment banks can cope with the rising investment demand. Volatility's contribution to market stability and investor behaviour is a key concern in the stock environment [3]. Investment banks should analyze investor behaviour to predict future market

trends. With that, investment banks can make expected profits by trading in the right direction and at the right time.

There are various models implemented by investment banks to analyze stock trends and predict future stock prices. Among all the models, Artificial Neural Network (ANN) and Neural Network (NN) techniques are widely used to predict stock prices accurately [8]. Besides that, the combination of statistical techniques is popular. These statistical techniques include autoregressive integrated moving average (ARIMA) and clustering which can provide historical evidence and theories for normality postulates [8]. The Artificial Intelligence-based model usually takes past prices of stock and other variables as input and does not include news analytics [7,21]. To gain comprehensive analytics of stock price prediction, it is encouraged to delve into news analytics. In the following sections, this paper will elaborate on some of the current prediction models and the integration of NLP in textual analytics.

2.1 CEEMDAN-LSTM

The full abbreviation of the CEEMDAN-LSTM model is the combination of Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Long-Short Term Memory. Ensemble Empirical Mode Decomposition (EEMD) is known as a type of algorithm and is being proposed as an approach to adaptive time-frequency data analysis [27]. LSTM integrated a neural network with good memory performance to process and predict events. According to Lin *et al.*, [10], the authors selected the daily closing price of CSI300 and S&P500 as the input to test the prediction on decomposed data. A positive result was obtained but behind the high accuracy and robustness of the model, sentiments are ignored in the research.

2.2 Support Vector Regression

Support vector regression is one of the popular machine-learning techniques. According to Dash *et al.*, [4] (2021), this support vector regression proposal for stock prediction of time series data is a fine-tuned version. The fine-tuned here represents the application of the grid search technique in the dataset training. It can help to choose the best kernel function and optimize the parameters, and this results in to increase in the accuracy of prediction outcome. Examples of parameters could be daily return, monthly return, cumulative monthly return, volatility, and risk associated with the stock market. Dash *et al.*, [4] underwent some observations and noticed that stock forecasting is difficult due to its nature involving time series data which has highly unpredictable errors. The integration of various machine learning algorithms may result in outperformed but high requirements in hyperparameters. In addition, even though the original support vector regression model has excellent accuracy in predicting the stock. However, it was very time-consuming due to parameters issues. Therefore, the authors integrated the grid search technique to reduce the time spent by optimizing the parameters for the entire dataset.

2.3 Deep Neural Networks (DNNs)

Undeniably, one of the common issues faced by machine learning or statistical models to predict the stock price is the feature extraction process [22]. In the stock market, trends are predicted using historical data such as closing price, opening price, volume, and so on. To some extent, investors analyze the outcome of the company's financial reports. This paper encourages the application of Deep Neural Networks (DNNs) as they can handle nonlinear high-dimensional manifolds of stock data

by separating them into several abstraction layers. DNNs have 2 layers: the input layer and the output layer. The input layer serves as numerous nodes that benchmark the number of features. After going through multiple layers of the input layer, the data will be presented at the output layer. Random forest is hired in the output layer as it is an example of an ensemble method that deals with classification trees. This method works more efficiently especially when the data size is large. Forest tree is observed to help in reducing variance and the classification process assists in optimal prediction over all the trees.

Stock prediction is complex and time-consuming. Stock price predicting is an intricate work, especially in the long term [31]. The variables of distinct events might cause the stock price to fluctuate occasionally. The fluctuations are influenced by various economic factors including macroeconomics, commodity price index, sentiments of investors, political events, and so on [31]. Therefore, close tracking and observation are essential. In the old times, we had different technical analyses and indicators. For instance, the stock analysts will observe the pattern or trend using SMA (Simple Moving Average), EMA (Exponential Moving Average), Bollinger Band, RSI (Relative Strength Index), and MACD (Moving Average Convergence/Divergence) [32]. From there, they will make predictions based on the result. For example, when the MACD line crosses upward to the signal line, it is bullish in the market. However, the process does not seem to be very efficient, and the accuracy is low.

The proposed solution to predict stock prices is the integration of NLP in the stock prediction model. NLP is an advanced tool that transformed technology and business by enabling the invention of smart devices, simultaneous translation, speech-to-text recognition, sentiment-based market predictions, intelligent shopping guides, and robotic customer service attendants [6]. The application of NLP is wide and diverse, especially to enhance the efficiency of systems. NLP can conduct textual analytics, and this is useful to understand market sentiments. For example, NLP can understand emotions based on the text retrieved from social media or customer feedback analysis for business. Automation of news analysis with NLP provides stock analysts with more comprehensive information, instead of focusing on past stock prices. According to Khalil and Pipa [7], the process of collecting news information is not an easy job, especially to analyze the hidden sentiments derived from the information. Therefore, it is reasonable to apply NLP as it is efficient in converting the texts into the required input such as emotions, feelings, and sentiments [7]. In this research paper, the focus is to improve and expedite the process of Natural Language Processing (NLP) by providing a sample dataset.

2.4 Framework Literature

NLP is a crucial component of machine translation applications. It enables seamless communication across language barriers. Additionally, it can effectively summarize text and retrieve the necessary information. These capabilities are essential for leveraging the advancement of chatbots and virtual assistants. By improving interactions between machines and humans, chatbots and virtual assistants provide a bridge that enables machines to understand and respond appropriately to natural language queries.

NLP technology plays a significant role in supporting data analysis from a management perspective. By extracting meaningful insights from large datasets, NLP enables constructive decision-making at the management level. In addition, chatbots and virtual assistants powered by NLP can help companies improve their customer service by handling multiple queries simultaneously and providing faster responses. From a business standpoint, NLP can facilitate investigating and analyzing market trends, customer preferences, product satisfaction levels, and so forth.

Table 2
Framework Literature Summary

Framework	CEEMDAN-LSTM	Fine-tuned Support Vector Regression	Deep Neural Networks	LSTM and NLP
Author	Yu, L., Yan, Y., Xu, J., Liao, Y., & Ma, F. (2021). Forecasting stock index price using the CEEMDAN-LSTM model. <i>The North American Journal of Economics and Finance</i> , 57, 101421. https://doi.org/10.1016/j.najef.2021.101421	Dash, R. K., Nguyen, T. N., Cengiz, K., & Sharma, A. (2021). Fine-tuned support vector regression model for stock predictions. <i>Neural Computing and Applications</i> , 35(32), 23295–23309. https://doi.org/10.1007/s00521-021-05842-w	Yu, P., & Yan, X. (2019). Stock price prediction based on deep neural networks. <i>Neural Computing and Applications</i> , 32(6), 1609–1628. https://doi.org/10.1007/s00521-019-04212-x	Khalil, F., & Pipa, G. (2021c). Is deep learning and natural language processing transcending financial forecasting? investigation through the lens of the news analytic process. <i>Computational Economics</i> , 60(1), 147–171. https://doi.org/10.1007/s10614-021-10145-2
Problem Statement	Statistical approaches are unable to handle nonlinear, dynamic, and noisy financial time series. There is noise in the original data of stock index prices. LSTM has relatively	Irrespective of the type of learning algorithms, the time to converge to an optimal solution by using an Artificial neural network (ANN) or its variants is a major concern. Fusion or a combination of machine learning algorithms require much more hyperparameters, more memory, as well as time. Lack of fine-tuning on the parameters of support vector regression (SVR) may lead to time-consuming.	Data series change continuously within the trade range. Due to there being certain time intervals within the trading range, all financial time series cannot be treated as completely continuous data. There is a relatively significant difference in trends between series observed in different intervals.	News for individual-level stock and news information is difficult to accumulate as it comes from various sources. The process of sentiment analysis in NLP involves using raw news text.
Objective	To apply artificial intelligence methods. To improve the forecasting performance using the data with noise while forecasting financial time series. Improve the forecasting accuracy of the LSTM model.	To optimize hyperparameters (performance metrics) avoid data over-fitting and local minima issues, thus, improving the predictions of the precision.	To increase the efficiency on a time scale and prevent the occurrence of vanishing gradient issues.	Maximize the accuracy of the prediction model by including sentiment analysis.
Methodology	The empirical mode decomposition (EMD) algorithm decomposes	Grid search technique is applied over the training dataset to	Optimize the nodes to deep RNN using LSTM.	Application of hourly stock returns, News Analytics

	complex signals into Intrinsic Mode Functions (IMFs) and residuals, transforming non-linear signals into linear ones. The Ensemble Empirical Mode Decomposition (EEMD) reduces mode mixing by adding white noise, while the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) further improves by adding white noise, overcoming completeness and reconstruction errors.	select the best kernel function and optimize its parameters.		preprocessing, Naive Bayes Classifier, and sentiment index development in data preprocessing.
Contribution	The forecasting result gains higher accuracy compared to other models such as Support Vector Machines (SVM) and has great robustness.	Increases the model's overall accuracy and reducing time and memory.	Higher prediction capacity and high efficiency.	Improves accuracy and time saving. Beneficial to all institutional and individual investors, traders, and portfolio managers who dive into the equity market.
Limitation	Limited to the historical data of 2 indices which are S&P500 and CSI300. The accuracy might only be relevant to predicting the stocks that are listed in these 2 indices. Historical data does not reflect the current market situation. Sentiments may change when similar events happen again.	Prediction in closing price does not reflect the real trend of the stock market.	Sensitive to hyperparameters and market sentiments are ignored.	Waste of a lot of useful information due to matching the time of news and information. Sentiment information comes anytime but the stock exchange only works for a specific time range.
Recommendation	To include more sets of historical data retrieved from other indexes. To include the analysis of market sentiments.	To explore the potential to include more metrics especially to consider the market sentiments such as the Volatility Index (VIX).	To optimize the parameters and to comprehend the research with market sentiment analysis.	To use time zone agnostic systems. Coordinated Universal Time (UTC) can resolve the time-matching issue without geographical constraints.

2.5 Change Management

The integration of NLP aims to enhance and comprehend the stock price prediction model. Most of the prediction model is built on technical analysis and statistical analysis. However, market sentiment plays a crucial role in deciding the stock market trend. NLP brings multiple benefits to the stock forecasting model. Due to its excellence in extracting information, NLP is widely used to extract textual data from various sources such as news articles, financial reports, social media posts, and so forth. This is very helpful in analyzing and understanding the market sentiment, the mood, and opinions of the investors, and thus the management level can make informed decisions. For instance, a positive sentiment may show an uptrend of an individual stock since the buying signal is relatively strong. Since NLP is one of the advanced Artificial Intelligence tools, it requires individuals who are professional and knowledgeable in NLP to use and control. Therefore, investment banks with experienced labor may have an advantage in utilizing prediction models. In addition, having a dataset for the model to capture specific keywords would make the process easier and faster.

3. Methodology

The financial industry can benefit from integrating Natural Language Processing (NLP) into the stock prediction model better to understand the implications, effects, and outcomes. This chapter outlines the methodologies used to attain the research objectives as well as address the formulated questions. The methodology is based on:

3.1 Research Design

The research design is established based on a structured approach to provide answers to the research questions. A mixed-method approach is employed in this study to discover and understand the impact of NLP integration on the financial industry from omni-bearing.

Table 3

Research Methodology Framework

Activity/Phase	Phase 1	Phase 2
	Quantitative Generalization	Qualitative Reasoning
Research Dimension	Phenomena Explanatory Sequential Dimension	
Research Design	Random Survey	
Data Collection	Online Across Malaysia 102 respondents	
Research Methods	Convenient sampling who are willing to participate and share information	

3.2 Research Method Type

The mixed-method approach consists of quantitative and qualitative methods [23]. Quantitative data provides insights using measurable data related to the integration of NLP, interest, experience, effectiveness, as well as acceptance. On each question, there will be a 7-point Likert scale provided for selection. On the other hand, qualitative data emphasizes the understanding of investors' experiences, perceptions, and factors that drive the aspiration in stock investment that might not be accurately rated through quantitative measures alone. This requires the interviewees to answer in the form of short paragraphs that could express their feelings or opinions comprehensively. The main reason of choosing mixed-method approach is it provides a more comprehensive and detailed result.

The availability of feelings, thoughts, and behaviours data (qualitative) and the combination of measurable and numerical data (quantitative) could portrait a richer insight at a larger picture.

3.2.1 Quantitative research: survey

Aim: To gather measurable data from a large sample of financial experts, investors as well as anyone who works in the financial industry to answer the research questions.

Instrument: Structure Equation Modelling (SEM) is employed to gain insights from the quantitative data. With SEM, the relationship and impacts of datasets can be easily understood.

3.2.2 Qualitative research: interview

Aim: To gain in-depth insights into the experiences and opinions of financial experts, investors as well as anyone who works in the financial industry concerning the integration and use of NLP.

Instrument: Thematic analysis will be utilized to interpret the qualitative data. By using thematic analysis, the patterns or themes of the collected data will be observed and recognized.

3.3 Sampling and Instrument

The target audience of this survey is all individuals with an interest in the current and future state of the financial industry, especially related to Artificial Intelligence technology such as NLP. The participants include financial professionals, investors as well as individuals who engage in the financial sector. 102 respondents have responded to the questionnaire [5]. The questionnaire will be conducted using Google Forms which can be easily distributed online such as email and social media. This helps to ensure the efficiency of the data collection process and a sample that accurately reflects willing participants' opinions.

The data collection tool employed, which is the questionnaire, has been purposely designed to cover various kinds of question formats to obtain a comprehensive result. The first part of the questionnaire has been set up to gather the basic demographic information of the participants. Participants are required to answer using single-choice queries. There are 7-point Likert scale items prepared to discover and explore the participants' perceptions and acceptance of the integration of NLP. Lastly, open-ended questions are needful for the respondents to express their thoughts and opinions, which will be the qualitative responses. According to LaDonna et al., [9], open-ended questions could potentially generate greater and more comprehensive data.

3.4 Data Collection

The questionnaire is the main tool used to collect data. It is a crucial and popular tool used for collecting data and gathering information from the respondents [25]. It is expected that the participants will take approximately 10 minutes to complete the survey. The survey duration is appropriate as this is a mixed-mode questionnaire that combines quantitative and qualitative questions.

The questionnaire starts with demographic questions to collect the background of the respondents. The main part of the questionnaire is to discover the key research areas that focus on the interest, expectation, and acceptance of the respondents towards the NLP-integrated stock prediction model. The main part is using Likert-scale questions and diving deeper by using open-

ended questions. This combination allows a thorough understanding of the opinions of the respondents regarding the application of NLP.

3.5 Data Analysis Techniques

3.5.1 Unit of measurement

In this study, the 7-point Likert scale is employed to collect and measure the respondents' attitudes and perceptions toward various aspects of NLP integration with the stock prediction model in the financial industry. The 7-point Likert scale is a kind of psychometric scale and has been widely used in psychological research as well as social science. On this scale, respondents are provided with a range of options to present their opinions. There are 7 scales ranging from 1 to 7, reflecting from strongly disagree to strongly agree.

From the research and studies, it is found that the 7-point Likert scale can result in the best psychometric outcomes [2]. Unlike the 5-point Likert scale, the 7-point Likert scale offers greater granularity as it allows relatively more nuanced responses, capturing the varying degrees of responses from agreement to disagreement, as well as neutral stances. In this survey, the relationship between the scales and the degree of response is as follows: 1 – Strongly Disagree; 2 – Disagree; 3 – Somewhat Disagree; 4 – Neutral; 5 – Somewhat Agree; 6 – Agree; 7 – Strongly Agree.

3.5.3 Statistical analysis tools

After the data has been collected and preprocessed, the data file is exported to the SmartPLS tool for statistical analysis. In SmartPLS, the data will be converted into a graphical analysis which enables users to understand the complex relationships within the data. There are many reasons that SmartPLS has been chosen, however, this study will just list three of them. The first reason to highlight is that SmartPLS offers a user-friendly interface. This has greatly simplified the process of conducting complex analyses with various statistical levels. For example, graphical analysis allows the users to build and analyze the models efficiently. In addition, SmartPLS has the functionality to handle complex and unrelated datasets even if the sample size is relatively small. By using the algorithms, SmartPLS can generate different Structure Equation Models such as Cronbach Alpha Coefficient, Cross-Loading, Fornell-Larcker, and so on, for analysis purposes. Besides, SmartPLS provides comprehensive output and visualization options. These functions allow users to interpret their results better.

3.6 Instrument Development

In this study, the questionnaire is adopted from the existing questionnaires, instead of creating a new questionnaire. The reason for choosing to adopt a questionnaire is due to several considerations. The first consideration is time. Adopting questionnaires helps to save time as it does not need to begin from scratch, pilot-test, and validate the instrument. Secondly, the expertise requirement will be lower since all the valid resources must be developed and validated by the experts. Additionally, the reliability of the questionnaire is relatively higher since the adopted questionnaires have been tested and validated in previous studies. In this study, 3 articles have been selected and modified according to the research purpose.

Table 4

Reference Table for Instrument Development

No	Reference
1	Elhussein, N. H. A., & Abdelgadir, J. N. A. (2020). Behavioral Bias in Individual Investment Decisions: Is It a Common Phenomenon in Stock Markets? International Journal of Financial Research (Print), 11(6), 25. https://doi.org/10.5430/ijfr.v11n6p25
2	Information Extraction from Unstructured Big Data: A Case Study of Deep Natural Language Processing in Fintech - ProQuest. (2022). https://www.proquest.com/openview/0a7fccfdbc798ac514d956893df11810/1?pq-origsite=gscholar&cbl=18750&diss=y
3	Nugraha, B. A., & Rahadi, R. A. (2021). Analysis of Young Generations Toward Stock Investment Intention: A Preliminary Study in an Emerging Market. Journal of Accounting and Investment, 22(1), 80–103. https://doi.org/10.18196/jai.v22i1.9606

4. Findings

This chapter presents the findings from a systematic analysis of survey data regarding the position of Natural Language Processing (NLP) in the financial industry.

4.1 Data Collection Findings

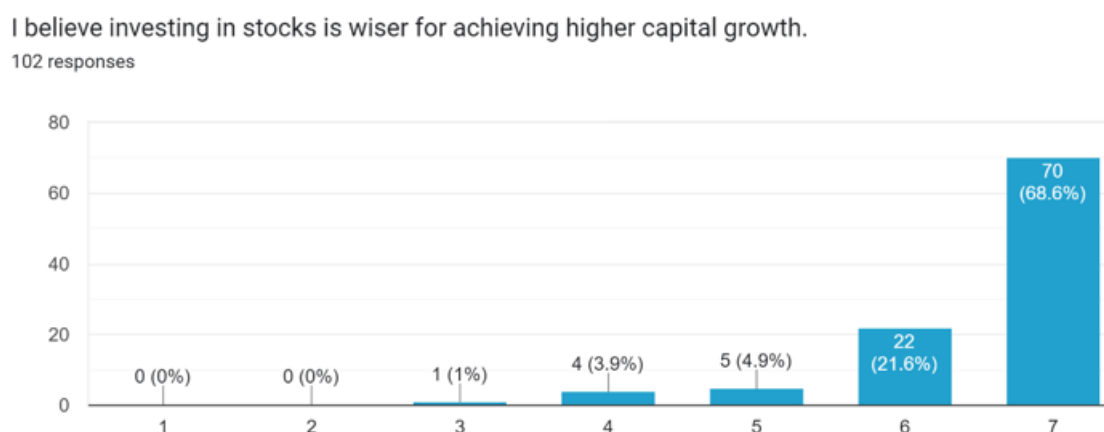


Fig. 3. Number of respondents who think investing in stocks can achieve higher capital growth

Quantitative: From the chart above, it can be observed that more than 95% of respondents agreed that investing in stocks can help them achieve higher capital growth. This stated that most respondents are willing to invest their money in the stock markets and believe it is a wise move. In contrast, 1% of respondents disagree at a relatively low level with this statement. This might be because they are not profiting from their investment.

Qualitative: To understand further whether NLP can help the respondents, the study collected the opinions from the respondents on the question of whether an NLP tool can save their time from reading news and better understand market trends. It can be concluded that most respondents agree that NLP can effectively reduce the time taken to read articles. This can be proved when respondents responded that NLP has a greater speed in reading articles as well as can handle multiple articles at one time. Some respondents also voted a 'No' as they are more confident with their justification and trust concerns towards using NLP.

You are interested in applying Natural Language Processing (NLP) if it has the functionality of easy access to information.

102 responses

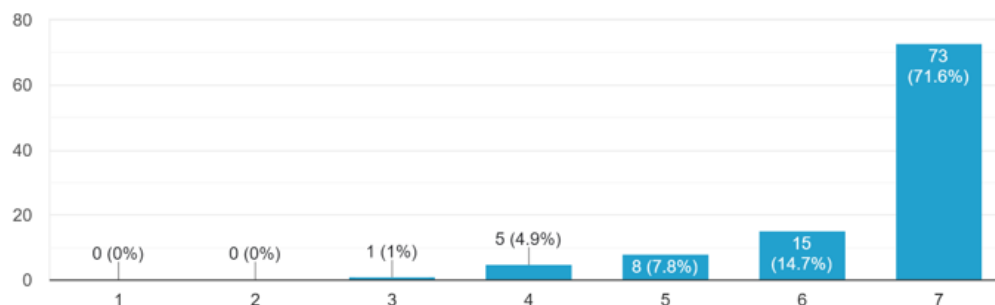


Fig. 4. Number of respondents who are interested in applying NLP if it has the functionality of easy access to information

Quantitative: 71.6% strongly agree that they will apply NLP if the NLP consists of the function of easy access to information. Overall, it reaches approximately 94% of respondents agree with this statement. Therefore, the function of easy access to information has become one of the highlights of NLP to attract users. Whereby, there are 1% and 4.9% of respondents rated slightly disagree and neutral respectively.

Qualitative: In terms of the aspects that attract the interest of respondents and motivate them to apply NLP in their investments, there are several reasons. Approximately 80% out of 102 respondents expect NLP can help them save more time in the investment analysis process and have high expectations of the accuracy of information. While there are also partial respondents who emphasize the comprehensiveness of NLP. This shows that respondents have a certain level of acceptance of NLP but based on functionality.

You can achieve better investment results by utilizing Natural Language Processing (NLP) to comprehend, generate and manipulate human language.

102 responses

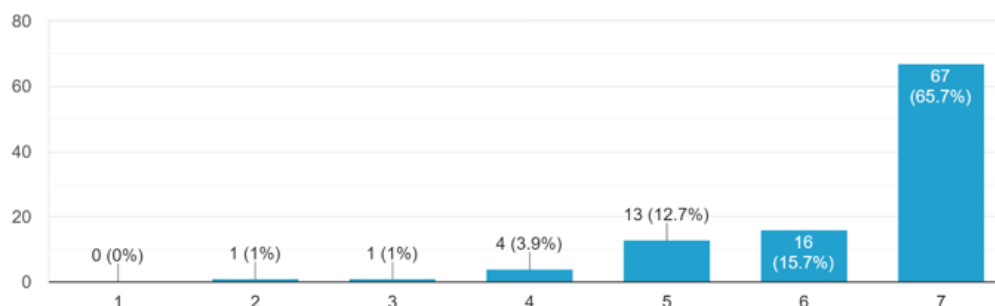


Fig. 5. Number of respondents who think NLP can help to achieve better investment results when it can comprehend, generate, and manipulate human language

Quantitative: According to the statistics of the responses to the above question, 65.7% of respondents strongly agree that NLP can improve investment results with its functionality of comprehending, generating, and manipulating human language. Overall, 94.1% of respondents hold a positive view of the improvement in investment results using NLP. There are also 2% of respondents who do not think NLP can achieve better investment results with such functionality.

Qualitative: To dive deeper into whether NLP would improve investment experience, this study investigates what improvements or benefits respondents think NLP could bring. From the collected

responses, it has been discovered that most of the respondents agree that NLP effectively serves in cost and time reduction. With the shorter time required in analysis, they can spend more quality time to accompany their family, friends, and personal time. Besides that, there are around 27 respondents who deem NLP could assist them in making more informed decisions. With that, they can achieve higher profitability in their investment return.

I am satisfied with using Natural Language Processing (NLP) when the rate of return on my investment meets my expectations.

102 responses

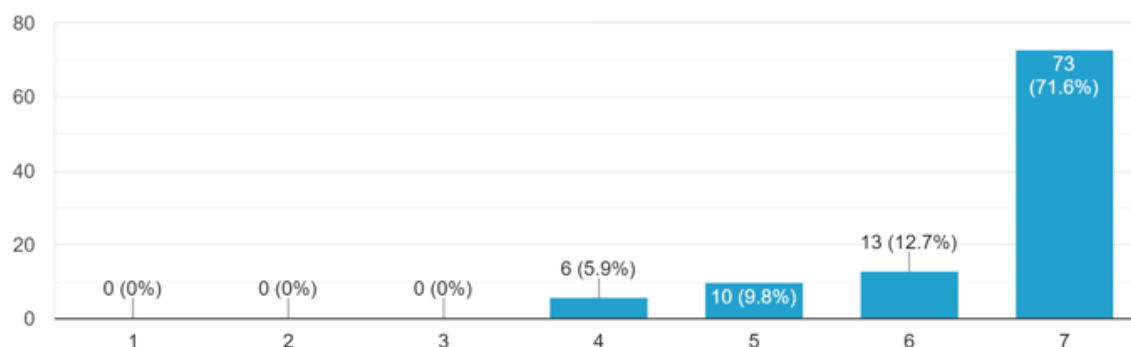


Fig. 6. Number of respondents who are satisfied with using NLP if the rate of return on investment meets their expectations

Quantitative: It is observed that 71.6% of respondents strongly agree that they will be satisfied to apply NLP if it can help them to achieve their profiting goal. To highlight from this chart, there is none of the respondents disagree with this statement. Therefore, it can be assumed that respondents emphasize and care about whether NLP could bring them a satisfactory rate of return. However, 6 respondents rated it neutral. This might be because they have doubts about the effectiveness of NLP could meet their expectations.

Qualitative: Apart from that, this study looked at the additional functions provided by NLP. Respondents have given various investment inputs or advice they are expecting. The responses can be concluded in 3 main points which respondents mainly focus on action to be taken, future trend of stock, and the precise prediction of stock price. It is expected that the forecasting model, integrated with NLP, can provide structured advice after analyzing the market sentiments.

I prefer to do my investment with Natural Language Processing (NLP) when it helps to efficiently track the market trend.

102 responses

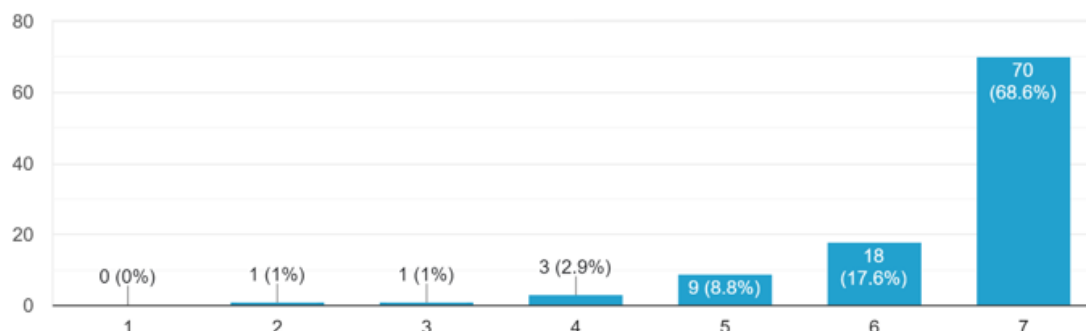


Fig. 7. Number of respondents who prefer to invest with NLP when it helps to track the market trend efficiently

Quantitative: The chart above shows that 68.6% of respondents strongly agree that they are likely to utilize NLP in their investment analysis process when the model can efficiently help track the market trend. Overall, there is a total of 95% of respondents who rated agree with this statement. While there are still 2% of respondents who do not prefer to use NLP, this might be due to reasons of insecurity, non-reliability, and so on.

Qualitative: In the last question of the questionnaire, the study explores a deeper understanding of the factors that influence the respondents to trust on NLP-integrated prediction model and the reasons why they do not trust it. As a result, most respondents (99 respondents) will trust NLP when it is effective with its functions, high accuracy, ability to make informed decisions, high reliability, and low error rate. Therefore, the performance of NLP will decide the initiative whether the respondents will apply NLP in their investment process. The other 3 respondents stated that they will refuse to apply NLP for reasons such as they are not in need currently and not fully trust AI, are worried about the error rate as the stock market fluctuates from time to time, and are the untrustworthy source of news might be included in the analysis stage, and inaccuracy results.

4.2 Structure Equation Model Analysis

Table 5
Cronbach Alpha coefficient

	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (AVE)
Acceptance	0.944	0.945	0.96	0.857
Effectiveness	0.717	0.839	0.87	0.77
Experience	0.82	0.87	0.916	0.845
Factors to Invest	0.406	0.818	0.722	0.589

From the table above, Cronbach's alpha for acceptance and experience are outperformed from the standard of 0.75. This indicates that the data collected possesses high consistency for these 2 variables. On the other hand, the effectiveness has a relatively lower Cronbach's alpha which is 0.717, however, it is still within the acceptable range. Unfortunately, factors to invest have the lowest alpha value (0.406) which is outlined in the standard. This might be due to lower internal consistency in the data collection process. To highlight from the table, it is observed that the Cronbach Alpha Coefficient value of Factors to Invest is 0.406 which is far below the bottom line, 0.70. This is mostly due to there are many investors do not strongly agree that they are inspired by their spouse, family, friends, colleagues, or attracted by the advertisement to engage in stock investment. The respondents most likely took a self-initiative into investments. This might be due to individuals being self-motivated to take investment actions when investments can fulfill their substantial needs. (Maharani & Saputra, 2021).

Table 6
Cross-Loading

	Acceptance	Effectiveness	Experience	Factors to Invest
A1	0.893	0.745	0.763	0.552
A2	0.962	0.753	0.734	0.657
A3	0.92	0.773	0.67	0.586
A4	0.926	0.735	0.67	0.586
E1	0.521	0.816	0.694	0.316
E2	0.846	0.936	0.719	0.528
EX1	0.63	0.757	0.895	0.375
EX2	0.765	0.719	0.943	0.612
F1	0.619	0.498	0.581	0.966
F2	0.329	0.209	0.118	0.496

Cross-loading shows the correlation between an item loading onto more than one factor. It is utilized frequently in factor analysis. By taking the example of the scenario above, item A1 has substantial loading with all the listed factors which are acceptance, effectiveness, experience, and factors to invest respectively. However, it can be observed that the correlation with acceptance is the highest, which means that acceptance has the strongest correlation with the A1 item.

Table 7
Fornell-Larcker Criterion

	Acceptance	Effectiveness	Experience	Factors to Invest
Acceptance	0.926			
Effectiveness	0.812	0.878		
Experience	0.767	0.797	0.919	
Factors to Invest	0.643	0.503	0.553	0.768

Fornell-Larcker criterion has been employed to assess the discriminant validity of the factors. The figures in the table represent the average variance extracted (AVE) values. This method can effectively evaluate whether the AVE values are greater than the correlation coefficients between any 2 of the items. For example, in the Acceptance column, the highest correlation is 0.926 and belongs to the Acceptance row. Therefore, the discriminant validity is supported for Acceptance.

Table 8
HTMT Discriminant

	Acceptance	Effectiveness	Experience	Factors to Invest
Acceptance				
Effectiveness	0.941			
Experience	0.861	1.052		
Factors to Invest	0.967	0.764	0.737	

Heterotrait-Monotrait (HTMT) is also one of the methods used to evaluate the discriminant validity. Like Fornell-Larcker, it distinguishes the correlation between any 2 of the items. However, HTMT discriminant compares items between different constructs (heterotrait correlations) and items within the same constructs (monotrait correlations). To calculate the average of HTMT ratios, all the values in the table are summed up and divided by the total quantity of values: $(0.941 + 0.861 + 1.052 + 0.967 + 0.764 + 0.737) / 6 \approx 0.897$. The formula to calculate the HTMT ratio of this table would require the average heterotrait correlation divided by the monotrait correlation. This is roughly equal to $0.897/1 \approx 0.897$. In this case, the monotrait correlation is assumed to be the maximum value since there is no monotrait correlation in this table. As a result, the outcome of the HTMT ratio is less than 1, which suggests that discriminant validity is supported.

Table 9
R-Square

	R-square	R-square adjusted
Acceptance	0.699	0.693
Effectiveness	0.253	0.245
Experience	0.306	0.299

R-square (R²) is a statistical measure mainly applied in regression analysis. The values contained in the table above represent the proportion of variance in each dependent variable. As shown in the table, the R-square values for acceptance, effectiveness, and experience are 0.699, 0.253, and 0.306

respectively. R2 ranges from 0 to 1, 0 is defined as the model does not fit the data well, whereas 1 indicates the dependent variable is supported by the independent variable and shows a perfect fit.

Table 10

F-Square

	Acceptance	Effectiveness	Experience	Factors to Invest
Acceptance				
Effectiveness	0.367			
Experience	0.131			
Factors to Invest		0.338	0.442	

F-square (F2) is a measure that is utilized in a regression model. The main purpose of F-square is to evaluate the effect size of individual predictors. The proportion of variances is quantified in the dependent variable that is specifically attributed to each predictor, after considering the influences of other predictors. In the context of the table above, there is no F-square value observed for the variable "Acceptance", whilst the F-square value for the variable "Effectiveness" is 0.367. This indicates that the variance falls by approximately 36.7%. Therefore, in this case, it means that there is a 36.7% variance in "Effectiveness" supported by the predictor(s) within the model, after including the considerations of other predictors. Similar to R-square, the maximum value of F-square is 1 and the minimum is 0. The higher the value of the F-square, the stronger the effects of predictors on the dependent variable.

Table 11

Hypothetical Structure Acceptance

Hypo	PLS Paths	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (O/STDEV)	2.50%	97.50%	P values	Hypothesis Accepted
	Effectiveness ->								
H1	Acceptance	0.55	0.533	0.135	4.08	0.243	0.768	0	Yes
	Experience ->								
H2	Acceptance	0.329	0.347	0.14	2.353	0.104	0.64	0.019	Yes
	Factors to Invest								
H3	-> Effectiveness	0.503	0.508	0.086	5.825	0.328	0.662	0	Yes
	Factors to Invest								
H4	-> Experience	0.553	0.549	0.105	5.264	0.32	0.726	0	Yes

Hypothetical Structure Acceptance is employed in this study to identify the robustness between NLP and the collected data. Hypotheses below in the table have been tested within the model and shown a result that the proposed model is supported by the data.

In the table below, the study shows the results of 4 hypotheses. In H1, it is observed that the path coefficient (Original Sample) is 0.55. This reflects a positive relationship between effectiveness and acceptance. The relationship between these two variables is significant as the T-statistics show 4.08 and a P-value of 0 (which is less than 0.05). Therefore, it can be concluded that when the effectiveness of the model increases, the acceptance will increase accordingly. H2 reveals a strong bond between experience and acceptance as the path coefficient is shown as 0.329. T-statistics is 2.353 and the P-value is 0.019, slightly higher than H1. However, if the P values are still below 0.05, it is considered statistically significant. Therefore, a better experience will lead to an increase in acceptance. In addition, H3 has a 0.503 path coefficient which is similar to H1. The relationship between factors to invest and effectiveness is statistically significant. This means that certain factors are positively related to the effectiveness. While H4 also has a similar path coefficient which falls at 0.553. This

implies a positive relationship between the factors to invest and experience. T-statistics and p-value are 5.264 and 0 respectively, considering a statistically significant relationship.

Overall, all these 4 variables are interconnected, and they statistically have a significant relationship in between. With the indication of the result below, it shall help to understand the structure of the model and potential decision-making. All hypotheses are accepted as the p-value is less than 0.05, showing that the relationship between the variables is supported by strong evidence.

5. Discussion

The study of NLP integration contributes comprehensively concerning theoretical, practical, managerial, societal, and sustainability.

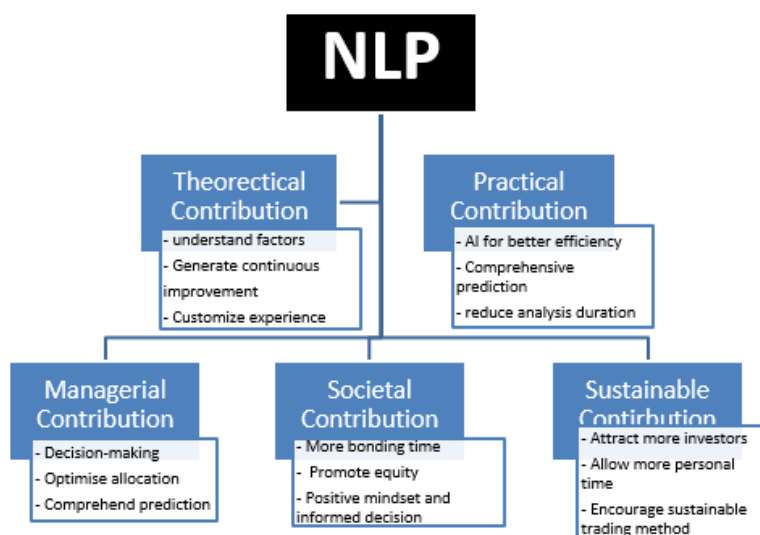


Fig. 8. Contribution breakdown of NLP

5.1 Theoretical contribution

In the dimension of theoretical, the study dives into exploring the level of accuracy of the NLP-integrated stock prediction model from various perspectives. It is specifically focused on:

- Understanding the factors of investors investing in stock.
- Generating continuous improvement using advanced technology in the investing environment.
- Customizing investment experience to match investors' preferences.

The theoretical contribution of this study provides a holistic view of the integration of NLP in the stock prediction model and its impact on investors. By taking account of the investors' benefits, researchers can focus on developing more robust and comprehensive prediction models to improve profitability in the high-complexity stock market. These AI-integrated prediction models can elevate the experience, interest, and acceptance of investors in stock markets.

5.2 Practical contribution

By observing the survey results, most of the respondents agreed that the integration of NLP into the stock prediction model is beneficial to them. This can deliberately explain that:

- Embody digitalization, automation, artificial intelligence, and efficiency have made positive progress within the financial industry.

- The integration provides comprehensive prediction to elevate the investment performance.
- Investors' demands in reducing analysis duration are well-addressed.

The practical contribution allows the improvements to nourish the investors' experience, thus increasing their intention and motivation to invest more in stock markets. In such cases, Fintech holds a fixed weightage in the financial industry and is steadily growing.

5.3 Managerial contribution

The implementation of NLP brings value to the investors. In terms of managerial contribution, it is focused on insights, recommendations, or implications. This will define how the integration of NLP helps the investors to make informed decisions and achieve their ultimate goals. It is observed to:

- Be a supportive tool for investment decision-making.
- Optimize the allocation of investment funds.
- Comprehend the prediction of stock price.

NLP allows investors to make better decisions with less time consumption. This improves the investment process to be more efficient and effective. With that, investors have ample opportunity to opt for the most suitable investment plan based on the information and funds.

5.4 Societal contribution

The impact of NLP integration could broaden to the societal level. Adopting NLP technology promotes high value of benefits to society. This is resulted in:

- Allow more time for the investors to bond with their family and enhance relationships.
- Promoting investment fairness and equity.
- Planting a positive mindset and making informed decisions before real investments take place.

At the societal level, NLP can effectively help to enhance the relationships between different parties such as families, friends, colleagues, and so on. This is benefiting from the time saved using NLP. The bonding can directly or indirectly improve personal communication skills rather than spending time collecting information. In addition, there are equal opportunities for different levels of investors, especially those individuals. Nevertheless, a positive investing environment can help to reduce the possibility of bankruptcy, thus reducing criminal cases.

5.5 Sustainable contribution

It is crucial to use strong encryption methods like AES in personal social media platforms because of their positive social impact. This initiative addresses growing concerns about the compromise of sensitive information and vulnerability to data breaches in the digital age by upholding privacy and data protection. By establishing a robust encryption framework, user communications and data are kept secure, potentially reducing the risks of unauthorized access. This action promotes user confidence in the platform's dedication to secure communication and data protection, which fosters user trust. Such a societal contribution improves the overall security environment and encourages responsible data handling practices, which ultimately makes the internet safer for all users [60-64].

5.6 Integration and Value Creation

In terms of sustainable contribution, this study dives into the perspectives of economic growth, health, and well-being as well as social equity delivered by NLP. These include:

- NLP attracts more investors to invest in stock markets, thus more capital used in business expansion, leading to more job creation.
- With the benefit of time-saving, it allows more personal time to release time and spend time to tie relationships.
- NLP encourages a sustainable trading method in which all investors deserve equity in possessing complete information.

6. Conclusions and Limitations

In this study, the main objective is to explore and discover the acceptance of investors in the NLP-integrated stock prediction model. Natural Language Processing (NLP) stands as a paramount AI tool renowned for its efficacy in data collection and processing. Its absence would necessitate investors to devote extensive periods towards information gathering and analysis, potentially spanning hours or even days. This prolonged duration raises the risk of missing time-sensitive opportunities, compromising investment decisions. Furthermore, the demanding nature of these analytical tasks may encroach upon personal time, imposing emotional strain and hindering relationships with loved ones. Such strains could subsequently impact investors' judgment and decision-making abilities, underlining the pivotal role NLP plays in streamlining processes and preserving work-life balance.

The research dives deeper into the correlation of different factors which are the effectiveness of the model and the investors' experience towards the acceptance level. Besides, it also investigates the factors that drive investors' intention to invest in stock markets. This project employed Google Forms to collect survey responses from the targeted respondents and utilized the SmartPLS tool to analyze the results. Based on the findings, all the factors positively correlate with the Cronbach Alpha Coefficient. The least positively correlated item belongs to factors to invest; however, it still displayed a correlation coefficient of 0.406. This shows that the effectiveness and experience of investors are the most valued and impactful in generating high acceptance rates in the application of NLP. If NLP has a higher effectiveness, the acceptance rate would be higher. The same goes for investors' experience, the more positive the experience from investors, the higher the acceptance level.

This research identifies several limitations that may impede the effective application of Natural Language Processing (NLP). Firstly, the challenge of market efficiency poses a significant obstacle, as the rapid integration of news and information often surpasses NLP's processing capabilities, limiting its ability to consistently outperform the market. Secondly, ethical concerns arise from the potential for NLP algorithms to inadvertently analyze misleading news or rumors, resulting in erroneous assessments and adverse effects on investment performance. This could cause breach of privacy and data insecurity to happen. To address these limitations, future enhancements should focus on augmenting NLP functionality with robust fake news or rumor filtering mechanisms, thereby enhancing the accuracy and reliability of outcomes. It is suggested to establish ethical guidelines as boundaries to act as a firewall and ensure the development process is conducted responsibly.

In future research, it should focus on mitigating the delays in information reception. The global sharing and dissemination of real-time information can enhance the efficiency of Natural Language Processing (NLP) systems. This enables NLP to surpass traditional investment methods. However, to overcome the time constraints is challenging. One potential approach is to centralize and consolidate all the information on a single platform before distributing to other platforms, such as social media.

Acknowledgment

This research was not funded by any grant.

References

- [1] Aggarwal, Divya. "Defining and measuring market sentiments: A review of the literature." *Qualitative Research in Financial Markets* 14, no. 2 (2022): 270-288. <https://doi.org/10.1108/QRFM-03-2018-0033>
- [2] Aybek, Eren Can, and Cetin Toraman. "How many response categories are sufficient for Likert type scales? An empirical study based on the Item Response Theory." *International Journal of Assessment Tools in Education* 9, no. 2 (2022): 534-547. <https://doi.org/10.21449/ijate.1132931>
- [3] Bhowmik, Roni, and Shouyang Wang. "Stock market volatility and return analysis: A systematic literature review." *Entropy* 22, no. 5 (2020): 522. <https://doi.org/10.3390/e22050522>
- [4] Dash, Ranjan Kumar, Tu N. Nguyen, Korhan Cengiz, and Aditi Sharma. "Fine-tuned support vector regression model for stock predictions." *Neural Computing and Applications* 35, no. 32 (2023): 23295-23309. <https://doi.org/10.1007/s00521-021-05842-w>
- [5] Fryett, James J., Andrew P. Morris, and Heather J. Cordell. "Investigation of prediction accuracy and the impact of sample size, ancestry, and tissue in transcriptome-wide association studies." *Genetic epidemiology* 44, no. 5 (2020): 425-441. <https://doi.org/10.1002/gepi.22290>
- [6] Kang, Yue, Zhao Cai, Chee-Wee Tan, Qian Huang, and Hefu Liu. "Natural language processing (NLP) in management research: A literature review." *Journal of Management Analytics* 7, no. 2 (2020): 139-172. <https://doi.org/10.1080/23270012.2020.1756939>
- [7] Khalil, Faisal, and Gordon Pipa. "Is deep-learning and natural language processing transcending the financial forecasting? Investigation through lens of news analytic process." *Computational Economics* 60, no. 1 (2022): 147-171. <https://doi.org/10.1007/s10614-021-10145-2>
- [8] Kumar, Deepak, Pradeepta Kumar Sarangi, and Rajit Verma. "A systematic review of stock market prediction using machine learning and statistical techniques." *Materials Today: Proceedings* 49 (2022): 3187-3191. <https://doi.org/10.1016/j.matpr.2020.11.399>
- [9] LaDonna, Kori A., Anthony R. Artino Jr, and Dorene F. Balmer. "Beyond the guise of saturation: rigor and qualitative interview data." *Journal of Graduate Medical Education* 13, no. 5 (2021): 607-611. <https://doi.org/10.4300/JGME-D-21-00752.1>
- [10] Lin, Yu, Yan Yan, Jiali Xu, Ying Liao, and Feng Ma. "Forecasting stock index price using the CEEMDAN-LSTM model." *The North American Journal of Economics and Finance* 57 (2021): 101421. <https://doi.org/10.1016/j.najef.2021.101421>
- [11] Lobe, Bojana, David Morgan, and Kim A. Hoffman. "Qualitative data collection in an era of social distancing." *International journal of qualitative methods* 19 (2020): 1609406920937875. <https://doi.org/10.1177/1609406920937875>
- [12] Machkour, Badr, and Ahmed Abriane. "Industry 4.0 and its Implications for the Financial Sector." *Procedia Computer Science* 177 (2020): 496-502. <https://doi.org/10.1016/j.procs.2020.10.068>
- [13] JosephNg, Poh Soon, Zhaoji Fu, Rui Zhang, and Koo Yuen Phan. "The Impact and Acceptance of Large Language Models in Healthcare: A Perspective from China." *Journal of Advanced Research in Applied Sciences and Engineering Technology*.
- [14] JosephNg, P. S. "Hotel room access control: an NFC approach ecotourism framework." *Journal of Science and Technology Policy Management* 15, no. 3 (2023): 530-551. <https://doi.org/10.1108/JSTPM-10-2021-0153>
- [15] JosephNg, Poh Soon. "EaaS Optimization: Available yet hidden information technology infrastructure inside medium size enterprise." *Technological Forecasting and Social Change* 132 (2018): 165-173. <https://doi.org/10.1016/j.techfore.2018.01.030>
- [16] Maharani, Amalina, and Farhan Saputra. "Relationship of investment motivation, investment knowledge and minimum capital to investment interest." *Journal of Law, Politic and Humanities* 2, no. 1 (2021): 23-32. <https://doi.org/10.38035/ilph.v2i1.84>
- [17] Maharani, Amalina, and Farhan Saputra. "Relationship of investment motivation, investment knowledge and minimum capital to investment interest." *Journal of Law, Politic and Humanities* 2, no. 1 (2021): 23-32. <https://doi.org/10.38035/ilph.v2i1.84>
- [18] Mhlanga, David. "Industry 4.0 in finance: the impact of artificial intelligence (ai) on digital financial inclusion." *International Journal of Financial Studies* 8, no. 3 (2020): 45. <https://doi.org/10.3390/iifs8030045>

- [19] Mishra, Puneet, Alessandra Biancolillo, Jean Michel Roger, Federico Marini, and Douglas N. Rutledge. "New data preprocessing trends based on ensemble of multiple preprocessing techniques." *TrAC Trends in Analytical Chemistry* 132 (2020): 116045. <https://doi.org/10.1016/j.trac.2020.116045>
- [20] JosephNg, P. S., Z. C. EricMok, K. Y. Phan, J. Sun, and Z. Wei. "Mitigating Social Media Cybercrime Revolutionising with AES Encryption and Generative AI." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 46, no. 2 (2024): 124-154. <https://doi.org/10.37934/araset.46.2.124154>
- [21] JosephNg, Poh Soon, Raenu Kolandaisamy, Li Dan, Tang Xiang, Lei Meng, Wang Yi Peng, Koo Yuen Phan, and Felix Pasila. "Policy Development and Application of an Ai-Based Fitness Posture Recognition System: A Cost-Effective Approach Using Cnn and Lstm Models." *Available at SSRN* 5159265.
- [22] Roy, Sanjiban Sekhar, Rohan Chopra, Kun Chang Lee, Concetto Spampinato, and Behnam Mohammadi-ivatlood. "Random forest, gradient boosted machines and deep neural network for stock price forecasting: a comparative analysis on South Korean companies." *International Journal of Ad Hoc and Ubiquitous Computing* 33, no. 1 (2020): 62-71. <https://doi.org/10.1504/IJAHUC.2020.104715>
- [23] Neshor Shoshan, Hadar, and Wilken Wehrt. "Understanding "Zoom fatigue": A mixed-method approach." *Applied Psychology* 71, no. 3 (2022): 827-852. <https://doi.org/10.1111/apps.12360>
- [24] Siang, Chan Ching, and Patricia Rayappan. "A study on the effect of macroeconomic factors on stock market performance in Malaysia." In *E3S Web of Conferences*, vol. 389, p. 09037. EDP Sciences, 2023. <https://doi.org/10.1051/e3sconf/202338909037>
- [25] Rath, Trupti, and Bindu Ronald. "Questionnaire as a tool of data collection in empirical research." *Journal of Positive School Psychology* (2022): 7697-7699.
- [26] Vijh, Mehar, Deeksha Chandola, Vinay Anand Tikkiwal, and Arun Kumar. "Stock closing price prediction using machine learning techniques." *Procedia computer science* 167 (2020): 599-606. <https://doi.org/10.1016/j.procs.2020.03.326>
- [27] Wu, Zhaohua, and Norden E. Huang. "Ensemble empirical mode decomposition: a noise-assisted data analysis method." *Advances in adaptive data analysis* 1, no. 01 (2009): 1-41. <https://doi.org/10.1142/S1793536909000047>
- [28] Yu, Pengfei, and Xuesong Yan. "Stock price prediction based on deep neural networks." *Neural Computing and Applications* 32, no. 6 (2020): 1609-1628. <https://doi.org/10.1007/s00521-019-04212-x>
- [29] Ding, Guangyu, and Liangxi Qin. "Study on the prediction of stock price based on the associated network model of LSTM." *International Journal of Machine Learning and Cybernetics* 11, no. 6 (2020): 1307-1317. <https://doi.org/10.1007/s13042-019-01041-1>
- [30] Strader, Troy J., John J. Rozycki, Thomas H. Root, and Yu-Hsiang John Huang. "Machine learning stock market prediction studies: review and research directions." *Journal of International Technology and Information Management* 28, no. 4 (2020): 63-83. <https://doi.org/10.58729/1941-6679.1435>
- [31] Nabipour, Mojtaba, Pooyan Nayyeri, Hamed Jabani, Amir Mosavi, Ely Salwana, and Shahab S. "Deep learning for stock market prediction." *Entropy* 22, no. 8 (2020): 840. <https://doi.org/10.3390/e22080840>
- [32] Nti, Isaac Kofi, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. "A systematic review of fundamental and technical analysis of stock market predictions." *Artificial Intelligence Review* 53, no. 4 (2020): 3007-3057. <https://doi.org/10.1007/s10462-019-09754-z>
- [33] Elhussein, Nawal Hussein Abbas, and Jarel Nabi Ahmed Abdelgadir. "Behavioral bias in individual investment decisions: Is it a common phenomenon in stock markets." *International Journal of Financial Research* 11, no. 6 (2020): 25. <https://doi.org/10.5430/ijfr.v11n6p25>
- [34] Nugraha, Bagus Aditya, and Raden Aswin Rahadi. "Analysis of young generations toward stock investment intention: A preliminary study in an emerging market." *Journal of Accounting and Investment* 22, no. 1 (2021): 80-103. <https://doi.org/10.18196/jai.v22i1.9606>
- [35] Raparathi, M., S. B. Dodda, S. R. B. Reddy, P. Thunki, S. Maruthi, and P. Ravichandran. "Advancements in natural language processing-A comprehensive review of AI techniques." *Journal of Bioinformatics and Artificial Intelligence* 1, no. 1 (2021): 1-10.