Comparative Analysis of Language Models and Machine Learning Models in Predicting Market Volatility During Financial Crises

Student name: Hengbo Huang

Student ID number: 2910004

Supervisor: Dr. Ioannis Psaradellis

A dissertation submitted in partial fulfilment of the requirements for the master's degree MSc in Financial Technology

Wordcount:11447

Adam Smith Business School
University of Glasgow

22nd August 2024

Abstract: In this paper, advanced language models (FinBERT, TextBlob, and VADER) are used to analyze the sentiment of the news during Covid-19 in the U.S. The results of the analysis of the different language models are used as training parameters and applied to the machine learning models (SVM, LSTM, and TCN) that are widely used for financial forecasting. After that, the volatility of the U.S. stock market during the 2020-2022 Covid-19 period is predicted and analyzed. The results of this study show that the combination of SVM with FinBERT and VADER (at lag 1) performs best in short-term and volatile market environments, with minimal error and excellent fit (R^2 0.99). The test of significance results for the p-value is less than 0.05 and for the F-statistic is about 800, which supports the robustness of this combination and indicates that the SVM model is statistically significant. These findings suggest that SVM models perform well in capturing market characteristics in short-term and volatile financial crisis environments and are able to predict market volatility accurately. This study provides new insights and references for researchers to select appropriate prediction models and language models to predict market volatility in financial crisis periods.

TABLE OF CONTENTS

1. Introduction

Predicting stock market volatility has been one of the core research topics in finance. Stock market volatility reflects the degree of fluctuation in market prices. For investors, significant market volatility often means uncertainty and risk, which may undermine their confidence in the market and lead to a reduction in stock investments. Especially for risk averse investors, sharp fluctuations in market returns may have a large negative impact. Therefore, accurately predicting stock market volatility can enhance investor confidence and help investors avoid blind decisions. In addition, stock market volatility is often viewed as an important indicator of macroeconomic health. Significant market fluctuations often signal economic instability or the occurrence of major events. Investors and economists can analyze market volatility to assess the economic situation and formulate appropriate countermeasures. For example, the recent Covid-19 pandemic has also triggered the financial crisis and market volatility. There is a research shows that the outbreak and spread of the COVID-19 virus during the COVID-19 epidemic in the United States had a profound effect on stock market volatility, leading to dramatic market swings and widespread economic panic (Albulescu, 2020). Therefore, it is essential to be able to predict market volatility in advance in order to deal with possible economic uncertainties and risks.

Stock market volatility is closely related to investor sentiment. When the market is strong, positive sentiment usually drives more individual investors to participate in trading, and this behavior has a significant impact on market prices (Karlsson, Loewenstein, and Seppi, 2009). On the contrary, when the market is depressed, negative sentiment tends to lead to low investor confidence, which in turn reduces stock investments. Therefore, the role of investor sentiment in the stock market is crucial. The study by Seo and Kim (2015) further demonstrated that investor sentiment has a major impact on the option market in predicting stock market volatility, especially in terms of persistence and magnitude of the prediction. Therefore, sentiment analysis is indispensable for predicting market volatility. In recent years, an increasing number of studies have used news sentiment analysis as a key factor in predicting market volatility. These studies show that the emotional content of news can significantly influence investors' decision-making behavior and market volatility. High levels of pessimism in the news media may lead to a downward trend in market prices, an effect that is particularly pronounced in small-cap stocks (Tetlock, 2007). In addition, Liang and his team (2020) showed that sentiment analysis of social media and Internet news media can effectively improve the prediction accuracy of the previous day's market volatility. These research indicate that news sentiment analysis plays a key role in predicting market volatility.

Exploring sentiment in texts has always been a complex subject. Sentiment analysis is usually accomplished by cross-referencing words in a text with sentiment lexicons that categorize words as positive, negative, or neutral. The overall emotional tendency of the text is based on the calculation of the number of emotions labeled in the text. However, Loughran and McDonald (2014) noted that the traditional sentiment lexicon was not applicable to financial texts, so they rewrote the finance-specific sentiment lexicon, which is now one of the most commonly used sentiment lexicons in finance. Nowadays, more studies use language models such as TextBlob and VADER to analyze the sentiment of texts and provide sentiment scores. BERT is also a language model introduced by Google that enhances sentiment analysis by analyzing based on sentence structure. Based on this, Araci (2019) fine-tuned the BERT model using a large number of financial texts to develop the FinBERT model, which is specialized for the financial domain. Today, FinBERT, TextBlob, and VADER become the most popular tools in the field of sentiment analysis and are widely used in research.

Machine learning has been a top topic in the field of financial forecasting, especially in stock market and volatility forecasting. Initially, researchers mainly used shallow machine learning models, which usually have only one or two feature extraction or transformation layers such as linear regression and support vector machine (SVM) (Xu et al., 2021). SVM models have performed very well in some experiments. For example, Tay and Cao (2001) found that SVM showed significant advantages in financial time series forecasting by comparing a multilayer back-propagation (BP) neural network with an SVM model. On the other hand, deep learning models have shown excellent performance in financial forecasting, especially when dealing with nonlinear dynamics in large datasets. For example, LSTM models are widely used in financial forecasting. It can effectively solve the problems of data forgetting and gradient explosion, and is suitable for processing large financial datasets and making accurate predictions (Wang et al., 2024). However in volatility forecasting, Lei, Zhang and Song (2021) find that the TCN model outperforms the LSTM model and the traditional generalized autoregressive conditional heteroskedasticity (GARCH) model, and the TCN's forecasting results are more accurate and robust.

Although there has been a significant amount of research comparing different machine learning models, research on stock market volatility, especially during financial crises, is still relatively limited. In addition, while it is common to make comparisons between language models, the research that combines and systematically compares sentiment analysis with machine learning models is still relatively few. Especially during the financial crisis, when the market was highly volatile, the performance of these models could be quite different. In order to fill this research gap, this study aims to use sentiment analysis results from different language models on U.S. news to train various machine learning models. The goal is to predict financial market volatility during Covid-19 in the United States and identify the best-performing combination. In this paper, the machine learning models SVM, LSTM and TCN and the language models FinBERT, TextBlob and VADER are used. This paper will construct several model combinations and systematically compare their performance using various evaluation metrics such as MSE, RMSE, R^2 and p-value. This paper hopes to analyze and identify the combination of models with the highest predictive accuracy, thus providing a more effective tool for predicting market volatility during financial crises.

The results of the study show that the combination of SVM, FinBERT and VADER in lag 1 performs best in predicting stock market volatility with the highest R^2 value (0.99) and the lowest prediction error. In addition, the combination has a significant p-value and the F-statistic close to 800. This finding suggests that this combination can be the most effective predictor of market volatility and also provides important support for risk management and decision-making during financial crises. The models that make predictions based on the results of FinBERT and VADER's sentiment analysis perform better than the other language model combinations. The predictive performance of the model combinations that contain TextBlob is improved after the results of the sentiment analysis of TextBlob and market volatility show Granger significant causality at lag 2. These results present investors and researchers with more detailed results on the comparative predictive ability of language models and machine learning models for market volatility during financial crises, providing a more reliable tool for effectively responding to possible future financial crises.

2. Literature Review

2.1 Sentiment analysis and market volatility

Many studies have used media and news sentiment to analyze and predict market volatility and found that news sentiment can positively or negatively influence market volatility. Antweiler and Frank (2004) found that messages on Internet stock message boards can help predict market volatility, confirming the relationship between online messages, trading volume and yield volatility. Tetlock (2007) examines the effect of pessimism in the news media on the market and finds that unusually high or low levels of pessimism can have a short positive effect on market volume. Sabherwal, Sarkar and Zhang (2011) further investigate that traders' sentiment indexes were positively correlated with current day stock returns, but showed negative correlations for returns in the coming days. Li et al. (2014) used machine learning and two lexicons to analysis sentiment of financial news and emphasized that sentiment analysis does help to improve the accuracy of predictions for stocks. By applying the MRS-FIEGARCH model to the RavenPack Dow Jones News Analytics database for sentiment analysis, the researchers derive more details on the specific impact of news sentiment on market volatility. Shi and Hos' (2021) results suggest that negative news may increase the state of high market volatility, while positive news, on the contrary, significantly reduces this possibility. These studies have confirmed that sentiment analysis has a strong link to market volatility and stock prices and can be used to make predictions about market volatility and stock prices. The impact of sentiment on volatility is more significant during crises. Moreover, sentiment is more predictive of realized volatility during a crisis (Maghyereh and Abdoh, 2022). Consequently, during financial crises, negative sentiment in the financial media is usually higher than the normal, which makes its correlation with market volatility become more pronounced.

2.2 Sentiment analysis methods

One of the most frequently used methods of sentiment analysis is lexicon-based sentiment analysis. Lexicon-based sentiment analysis is performed by identifying words and phrases in the text that contain specific sentiment. Lexicons characterize words and phrases, which are usually categorized as positive, negative, and neutral. This is done by matching the words and phrases in the text against the feature descriptions in the lexicon, and then counting and calculating the overall text sentiment (Kearney and Liu, 2014). In this way, the emotional tendencies in a text can be systematically assessed. Previously the most commonly used lexicons in the study were the Harvard IV-4 Dictionary and the text analysis program DICTION (Todd, Bowden, and Yashar Moshfiqi, 2024). However, it has been found that just annotating words and phrases with sentiment may not accurately reflect the sentiment of the whole sentence. For example, this approach is unable to correctly understand conflicting sentiment words or negative expressions in a sentence (Fan et al., 2022). In addition, currently commonly used sentiment lexicons such as Harvard IV-4 or Loughran-McDonald are not specifically designed for the financial domain. This situation leads to the fact that in financial text analysis, the sentiment tendency of many specialized terms does not match the labeling in the common lexicons, thus affecting the accuracy of sentiment analysis (Mishev et al., 2020). Therefore, Loughran and McDonald (2011) built a lexicon specifically for finance to better analyze finance-related textual sentiment. Loughran and McDonald (2014) also reclassified about 83% of the positive words and 70% of the negative words misclassified in DICTION according to financial semantics, resulting in words that corresponded more accurately to sentiment. TextBlob and VADER are both commonly used lexicon-based language models. TextBlob is a Python-based natural language processing library that scores the sentiment and subjectivity of words in text (Ahuja and Dubey, 2017). TextBlob is used in many areas of sentiment analysis and performs well in all of these areas. Ranjan (2019) used TextBlob to analyze sentiment in financial blogs and found a significant correlation between predicted and actual stock prices based on TextBlob sentiment analysis results. A study on American Airlines tweets showed that training with TextBlob sentiment analysis results improved the performance of deep learning models more significantly than using VADER and Afinn (Aljedaani et al., 2022).

Another language model that is commonly used for textual sentiment analysis is the VADER. VADER is a lexicon and rule-based sentiment analysis tool that specializes in handling sentiment analysis for a wide range of texts, especially in social media (Pano and Kashef, 2020). VADER also has an excellent performance in the financial sector. Prameswari Ekaputri and Akbar (2022) use the Vader model combined with a financial lexicon to analyze the sentiment of company news with an accuracy of 73%. Pano and Kashef (2020) investigated the relationship between the sentiment scores of Twitter texts and the price of Bitcoin during COVID-19 and found that the results of sentiment analysis using VADER can effectively reflect the movement of the price of Bitcoin in the short term.

Nowadays there are more studies are using machine learning based sentiment analysis methods, as machine learning is able to better isolate sentence structure and thus extract the sentiment in the sentence more accurately. Machine learning-based sentiment analysis generally has higher performance and accuracy. Kotelnikova and his team's research (2022) shows that the machine learning-based RuBERT model recognizes positive and message text at 0.5228 and 0.6654, compared to 0.2008 and 0.3395 for the lexicon-based approach. Frankel, Jennings and Lee (2021) found that machine learning methods are more consistently able to capture the sentiment of disclosures when analyzing complex documents such as 10-K, and that lexicon-based methods can struggle over time to capture the emotional content of 10-K messages. As a result researchers have shifted their approach to sentiment analysis more towards the study of machine learning based methods. In recent years researchers have been optimizing deep models for sentiment analysis. Deep learning is more effective than traditional machine learning methods in analyzing the semantics of a sentence, mainly because it captures the relationships between words in a sentence so that the meaning between word sequences is not lost. Deep models not only identify key words, but also understand how those words interact in a sentence to better capture the overall meaning of the sentence (Sohangir et al., 2018). There are many language models based on deep learning, such as Google's BERT model. BERT stands for Bidirectional Encoder Representation, and its model is based on the Transformer architecture, which is characterized by the ability to consider the dependencies of individual words on other words in the input sequence (Vaswani et al., 2017; Devlin et al., 2018). Araci (2019) introduced FinBERT, an improved model for financial sentiment analysis based on the BERT model. FinBERT is trained by a finance-specific lexicon to more accurately analyze the sentiment of a number of specific words in the financial domain, with a 15% increase in accuracy over BERT. Another study showed that FinBERT outperformed other machine learning algorithms and was more accurate in recognizing positive and negative emotions (Huang, Wang and Yang, 2022). Several studies have also used FinBERT to predict financial markets, they combined FinBERT with machine learning models such as LSTM and SVM for stock price prediction, and these studies have shown that the addition of FinBERT significantly improves the accuracy of the models in predicting stock prices (Kim, Kim and Choi, 2023; Liu, Leu and Holst, 2023). These studies show that FinBERT is one of the most advanced and accurate financial language models on the market today.

2.3 Mechanical Learning Models

There are three machine learning models are selected in this paper, which are : Support vector machines (SVMs), long-short-term memory networks (LSTMs), and temporal convolutional networks (TCNs). These three machine learning models that are commonly used in financial market forecasting and are often used in conjunction with sentiment analysis results.

Among the machine learning methods for predicting financial time series, support vector machine is one of the commonly used models. Support Vector Machine (SVM) is a powerful machine learning model widely used in classification and regression tasks. The core design of SVM is structural risk minimization, which helps to reduce the overfitting problem of the model (Kim, 2003). The advantage of SVM in preventing overfitting makes it particularly popular in financial market analysis. Huang, Nakamori and Wang (2005) discuss the ability of SVM in predicting the direction of financial market movements and show that SVM perform well with financial time series data. SVM have the excellent predictive ability compared to other forecasting methods. Kim (2003) study applies SVM to forecast stock price index, which also indicates that SVM has great potential for financial time series forecasting. A study by Ren, Wu, and Liu (2019) shows that SVM combined with sentiment analysis was used to predict the direction of movement of the SSE 50 Index, showing high prediction accuracy which up to 89.93%. These studies highlight the effectiveness of combining advanced machine learning techniques with sentiment analysis in the field of financial forecasting.

In the field of financial time series prediction, Long Short-Term Memory (LSTM) networks have become an important deep learning tool. Due to its excellent performance in predicting time series and the ability of memory units to determine the retention and forgetting of information over an arbitrary length of time, LSTM are particularly well suited for solving the gradient vanishing problem in traditional neural networks (Bao, Yue and Rao, 2017). LSTM networks have a better performance than standard deep networks and logistic regression in predicting complex sequences (e.g., financial market returns) and can effectively learn and extract information from time-series data (Fischer and Krauss, 2018). Many studies have also combined sentiment analysis with LSTM prediction. Ko and Chang (2021) used BERT for textual sentiment analysis and LSTM for combining historical stock trading information to predict stock prices. Amrita Ticku and her team (2023) used a Long Short-Term Memory network (LSTM) combined with sentiment analysis of stock market news to predict future stock prices. The results show that this combination outperforms traditional stock prediction methods in predicting stock prices.

These studies show that sentiment information in stock market news has significant predictive value for stock market dynamics.

TCN is a neural network architecture designed for time-series data that uses causal convolution to ensure that data after the current point in time does not influence the prediction and has a longer memory time. It also has lower computational complexity than recursive architectures (Bai, Kolter and Koltun, 2018). Guo and his team (2022) proposed a stock price prediction model based on an improved time-series convolutional network (TCN), which is better than the common time-series prediction models such as LSTM and GRU in prediction accuracy and response speed. Bai, Kolter, and Koltuns (2018) also showed that TCNs are clearer and better than typical recurrent networks such as LSTMs for a wider range of standard tasks. Zhang and Wang (2024) combined wavelet transform and TCN) models with sentiment analysis of stock news. The results of the study showed that by incorporating TCNs into the language model, it was able to predict the movement of stock data more effectively, with a 5.92% increase in accuracy compared to before.

2.4 Comparison of language models and machine learning models

Liapis, Karanikola, and Kotsiantis (2023) evaluated the performance of 30 machine learning models for time series prediction. Among the six key metrics (i.e. MAPE, R^2, and RMSLE), Temporal Convolutional Networks (TCNs) performed the best in three metrics, better than GRU_FCN and LSTM_FCN. Meanwhile, STM_FCN and LSTMPlus also demonstrated excellent performance, with LSTMPlus being in the top three in all five metrics. In addition, the study also tested the effectiveness of FinBERT, Vader and TextBlob for sentiment analysis, and the results showed that TextBlob and Vader outperformed FinBERT and contributed more to the prediction results. Shapiro, Sudhof and Wilson (2017) explored a variety of methods for text sentiment analysis and applied them to a large corpus of economic and financial news articles. This study found that VADER's performance in predicting the sentiment of news articles was comparable to that of a combined vocabulary model, even though it was not specifically designed for the financial domain. In contrast, the BERT model performs better, but BERT suffers from a more serious "black box" problem and is less transparent and interpretable. There are also some studies that compare the predictive ability of different machine learning models in more detail. Lakshminarayanan and McCrae (2019) compared the efficacy of Support Vector Machines (SVMs) and Long Short-Term Memory Networks (LSTMs) for stock price prediction. The results of the study showed that on the basic stock price dataset, SVM and LSTM each performed well in terms of their predictive ability, but on the combined dataset, LSTM outperformed SVM. Gopali and his group (2021) compared RNN-based LSTM and CNN-based TCN models and find that TCN is slightly better in performance and faster in training. However, most studies focus primarily on comparisons of stock forecasting ability rather than market volatility, this may cause machine learning models to display differences in forecasting performance

3. Methodology

3.1 Research Design

The aim of this study is to investigate the effectiveness of different language models (FinBERT, TextBlob, VADER) combined with machine learning models (SVM, LSTM, TCN) in predicting U.S. market volatility during Covid-19 epidemic. The news collection is a collection of U.S. financial news about Covid-19 for the period January 2020 through December 2022 through the LexisNexis and Factiva databases. These news data are sourced from multiple news organizations to ensure diversity and representativeness. The next step involves natural language processing of the news text. The specific cleaning steps include removing special symbols (e.g., @, #, etc.), deleting redundant text (e.g., excessively long blank text and meaningless strings), and removing formatting information (e.g., author names, word count statistics, and copyright notices) from the news. This paper performs text preprocessing also includes tokenization, stop word removal, lemmatization, stemming, removing punctuation and text normalization. However, it is worth noting that FinBERT can perform sentiment analysis based on whole sentences, so preprocessing steps that would affect whole-sentence analysis (e.g., removing punctuation and removing stop words, etc.) are not applicable to FinBERT new texts. After these processes, the text was categorized and stored by month for following sentiment analysis and model training.

After completing the text preprocessing, this experiment used three language models (FinBERT, TextBlob, and VADER) to analyze the sentiment of the news texts. The results of the FinBERT and VADER models included scores for positive, neutral, and negative sentiment. TextBlob's analysis resulted included scores for the subjectivity and polarity of the news. The results of the sentiment analysis were summarized on a monthly basis and the average monthly sentiment score was calculated. Also, in order to explore the relationship between the sentiment analysis results and market volatility, this paper examines the correlation between the sentiment analysis results and the U.S. volatility index VIX using Granger causality tests. The results of the Granger causality test will provide a reference basis for the adjustment of the input feature weights of the subsequent machine learning model.

In the pre-training stage of machine learning, this study combines the results of sentiment analysis with the US S&P 500 Total Return Index and the S&P 500 Composite Price Index as the input data for the machine learning model, and the VIX Volatility Index as the target variable of the model. The dataset is chronologically divided into training, validation, and testing sets with a ratio of 8:2, i.e., 80% of the data is used to train the model, 10% is used to validate the model performance, and the remaining 10% is used to test the predictive performance of the model. There are three main machine learning models for which time is conducted in this paper: SVM, LSTM, and TCN. These models have different structures and properties that capture linear and nonlinear relationships in market data. The specific settings of these three models are: the SVM has an RBF kernel chosen for its kernel function, with a penalty parameter of C=1e3; the LSTM model is set up with two hidden layers containing 32 neurons in each layer and uses L2 regularization and Dropout techniques to reduce overfitting; and the TCN model employs a three-layered convolutional structure, with convolutional kernel size of 2 in each layer, and also uses Dropout regularization. The model also uses 5-fold cross-validation and early stopping method to prevent model overfitting during model training.

Finally, after model training was completed, predictions are made on the test set data and the prediction results for each model are recorded. The predictive effectiveness of different combinations of language and machine learning models is evaluated by comparing the models' prediction error metrics (e.g., mean square error MSE, root mean square error RMSE, mean absolute error MAE, and coefficient of determination R^2) as well as statistical significance tests (e.g., p-value and f-statistic). The best performing combination was identified by analyzing the prediction error and statistical significance.

3.2 Dataset

This paper collects financial news from the U.S. The primary databases for news sources are LexisNexis and Factiva. The news sources cover The New York Times, Business Wire, The Wall Street Journal, USA Today, and Release Wire these well-known financial news organizations to ensure that the news sources are comprehensive. The news selection is all about the U.S. financial markets and the financial impact of the Covid-19 epidemic news, spanning January 2020 through December 2022, covering the main epidemic period. All data is sorted by date and month and contains a total of 5,627 articles. Additionally, to better predict volatility in the U.S. market, this paper uses the S&P 500 Total Return Index and the S&P 500 Composite Price Index for the U.S. The S&P 500 Total Return Index accurately reflects the actual investment returns of the top 500 U.S. stocks. It includes price changes and reinvested dividends, providing a more comprehensive view of market returns and helping to predict future market volatility. The S&P 500 Composite Price Index can reflect short and medium-term market trends. Therefore, when constructing the market volatility prediction model, the price index provides an important input feature for model training. Combining the above sentiment analysis results with the S&P 500 Total Return Index, the forecasting model can both more accurately predict future U.S. market volatility. The closing price (CLOSE) and 20-day moving average were used as training parameters for these two indicators in the experiment. 20-day moving averages smooth out the data by eliminating the noise of short-term price fluctuations. In addition, the 20-day moving average of the CBOE S&P 500 Volatility Index (VIX) is chosen as the actual predictive control for the experiment. The data for the three indicators are collected over the period from January 2020 through December 2022 with a daily frequency.

3.3 Model Selection

3.3.1 Language model

For the selection of models for sentiment analysis, three models that have been widely used in financial text sentiment analysis are selected for this experiment: FinBERT,TextBlob and Vader.

FinBERT is a model specialized for financial text analysis, based on the BERT architecture. FinBERT uses the same architecture in network structure as the native BERT released by Google, including a 12- or 24-layer Transformer structure. However, FinBERT uses a more financially relevant training corpus and uses a more sophisticated pre-training model, which enhances its performance in the financial domain. Compared with the BERT model, FinBERT improves on the pre-training model, which mainly consists of two pre-training methods: word-level pre-training and task-level pre-training. FinBERT introduces financial whole-word blocking technology, which ensures that when blocking a word, it blocks the entire word, not just a part of the word (Huang, Wang, and Yang, 2022). This helps the model to learn the full semantics of the word better rather than distracting on sub-words and also helps the model to understand the meaning of the word in context. In addition, FinBERT not only focuses on complete words, but also considers sub-word or character-level features of words. This is because financial texts contain a large number of specialized terms, abbreviations and unique expressions. By pre-training on the financial dataset, the FinBERT model is better able to handle these specific linguistic phenomena (Kim, Kim and Choi, 2023). Task-level pre-training, on the other hand, is designed to allow the model to better learn the financial domain knowledge at the semantic level. FinBERT also introduces two types of supervised learning tasks, which are the industry classification of research reports and the financial entity recognition task of financial news. In addition, FinBERT contains six types of self-supervised pre-training tasks: 1. range substitution, 2. capitalized word prediction, 3. word-paragraph prediction, 4. sentence position recovery, 5. sentence distance, and 6. conversational relations (Liu et al., 2021). By fine-tuning and multitasking learning on specific financial tasks, FinBERT is better able to perform tasks such as sentiment analysis, information extraction, and risk assessment. Through better analysis of sentences and more precise processing of finance-related word meanings, FinBERT has a superior performance in sentiment analysis in finance compared to the BERT model. This makes FinBERT a powerful tool in financial text analysis tasks. When using FinBert for sentiment analysis, FinBERT returns a ternary array containing the text's positive, negative and neutral sentiment scores, the higher the score, the higher the corresponding sentiment, and the sum of the 3 scores is usually 1.

TextBlob is a sentiment analysis tool based on a sentiment lexicon and pattern sentiment analysis engine (Sohangir, Petty and Wang, 2018). It analyzes the sentiment polarity and subjectivity of text by decomposing text data into separate lexical items and comparing them with positive and negative words in the lexicon (Asderis, 2022). Sentiment polarity is a floating-point value ranging from -1 to 1, where -1 indicates a strong negative sentiment and +1 indicates a strong positive sentiment. Subjectivity is a floating-point value ranging from 0 to 1, where a value closer to 1 indicates a more subjective sentence and a value closer to 0 indicates a more objective sentence (Illia, Eugenia and Rutba, 2022). TextBlob determines the overall sentiment tendency by rating each word and calculating the average of these scores. Due to its simplicity and powerful functionality, TextBlob is widely applied in social media and other text analysis fields.

Similar to TextBlob, VADER (Valence Aware Lexicon and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool. It is based on a lexicon of words with predefined sentiment scores. VADER returns positive, neutral and negative scores for textual sentiment analysis (Illia, Eugenia & Rutba, 2022). Each word in the VADER lexicon is assigned a sentiment score that ranging from -2 to +2 (ÇILGIN et al., 2022). Negative numbers indicate negative sentiment, positive numbers indicate positive sentiment, and scores close to zero indicate neutrality. VADER is unique in its ability to recognize and understand the sentiment impact of punctuation, capitalization, degree modifiers, and negatives (Asderis, 2022). VADER is widely used in a variety of sentiment analysis scenarios, including analyzing tweets, comments, and other forms of social media content. It is also popular in academic research and is often used for tasks such as financial market sentiment analysis and customer feedback analysis.

3.3.2 Machine Learning model

For the mechanical learning models for predicting the volatility of the U.S. financial markets, three models that are widely used for forecasting in the financial sector are also selected for this experiment:

Support Vector Machine (SVM) is a supervised learning binary classification model that performs particularly well in classification tasks. SVM constructs one or more hyperplanes by picking those sample points, called support vectors, that are closest to the classification boundaries to find the optimal separating planes in the high-dimensional feature space for efficient classification (Kurani et al., 2021). The core is to maximize the classification interval (margin maximization) thus improving the generalization ability of the model. The optimization problem of SVM can be expressed as:

$$\min_{w,b} \frac{1}{2} ||w||^2 \tag{1}$$

Satisfying constraints:

$$s.t. y_i(w^T x_i + b) \geq 1, i = 1,2, \dots, n \tag{2}$$

The above formula is the basic form of support vector machine, through this strategy, SVM can significantly improve the generalization ability of the model. In equation 1 and 2 w is the weight vector, b is the bias, x_i is the training sample, which is the sentiment analysis parameter and the U.S. stock market index in this experiment, and y_i is the sample label, which is the index VIX related to the volatility of the U.S. stock market in this experiment. The learning algorithm of the SVM is optimized by solving a convex quadratic programming problem, which makes it possible to effectively maximize the intervals between the different categories in the classification process. At the same time, real data usually cannot be completely categorized accurately, and some noise exists. Therefore, SVM introduces the concept of soft margin (soft interval) to make the sample classification less strict, and introduces a relaxation factor:

$$\min_{w,b} \frac{1}{2} ||w||^2 + C \sum_{i}^{n} \xi_i \tag{3}$$

$$s.t. y_i(w^T x_i + b) \geq 1, i = 1,2,\ldots,n.$$

$$\xi_i \geq 0, i = 1,2,\ldots,n$$

where ξ_i is the slack variable and C is the regularization parameter. In order to deal with nonlinear data, SVM introduces a kernel function (kernel function) that maps the data to a high-dimensional space in order to enable the construction of a linearly differentiable hyperplane in the high-dimensional space. With the kernel function trick, SVM is able to perform not only linear classification, but also handle tasks such as nonlinear classification, regression analysis, and outlier detection. In addition, SVM uses the principle of Structural Risk Minimization (SRM), which improves the generalization ability of the model by simultaneously minimizing the empirical error and model complexity (Tay and Cao, 2001). Due to its excellent learning ability and wide application prospects, SVM algorithms have been widely used in many fields, including face recognition, image classification, text categorization, and stock market prediction.

Long Short-Term Memory (LSTM) model is an improved version of Recurrent Neural Networks (RNN) designed to solve the gradient vanishing and gradient explosion problems of traditional RNNs (Siami-Namini and Akbar Siami Namin, 2018). LSTM introduces three key gating mechanisms: an input gate, a forgetting gate, and an output gate, as well as a cell state to achieve effective management of information and capture of long-term dependencies. The input gate controls the storage of new information i_t is the activation value of the input gate, W_i is the weight of the input gate, b_i is the bias, σ is the sigmoid function, the formula is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

The forgetting gate manages the updating or forgetting of old information, f_t is the activation value of the forgetting gate, W_f is the weight of the forgetting gate, and b_f is the bias:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (5)$$

In the candidate memory cell state, $\tilde{c}_t$ is the candidate memory cell, W_c is the weight of the candidate memory cell, and b_c is the bias:

$$\tilde{c}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

Update Memory Cell State (Cell State) combines the outputs of the forget gate and input gate to update the memory cell state:

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{7}$$

c_t is the state of the memory cell at the current time and c_{t-1} is the state of the memory cell at the previous time step.

The output gate, on the other hand, determines the output information for the current time step. o_t is the value of the output gate, W_o is the weight of the output gate, and b_o is the bias:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$

The final hidden state combines the memory cell state and the output of the output gate:

$$h_t = o_t * tanh(c_t) \tag{9}$$

Through the cell state and these three gating mechanisms, LSTM is able to save, read, reset, and update information over long periods of time (Cao, Li and Li, 2019). In this way, LSTM can efficiently capture and retain long-time dependencies, overcoming the limitations of traditional RNNs in dealing with long-time sequence data. Since financial market data are usually time series data with strong time dependencies, LSTMs are widely used in financial forecasting, such as stock price forecasting and market volatility forecasting. The capabilities of LSTM make it an important tool for financial analysis and decision making, providing investors and researchers with a more accurate forecasting and analysis tool.

Temporal Convolutional Network (TCN) builds on the key improvements of Convolutional Neural Network (CNN) to enhance its performance in sequence modeling tasks. First, TCN uses a Causal Convolution structure to ensure that the output at time point t depends only on previous input data, thus avoiding utilizing information from future times in prediction. Secondly, TCN extends the sensory field of the convolutional neural network by introducing Dilated Convolution without increasing the number of convolutional layers. The formula for the Dilated Convolution is:

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d\cdot i} \qquad (10)$$

where F(s) is the forecast data and in this experiment is the U.S. market volatility data VIX, x is the input sequence and in this experiment is the result of kernel sentiment analysis of the U.S. stock indices, d is the expansion rate, k is the convolution kernel size, and f is the convolution kernel. Dilated convolution enables TCN to effectively capture data dependencies over a longer time horizon while maintaining efficient computational performance (Dai, An and Long, 2022). In addition, TCN uses Residual Block to mitigate the gradient vanishing problem, which significantly improves the training stability and performance of the model. Finally, the parallel processing capability of TCN enables it to parallelize time series data into vectors, which significantly accelerates the training speed and improves the computational efficiency (Fu and Xiao, 2022). These features make TCN a powerful tool in the fields of financial market forecasting, risk management, and investment decision making.

3.4 Performance Test Indicators

The error indicators used in this paper to measure the ability of language models (FinBERT, TextBlob, and VADER) and machine learning techniques (SVM, LSTM, and TCN) in predicting market volatility are the following four: Mean Squared Error, MSE; Root Mean Squared Error, RMSE; Mean Absolute Error, MAE; R^2, Coefficient of Determination. First is the Mean Squared Error MSE, a measure of the regression problem:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{11}$$

MSE is the average of the sum of the squares of the differences between the true and predicted values and is used to detect deviations between the predicted and true values of the model. In equation 11, y_i is the actual value, in this experiment it is the data of VIX, $\hat{y}_i$ is the result of the prediction of the machine model, and n is the number of samples. A smaller MSE means that the model predicts a smaller error, and the model has a better performance.

Root Mean Squared Error, RMSE：

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

RMSE is the Root Mean Square Error, it calculates the sample standard deviation of the difference between the predicted and actual values and gives a more intuitive reflection of the prediction error. The lower the RMSE, the better the model and its predictions.

Mean Absolute Error, MAE：

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{13}$$

MAE is the average of the absolute errors of the actual and predicted values, a smaller MAE means that the model predicts a smaller error and the model performs better.

Coefficient of Determination, R^2:

$$\mathrm{R}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{14}$$

R^2 is the coefficient of determination, which reflects the degree of model fit. $\bar{y}$ is the mean of the actual values, which in this experiment is the mean of VIXD. R^2 ranges from 0 to 1, the closer its value is to 1, the better the variables of the equation explain y and the better this model fits the data.

The statistical significance test used in this study as well, consists of two main metrics: the F-statistic and the p-value. The F-statistic is used to assess the significance of the prediction results of the machine learning model. The calculation of the F-statistic is based on the sum of squares of regression (SSR) and the sum of squares of residuals (SSE). The formular of F-statistic is:

$$F = \frac{MSR}{MSE} \tag{15}$$

Where MSR stands for mean square of regression and MSE stands for mean square of residuals. The larger the value of F-statistic, the greater the explanatory power of the model for the dependent variable and the higher the significance of the model.

The p-value is a probability calculated based on the F-statistic to test the statistical significance of the model. P-value indicates the probability of observing the current outcome or a more extreme outcome if the null hypothesis is true. The null hypothesis (H_0) in this study is that all the input features in the model (sentiment analysis results and financial market indicators) do not have a significant effect on the target variable (VIXD). P-value is computed using the following formula:

$$p = P(F > F_{observed} | H_0) \tag{16}$$

where$F_{observed}$ is the observed F-statistic. H_0 is the null hypothesis. If the p-value is small (usually less than 0.05), the null hypothesis can be rejected, indicating that the model is statistically significant. In this case we can assume that the machine learning model has a strong ability to explain the US market volatility index VIX.

4. Results and Discussions

4.1 News sentiment analysis results

The main goal of this experiment is to predict the volatility of the U.S. stock market, and the experiment consists of two main parts. The first one is the sentiment analysis of 3 language models on the US financial news and Granger causality tests are performed on the results. Based on the results of the test it is possible to better show the correlation of the results of the different language models on the U.S. stock market volatility index (VIXD), thus determining the importance of the different sentiment analyses results.

In the sentiment analysis phase, this paper begins with a sentiment analysis of the news using FinBERT. The FinBERT model will return positive, neutral and negative sentiment scores for the text.

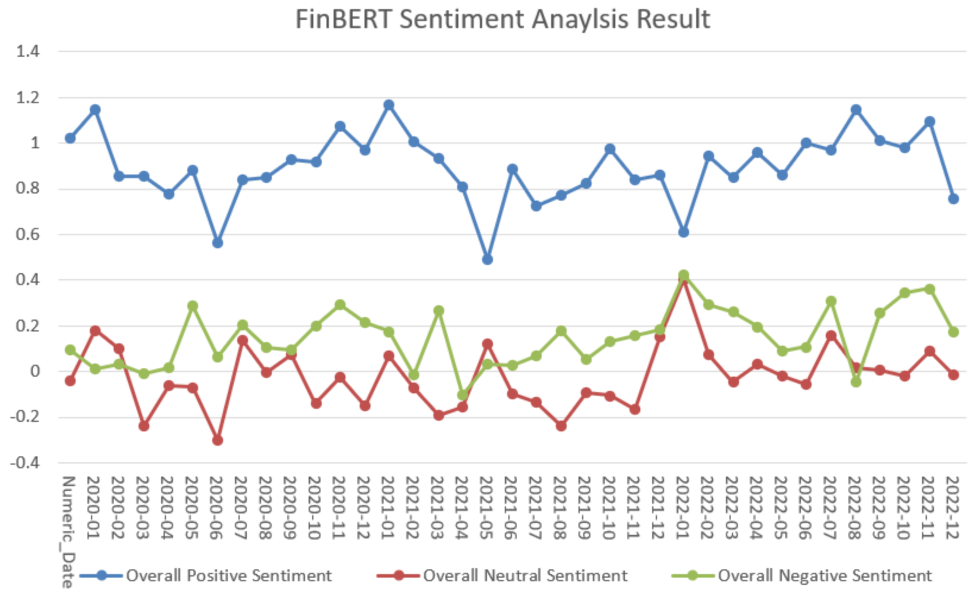

Figure 1 FinBERT model time-series results of news sentiment analysis for the US financial markets from January 2020 to December 2022

Figure 1 shows the results of sentiment analysis from January 2020 through December 2022 using the FinBERT model. The figure includes time series data for overall positive sentiment, overall neutral sentiment, and overall negative sentiment. As shown, overall positive sentiment is higher than negative and neutral sentiment. Positive sentiment is more volatile, with an overall downward trend. Neutral sentiment is more stable, remaining between 0.2 and 0.4, with a few points in time where there are large fluctuations, such as in early 2022. The negative sentiment index, on the other hand, is the lowest and slightly volatile, with higher negative sentiment in the early stages (2020), approaching the level of 0.2, but declining significantly after 2021, remaining below 0.1.

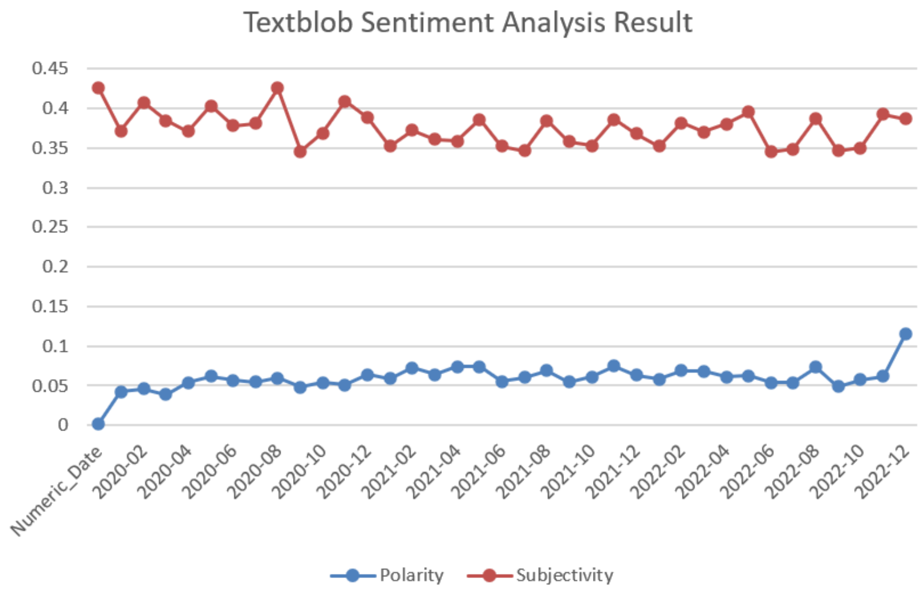

Figure 2 TextBlob Model Time Series Results for U.S. Financial Market News Sentiment Analysis for January 2020 to December 2022

Figure 2 presents the results of sentiment analysis using the TextBlob model. The figure includes time-series data on Sentiment Polarity and Subjectivity. The overall volatility presented by the Textblob data is low, with Sentiment Polarity remaining relatively low from 2020 to 2021, maintaining small fluctuations between 0 and 0.1, and peaking at the end of 2022. Subjectivity, is relatively stable, indicating that the subjectivity of market sentiment does not change much and stays at a relatively stable level most of the time.

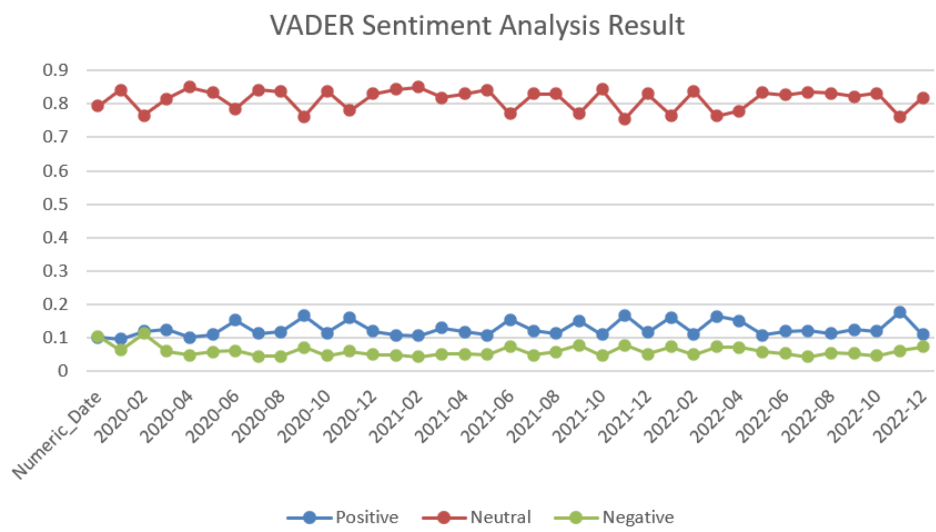

Figure 3 Time Series Results of VADER Model for Analyzing News Sentiment in U.S. Financial Markets from January 2020 to December 2022

Figure 3 presents the results of the sentiment analysis from January 2020 to December 2022 using the VADER model. Figure 3 includes time series data for Positive sentiment (Positive), Neutral sentiment (Neutral) and Negative sentiment (Negative). Positive and Negative sentiment are relatively stable throughout the time period and have small fluctuations, indicating that there are no dramatic changes in overall market sentiment. Neutral Sentiment remains high throughout the time period, indicating that the market is mostly on the sidelines or neutral.

The results of the three sentiment analyses show the highest score for news neutral sentiment. The FinBERT and TextBlob results show higher volatility in sentiment, indicating that the market sentiment experienced significant changes during the Covi-19 period, while the VADER model shows a more stable sentiment. There are some differences in the results of sentiment analysis among the three models.

4.2 Granger Causality Test

In order to test the results of different sentiment analysis language models, this paper uses the Granger causality test to analyze the results of sentiment analysis for FinBERT, TextBlob and VADER models, specifically including the sentiment variables (Positive, Neutral, Negative, Polarity, Subjectivity) at different lags. Granger causality test can test whether the data of one time series can predict another time series. Here it can be used to test whether the results of sentiment analysis can predict the market volatility the smaller the p-value, the stronger the predictive ability (Mittal and Goel, 2011).

	Min P-Value							
	FinBERT			TextBlob		VADER		
Lags	Positive	Neutral	Negative	Polarity	Subjectivity	Positive	Neutral	Negative
1	0.1144	**0.0005**	0.7750	0.2853	0.4037	0.4634	0.7217	0.0862
2	0.2072	**0.0059**	0.9067	**2.2247E-08**	**0.0105**	0.1960	0.9666	**0.0190**

3	0.6897	**0.0064**	0.3308	0.0780	0.5859	0.8557	0.2348	**0.0194**
4	0.1999	**0.0081**	0.0630	0.9545	0.6716	0.0911	0.2045	0.6348
5	0.3004	**0.0036**	0.0571	0.9522	0.7364	**0.0234**	**0.0461**	0.4887

Table 1. The Granger Causality Test Result of FinBERT, TextBlob VADER models with VIX stock market index.

The results of FinBERT's sentiment analysis show that the p-value of positive and negative sentiment is greater than 0.05, indicating that there is no significant Granger causality of positive as well as negative sentiment on market volatility. The p-values for neutral sentiment lags 1 to 5 are 0.0005, 0.0059, 0.0064, 0.0081, 0.0036, respectively, which are all less than 0.05 and have significant Granger causality. The TextBlob results show that the p-values of sentiment polarity and subjectivity at lag 2 are significantly less than 0.05 and have a significant Granger causality on market volatility. The results of the VADER show that the p-values of positive and neutral sentiment in lag 5 are 0.0234 and 0.0461, which are less than 0.05, and the p-values of negative sentiment in lags 2 and 3 are 0.0190 and 0.0194, which also less than 0.05 and have significant Granger causality on market volatility.

From the results, when in 2-period lag, the most significant causal relationships are observed, including FinBERT's neutrality, TextBlob's polarity and subjectivity, and VADER’s negativity, all of which show significance. These indicators demonstrate a certain ability to predict market fluctuations. Therefore, this experiment will use lag 1 and lag 2 to make predictions and compare the results of the experiment.

4.3 Machine Learning Model Volatility Prediction Result

The second part of this experiment is to test which models are effective in predicting market volatility during a financial crisis by synthesizing data using US market index data, volatility indices, economic indicators, and sentiment analysis of textual data from different news. In order to increase the accuracy of the model predictions, the machine learning dataset in this experiment is split into training set: 80%, validation set: 10%, and test set: 10%. Each lag of each model combination was evaluated using 6 different metrics for the prediction results. This experiment contains 3 language models: FinBRET, TextBlob, VADER and 3 machine learning models: SVM, TCN, LSTM, a total of 12 prediction scenarios are investigated at 1-day as well as 2-day time lags. To ensure comprehensive sentiment analysis, in this paper, the results of using two to three language models were chosen as the training parameters for the machine learning models, and the prediction results of the combination of different language models and machine learning algorithms were compared and analyzed. The MSE, RMSE, MAE, R^2, p-value and f-statistic for each combination are listed in detail. To make comparisons easier, Bar charts are used to show the error values for each group.

4.3.1 Lag 1 Machine Learning Models Market volatility prediction results:

Model	MSE	RMSE	MAE	R^2
SVM + FinBERT & TextBlob	0.2218	0.4712	0.333	0.982
SVM + FinBERT & VADER	0.0693	0.2632	0.2002	0.995
SVM + TextBlob & VADER	0.08566	0.2926	0.2309	0.9933
SVM + FinBERT &TextBlob & VADER	0.2269	0.4764	0.3356	0.9819
LSTM+ FinBERT & TextBlob	1.0216	1.0107	0.7487	0.9256
LSTM+ FinBERT & VADER	0.5360	0.7321	0.5844	0.9561
LSTM+ TextBlob & VADER	1.0184	1.009	0.7346	0.8858
LSTM+ FinBERT &TextBlob & VADER	0.5812	0.7624	0.5910	0.9577
TCN+ FinBERT & TextBlob	1.1140	1.0555	0.7591	0.9189
TCN+ FinBERT & VADER	0.9522	0.9758	0.8088	0.9221
TCN+ TextBlob & VADER	1.8643	1.3654	1.0974	0.7910
TCN+ FinBERT &TextBlob & VADER	1.3029	1.1414	0.8107	0.9052

Table 2. Prediction Error Results of Lag 1 Predictions Using Machine Learning Models Combined with Language models

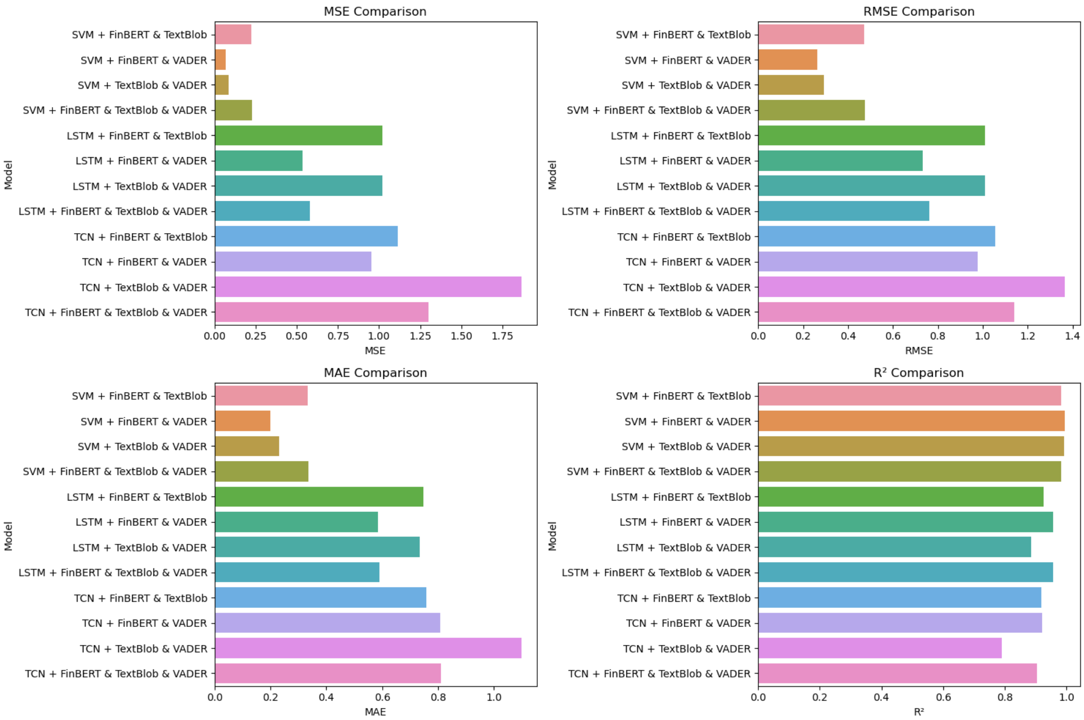

Figure 4. Lag 1 Bar chart of Prediction Error Results of Lag 1 Predictions Using Machine Learning Models Combined with Language models

Table 2 shows the detailed error results of the machine learning model predictions. The results show that in SVM combined with different sentiment analysis, SVM model trained based on FinBERT and VADER sentiment analysis performs best in predicting results, with the lowest MSE (0.069), RMSE (0.2632), and MAE (0.2002), and the highest R^2 (0.995). SVM with FinBERT and VADER performs well overall, with high accuracy in predicting market volatility and explaining a large portion of the variance in market volatility. Figure 5 displays the VIX prediction results of the SVM model trained on FinBERT and VADER sentiment analysis are highly similar to the actual VIX fluctuations.

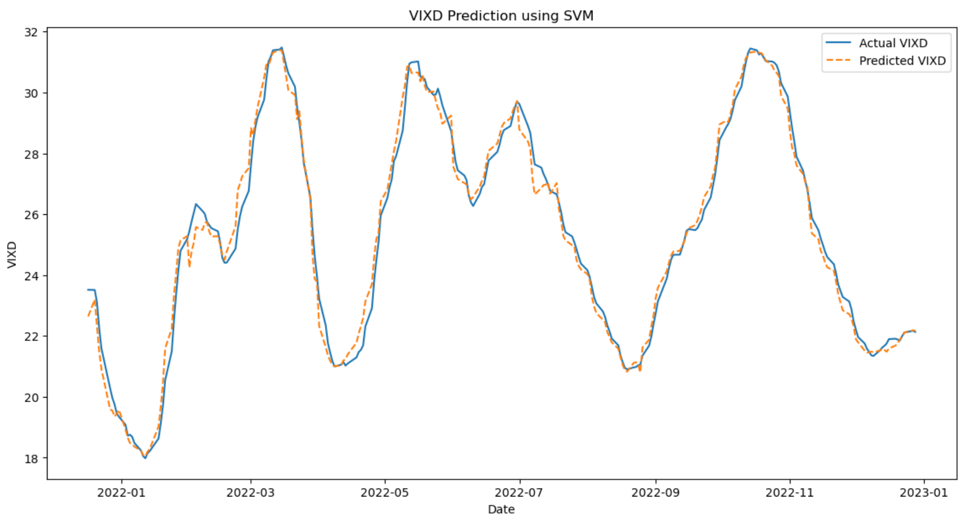

Figure 5. SVM Combined FinBERT and VADER's VIX Forecast Results vs. Actual VIX Line Chart

LSTM trained with FinBERT and VADER results outperforms with relatively low MSE (0.5360), RMSE (0.7321), and MAE (0.5844), and a high R^2 value (0.9561). But the overall performance of LSTM model is lower than SVM model. Figure 6 shows the VIX prediction results of LSTM with FinBERT and VADER. The prediction results show that LSTM cannot predict some fluctuations and shows a bit of smoothing when it is around the peaks and valleys.

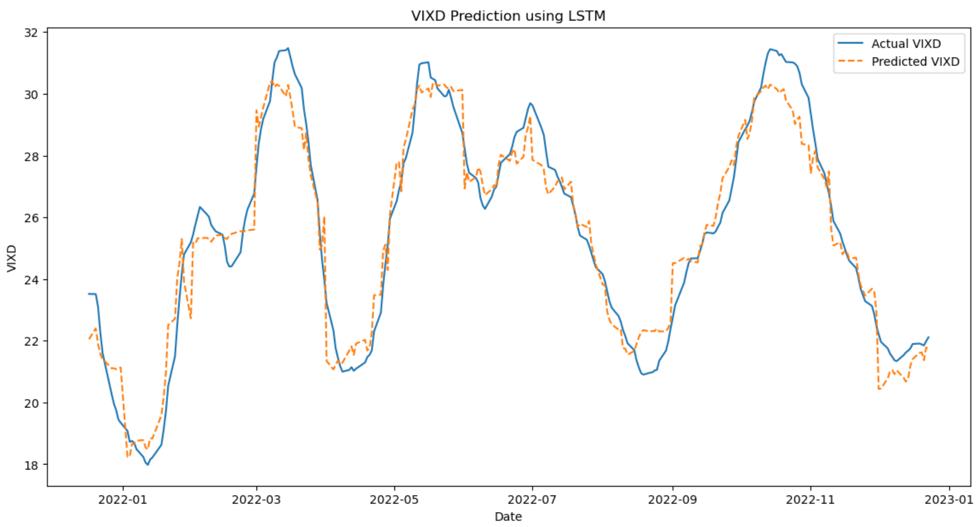

Figure 6. LSTM Combined FinBERT and VADER's VIX Forecast Results vs. Actual VIX Line Chart

The TCN model combinations had overall poorer predictive results than LSTM and SVM, especially the combination using TextBlob and VADER sentiment analysis results as parameters had the lowest R^2 of 0.7910. In the TCN model combinations, there are also model combinations that perform relatively well. TCN trained with FinBERT and TextBlob analysis results has lower MAE (0.7591). However, the overall performance of TCN with FinBERT and VADER is better, with relatively low MSE (0.9522) and RMSE (0.9758), and high R^2 (0.9221), but these two values are still high compared to SVM and LSTM results. Figure 7 shows the TCN model trained on FinBERT and VADER, and even though this combination performs the best among the TCN combinations, Figure 7 shows that this prediction differs from the actual VIX value.

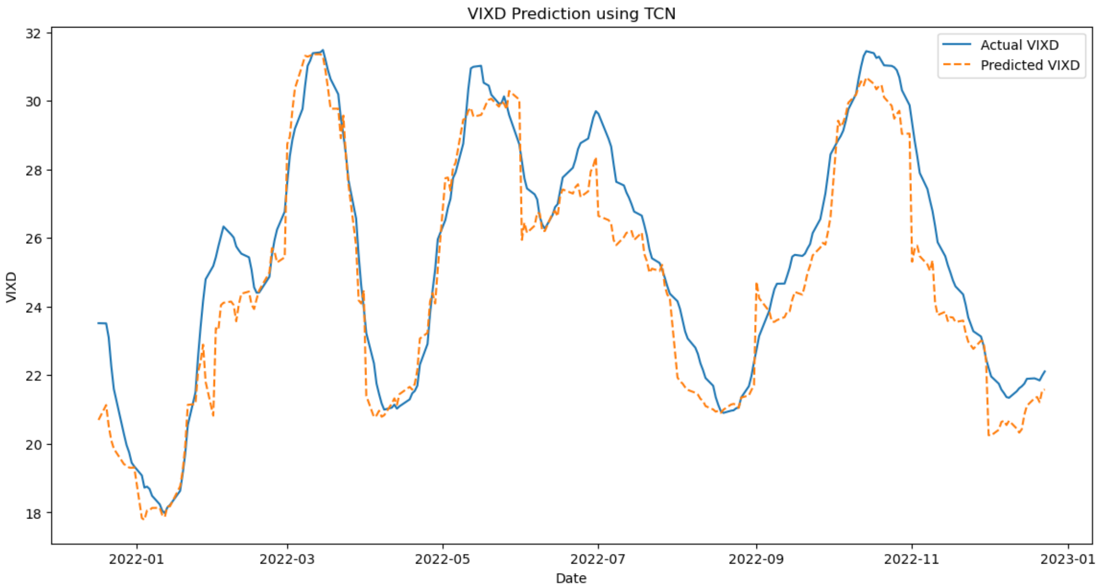

Figure 7. TCN Combined FinBERT and VADER's VIX Forecast Results vs. Actual VIX Line Chart

From the three best performing model combinations, it seems that FinBERT and VADER have the best combination of sentiment analysis and provide the best predictive data for the machine learning model. The reason for this may be that both FinBERT and VADER provide more detailed positive, negative, and neutral sentiment, while TextBlob can only provide the polarity of the data, not the detailed sentiment information. And too much sentiment analysis data may instead interfere with the predictive ability of machine learning models. As for the comparison of the predictive ability of the machine learning models, the combination of SVM and language models provides the best performance in most metrics, and LSTM does not perform as well as SVM in terms of MSE and

RMSE values, but it also demonstrates a strong predictive ability, utilizing its ability to capture temporal dependencies in the data. The TCN model, on the other hand, did not exceed the overall predictive power of SVM and LSTM and is not the best choice for making market volatility predictions. This comparative analysis shows that SVM combined with FinBERT and VADER in lag 1 are the most effective in predicting market volatility during the financial crisis and can predict market volatility largely correctly.

Model	P-value	F-statistic
SVM + FinBERT & TextBlob	3.7539e-43	268.5061
SVM + FinBERT & VADER	1.2108e-43	797.8608
SVM + TextBlob & VADER	1.0093e-42	695.6324
SVM + FinBERT &TextBlob & VADER	2.4841e-30	182.1355
LSTM+ FinBERT & TextBlob	2.3812e-19	48.5974
LSTM+ FinBERT & VADER	3.7853e-24	84.9817
LSTM+ TextBlob & VADER	9.5563e -17	36.3036
LSTM+ FinBERT &TextBlob & VADER	2.9565e-22	67.6837
TCN+ FinBERT & TextBlob	2.1160e-20	55.2050
TCN+ FinBERT & VADER	1.5094e-20	55.0775
TCN+ TextBlob & VADER	5.2866e-13	21.9778
TCN+ FinBERT &TextBlob & VADER	4.3191e-17	34.9055

Table 3. P-value and F-statistic Results of Lag 1 Predictions Using Machine Learning Models Combined with Language models

Table 3 shows the p-values and F-statistics of the different model combinations, which show the significance and goodness of fit of the models in predicting market volatility. P-values are used to assess the significance of the model. The results indicate that the p-values of all the models are very small and well below the standard significance level of 0.05, suggesting that all these models are significant in predicting market volatility. In particular, the p-value of SVM with FinBERT and VADER combination is 1.2108e-43, which indicates that this model has the highest significance in lagged period prediction, further proving the excellent performance of this model combination in predicting market volatility. The p-values of the LSTM models are generally larger than those of the SVM models but are also well below 0.05, especially the p-value of 3.7853e-24 for LSTM with FinBERT and VADER, which has a good significance. TCN combinations are still the weakest performer among the three models. TCN with FinBERT and TextBlob has the relatively low p-value of 2.116e -20, and TCN with FinBERT and VADER also performs well with p-value of 1.5094e-20. The F-statistic is used to evaluate the overall fit of the model. A larger F-statistic indicates a better fit, meaning the model can explain more of the market volatility. Among all the models, SVM with FinBERT and VADER combination has the highest F-statistic of 797.8608, shows that the model has the best fit. The F-statistics of other SVM combinations are also relatively high, further supporting the strong capability of the SVM model in predicting market volatility. Among the LSTM combinations, LSTM with FinBERT and VADER still performs the best with an f-statistic value of 84.9817. The TCN combinations perform slightly worse than LSTM. The TCN models trained on FinBERT and TextBlob results, as well as on FinBERT and VADER results, are both performing well, with both achieving scores around 55. Overall SVM models are still the best performers, with excellent significance and fit. While LSTM and TCN are not as good as SVM, they still provide good predictive power when combined with a suitable language model.

4.3.2 Lag 2 Machine Learning Models Market volatility prediction results:

Model	MSE	RMSE	MAE	R^2
SVM + FinBERT & TextBlob	0.1141	0.3377	0.2443	0.9866
SVM + FinBERT & VADER	0.1393	0.3732	0.2733	0.9843
SVM + TextBlob & VADER	0.2822	0.5312	0.3279	0.9729
SVM + FinBERT &TextBlob & VADER	0.1379	0.3713	0.2790	0.9899

LSTM+ FinBERT & TextBlob	1.2574	1.1213	0.8668	0.9026
LSTM+ FinBERT & VADER	0.4124	0.6420	0.5008	0.9685
LSTM+ TextBlob & VADER	0.6078	0.7796	0.6246	0.9535
LSTM+ FinBERT &TextBlob & VADER	0.5038	0.7098	0.5781	0.9610
TCN+ FinBERT & TextBlob	0.9353	0.9671	0.7496	0.9312
TCN+ FinBERT & VADER	1.1805	1.0865	0.9025	0.8973
TCN+ TextBlob & VADER	1.3824	1.1758	1.0137	0.8690
TCN+ FinBERT &TextBlob & VADER	1.2229	1.05967	0.8756	0.9130

Table 4. Prediction Error Results of Lag 2 Predictions Using Machine Learning Models Combined with Language models

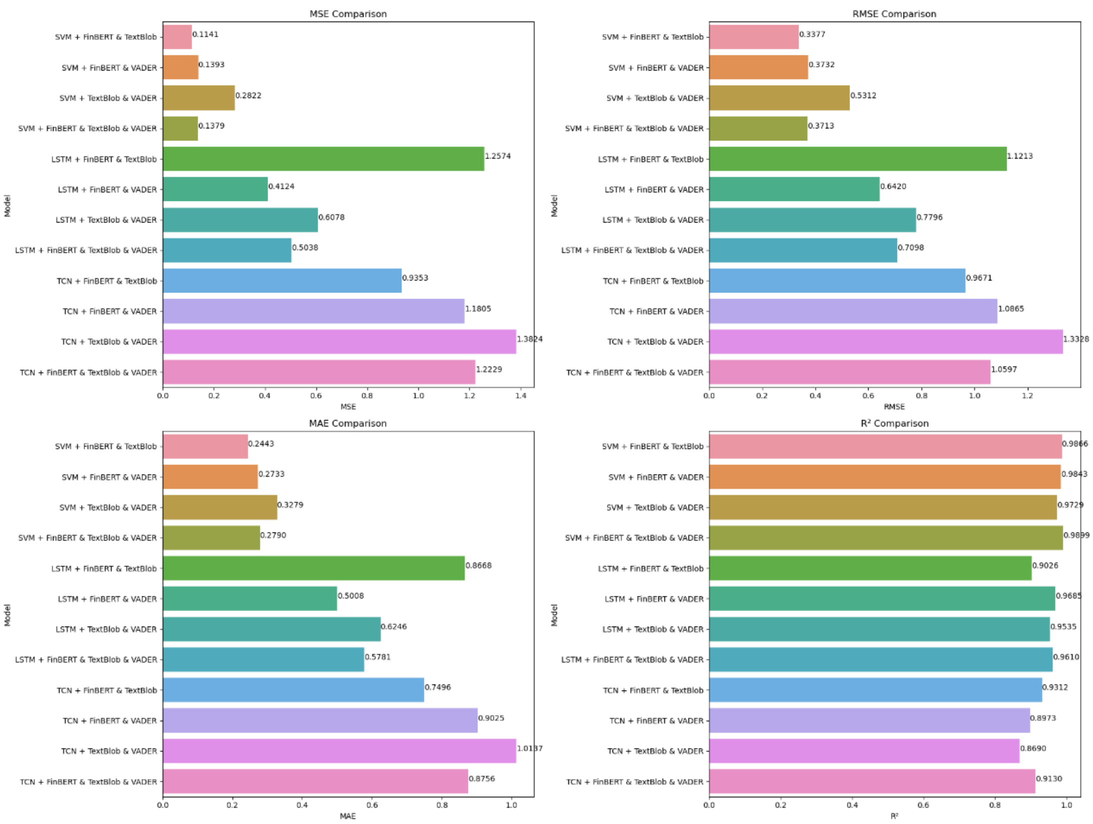

Figure 8. Bar chart of Prediction Error Results of Lag 2 Predictions Using Machine Learning Models Combined with Language models

In the Granger causality test, a significant number of indicators show causal significance at lag 2 periods. Therefore, the prediction results of machine learning model using sentiment analysis parameters with adjusted weights at a 2-period lag was testes in this experiment.

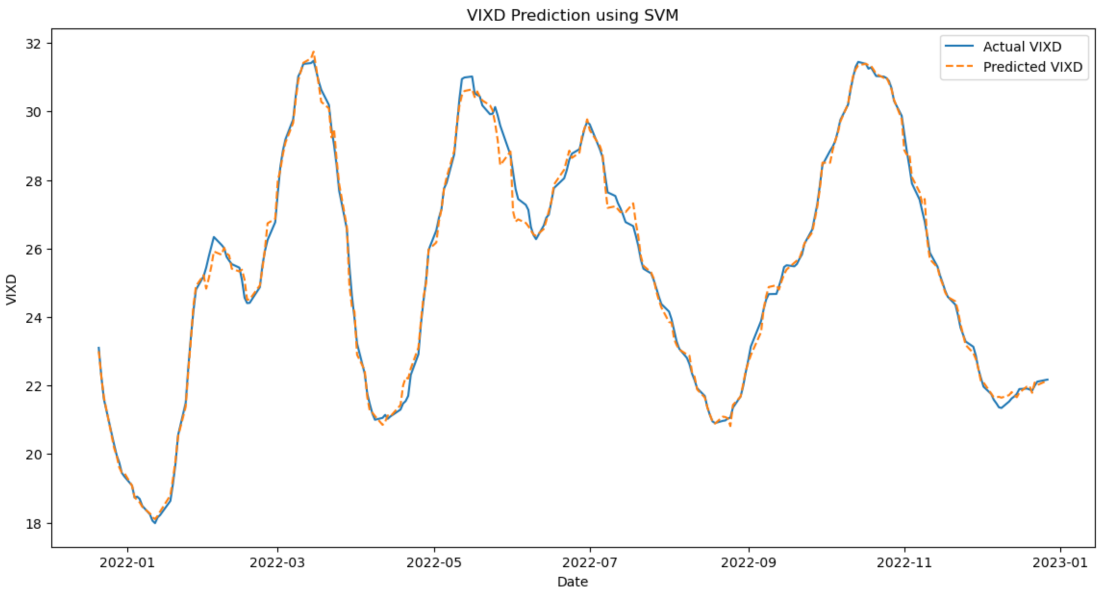

Figure 9. In lag 2 period SVM Combined FinBERT and TextBlob's VIX Forecast Results vs. Actual VIX Line Chart

The prediction results show that the SVM combinations has similar overall error values. Among the SVM combinations, the SVM with FinBERT and TextBlob combination has MSE of 0.1141, RMSE of 0.3377, and MAE of 0.2443, with an R^2 value of 0.9866, which are the smallest values compare to other SVM model combinations. Figure 9 exhibits the graph of SVM with FinBERT and VADER prediction results, and it can be observed that it still has an excellent performance in lag phase 2. This indicates that this combination has the highest accuracy and goodness of fit in lag 2 predictions.

LSTM that trained with FinBERT and VADER sentiment analysis results still has the best performance among the LSTM combinations, but still not as good as the SVM model. This combination has an MSE of 0.4124, an RMSE of 0.6420, an MAE of 0.5008, as well as an R^2 of 0.9685. Figure 10 exhibits a plot of LSTM with FinBERT and VADER prediction results, where LSTM can predict large market fluctuations, but still lacks in predicting peaks and valleys.

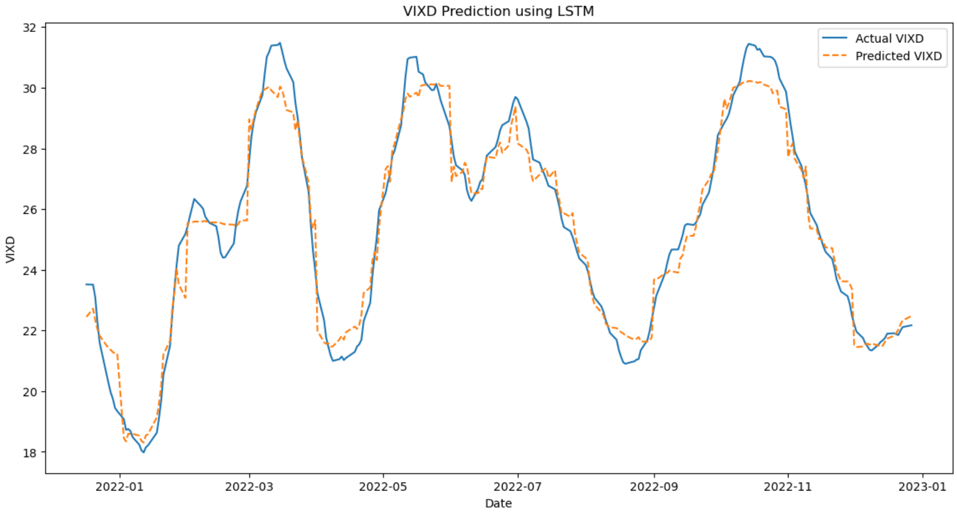

Figure 10. In lag 2 period LSTM Combined FinBERT and VADER's VIX Forecast Results vs. Actual VIX Line Chart

TCN with FinBERT and TextBlob performed best in the TCN model with MSE of 0.9353, RMSE of 0.9671, MAE of 0.7496, and R^2 of 0.9312. Figure 11 shows the prediction results of the TCN model trained on the results of FinBERT and TextBlob sentiment analysis, and it can be seen that the prediction results are basically consistent with the actual volatility trend, but there are still deficiencies in some of the peak prediction. In general, the TCN model is still not as good as SVM and LSTM.

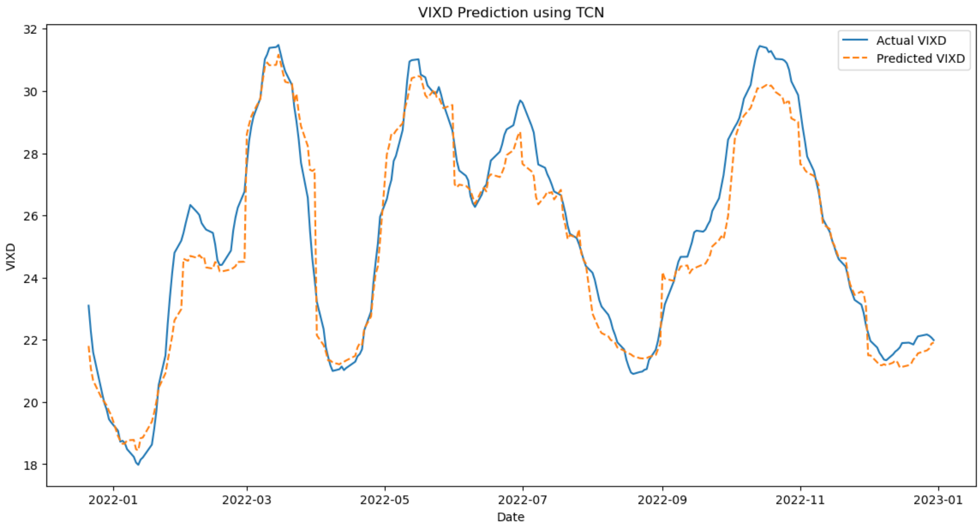

Figure 10. In lag 2 period TCN Combined FinBERT and TextBlob's VIX Forecast Results vs. Actual VIX Line Chart

Compared with the value of lag 1, lag 2 is still overall superior although the best performing model performs slightly lower than lag 1 in all metrics. Among the combinations of SVM with language models, only the combination of SVM with FinBERT, TextBlob, and VADER outperforms lag 2 on all error metrics with lag 1. Notably, MSE, RMSE, and MAE are significantly reduced. In the combination of LSTM and language models, LSTM with TextBlob and VADER combination and LSTM with FinBERT, TextBlob and VADER have significant improvement in lag 2, especially LSTM with TextBlob and VADER has an R^2 of 0.9535 in lag 2, which is significantly improved compared with the R^2 of 0.8858 in lag 1. Comparing the performance of TCN at lag 1 and lag 2, TCN with FinBERT and TextBlob has a significant improvement, and even the overall data is better than the best TCN model combination in lag 1: TCN with FinBERT and VADE. Overall, the combination that include TextBlob have a certain improve in their prediction results. However, the causality is stronger in lag 2, and lag 2 improves some of the model indexes, but the overall models' performance is not much better than the performance in lag 1. SVM with FinBERT and VADER is still the best model combination among all lag 1 and lag 2 combinations.

Model	P-value	F-statistic
SVM + FinBERT & TextBlob	2.6251e-36	341.5319
SVM + FinBERT & VADER	9.6322e-33	248.8562
SVM + TextBlob & VADER	9.6730e-30	173.9629
SVM + FinBERT &TextBlob & VADER	6.134e-35	316.7450
LSTM+ FinBERT & TextBlob	3.2637e-17	37.2974
LSTM+ FinBERT & VADER	1.1045e-26	114.5884
LSTM+ TextBlob & VADER	1.2388e-24	91.0341
LSTM+ FinBERT &TextBlob & VADER	1.8164e-22	69.4993
TCN+ FinBERT & TextBlob	8.5662e-21	57.8751
TCN+ FinBERT & VADER	2.4396e-17	36.9156
TCN+ TextBlob & VADER	3.55511e-17	37.1257
TCN+ FinBERT &TextBlob & VADER	5.7303e-17	34.3436

Table 5. P-value and F-statistic Results of Lag 2 Predictions Using Machine Learning Models Combined with Language models

While observing the p-value and f-statistic evaluation table 5, the overall model still performs well in both statistical significance test metrics. The p-value of SVM is still the smallest, and both SVM with FinBERT and TextBlob combination, and SVM with FinBERT and VADER combination have very low p-value of 2.6251e-36 and 9.6322e-33, respectively. P-values of LSTM model are also small, but slightly larger compared to SVM model. For example, LSTM with FinBERT and VADER has a p-value of 1.1045e-26, which is still significant but slightly not good as the SVM model. The TCN combinations have relatively large p-value among all the models but are still less than 0.05. TCN with FinBERT and TextBlob has the lowest p-value among the TCN combinations which is 8.5662e-21, indicating that the model are statistically significant. The F-statistic results show that the SVM model has higher F-statistics in lag 2, especially the F-statistic of SVM with FinBERT and TextBlob is 341.5319, which shows an excellent fit and indicates that the model is able to explain the market volatility very well. The F-statistics of the LSTM model are relatively low, for example, LSTM with FinBERT and TextBlob has an F-statistic of 37.2974, shows that it is not as good as the SVM model in explaining market volatility. The TCN model has the lowest F-statistic among all the models, in particular, the F-statistic of TCN with TextBlob and VADER is only 34.3436, which indicates that the model is poorly fitted in lag 2 periods and has a low explanatory power. Overall, the SVM models still perform optimally in lag 2, and they perform well in terms of significance and overall fit, especially when combined with FinBERT and TextBlob. The LSTM and TCN models, on the other hand, do not perform as well as the SVM, but both have lower overall significance and some good fit.

FinBERT and TextBlob combination in lag 2 period perform better than FinBERT and VADER in SVM and TCN model error tests and statistical significance test metrics. The main possible reason is firstly that compared to lag 1, the sentiment analysis values of TextBlob, subjectivity and polarity, have significance in lag 2, and therefore in the lag 2 TextBlob can provide more valid sentiment features to help the model to make predictions. However, in lag 1, due to the low correlation of TextBlob, it may instead over-capture some details and noise in the data, resulting in higher variance and error. In addition, although combining FinBERT with VADER or TextBlob as training works well, using the results of FinBERT, VADER, and TextBlob at the same time will create a situation that excessive features may have a negative impact on the prediction.

Based on the lag 1 and lag 2 errors as well as the significance test results, the SVM model performs much better than the LSTM and TCN models in predicting the volatility of the U.S. market during the Covid-19 period. Interestingly even though the computational complexity of the SVM model is lower than the LSTM and TCN models, it performs very well. The possible reason is that the main duration of the epidemic is 2020-2022, the overall market data size is small compared to other long time forecasts, especially the financial stock market data during the epidemic may be relatively short or drastically fluctuating. SVM is particularly effective at handling small datasets with well-defined features. However, the limited training data or short time series may prevent LSTMs and TCNs from effectively capturing the long-term market dependencies that they are designed to specialize in, leading to poorer forecasting performance compared to SVM models. On the other hand, unexpected events such as financial crisis can affect the market patterns and behaviors of regular time series, which may influence the performance of LSTM and TCN models. In contrast, rapid market fluctuations generate a large number of features, giving SVM models that rely on feature learning a competitive advantage. This experiment results show that the predictive performance of SVM can be significantly enhanced through effective feature engineering, such as selecting appropriate market indicators and incorporating sentiment analysis data. Therefore, in short-term and special market environments, SVM performs more robustly and effectively in dealing with rapid market volatility and uncertainty.

5. Conclusion

Market volatility is important to financial researchers and investors, and effective prediction of market volatility can not only help to minimize risk, but also lead to early preventive measures. However, market volatility during financial crises is more difficult to manage and is often accompanied by great uncertainty and high volatility. Therefore, in order to better forecast market volatility during financial crises, this paper empirically investigates the market volatility in the United States during the COVID-19 period by selecting three advanced language models, FinBERT, TextBlob, and VADER, and combining them with three machine learning models, SVM, LSTM and TCN which are widely used in financial forecasting. In this experiment, U.S. financial news during Covid-19 was selected and sentiment was analyzed through 3 language models. Next, different machine learning models were combined with analysis results from various language models for training. A total of 12 forecasting combinations are constructed, and their performance in predicting market volatility at lag 1 and lag 2 is compared.

Overall, the results show that the p-values of the model predictions are less than 0.05 and the F-statistics are relatively high, indicating that these predictions have significant validity in predicting the volatility of the financial markets during financial crises. However, for the error metrics, there is a large difference between the results of different model combinations. Among all the model combinations, the SVM model using FinBERT and VADER sentiment analysis results as training parameters in lag 1 period performs the best and has the most accurate predictions. This combination has the lowest mean square error (MSE, 0.069), root mean square error (RMSE, 0.263), and mean absolute error (MAE, 0.200), as well as the highest coefficient of determination (R^2, 0.995), showing optimal prediction accuracy. The p-value of this combination is less than 0.05, indicating a significant significance, while the F-statistic is close to 800, showing excellent model fitting. The prediction results of the SVM model are overall better than the LSTM and TCN models. This demonstrates that during periods of financial crisis, when the market experiences significant volatility and the overall duration is relatively short, the predictive capability of the SVM model outperforms that of the LSTM and TCN models. Meanwhile, ML models trained using the results of FinBERT and VADER sentiment analysis have higher prediction accuracy than models trained based on a combination of other language models in lag 1. In lag 2, the accuracy of the models based on FinBERT and TextBlob sentiment analysis results as training parameters has been improved, especially the MSE, RMSE error values of LSTM and TCN models have been reduced, and R^2 values have been improved. This situation may be due to the TextBlob model sentiment analysis results have a significant Granger causality significance at lag 2. However, although the Granger causality test shows more significant causality for the sentiment parameters in lag 2, the SVM model that includes TextBlob as a training parameter does not have significant enhancements. Also, the overall performance in lag 2 is not significantly better than that in lag 1, some model combinations even regress on some metrics.

This paper also has some limitations. First, this study was not able to further validate the sentiment analysis accuracy of the language models. In order to accurately assess the sentiment analysis results of the language model, future research needs to manually label the collected financial news with sentiment as real sentiment labels. By comparing the manually labeled sentiment results with the sentiment analysis results of the language model, the accuracy of the model can be more effectively assessed. This limitation suggests that the current findings may be biased in terms of sentiment analysis accuracy, which should be improved in subsequent studies. Secondly, the study focuses on the U.S. market during the COVID-19 period, and it can be extended to other financial crises or markets in different countries in the future to verify the broad applicability of the model. In addition, although this paper combines a variety of language models and machine learning models, future research can further explore more model combinations as well as improve feature engineering methods to enhance prediction performance. In conclusion, this study demonstrates that the SVM model, trained using sentiment analysis results from FinBERT and VADER in lag 1, shows significantly accuracy in predicting market volatility during the Covid-19 financial crisis. This finding provides an important reference and guidance for future model selection for market volatility forecasting during financial crises, helping future researchers and investors to more accurately predict market volatility during financial crises.

6. Reference

Ahuja, S. and Dubey, G. (2017). Clustering and sentiment analysis on Twitter data. *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*. doi:https://doi.org/10.1109/tel-net.2017.8343568.

Albulescu, C.T. (2020). COVID-19 and the United States financial markets' volatility. *Finance Research Letters*, 38, p.101699. doi:https://doi.org/10.1016/j.frl.2020.101699.

Aljedaani, W., Rustam, F., Mkaouer, M.W., Ghallab, A., Rupapara, V., Washington, P.B., Lee, E. and Ashraf, I. (2022). Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowledge-Based Systems*, 255, p.109780. doi:https://doi.org/10.1016/j.knosys.2022.109780.

Amrita Ticku, Tripathy, N., Pankaj Kumar Mishra, Sinha, A., Shivangi Jadon, Raj, A., Harsh Kulshrestha, Rai, V. and Shamim, R. (2023). Vader protocol based sentiment analysis technique using LSTM for Stock Trend prediction. doi:https://doi.org/10.1109/icccnt56998.2023.10307985.

Antweiler, W. and Frank, M.Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 59(3), pp.1259–1294. doi:https://doi.org/10.1111/j.1540-6261.2004.00662.x.

Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.1908.10063.

Asderis, G.-A. (2022). Sentiment Analysis on Twitter Data, a Detailed Comparison of TextBlob and VADER. *Ihu.edu.gr*. [online] doi:https://repository.ihu.edu.gr//xmlui/handle/11544/29947.

Bai, S., Kolter, J.Z. and Koltun, V. (2018). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.1803.01271.

Bao, W., Yue, J. and Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE*, 12(7), p.e0180944. doi:https://doi.org/10.1371/journal.pone.0180944.

Cao, J., Li, Z. and Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications*, 519, pp.127–139. doi:https://doi.org/10.1016/j.physa.2018.11.061.

ÇILGIN, C., BAŞ, M., BİLGEHAN, H. and ÜNAL, C. (2022). Twitter Sentiment Analysis During Covid-19 Outbreak with VADER. *AJIT-e: Academic Journal of Information Technology*, 13(49), pp.72–89. doi:https://doi.org/10.5824/ajite.2022.02.001.x.

Dai, W., An, Y. and Long, W. (2022). Price change prediction of Ultra high frequency financial data based on temporal convolutional network. *Procedia Computer Science*, 199, pp.1177–1183. doi:https://doi.org/10.1016/j.procs.2022.01.149.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. [online] arXiv.org. Available at: https://arxiv.org/abs/1810.04805.

Fan, S., Chen, L., Li, H., Lin, Z., Su, J., Zhang, H., Gong, Y., Guo, J. and Duan, N. (2022). Sentiment-Aware Word and Sentence Level Pre-training for Sentiment Analysis. *arXiv (Cornell University)*. doi:https://doi.org/10.18653/v1/2022.emnlp-main.332.

Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), pp.654–669. doi:https://doi.org/10.1016/j.ejor.2017.11.054.

Frankel, R., Jennings, J. and Lee, J. (2021). Disclosure Sentiment: Machine Learning vs. Dictionary Methods. *Management Science*. doi:https://doi.org/10.1287/mnsc.2021.4156.

Fu, Y. and Xiao, H. (2022). Stock Price Prediction Model Based on Dual Attention and Tcn. *SSRN Electronic Journal*. doi:https://doi.org/10.2139/ssrn.4282842.

Gopali, S., Abri, F., Siami-Namini, S. and Siami, N.A. (2021). A Comparative Study of Detecting Anomalies in Time Series Data Using LSTM and TCN Models. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.2112.09293.

Guo, W., Li, Z., Gao, C. and Yang, Y. (2022). Stock price forecasting based on improved time convolution network. *Computational Intelligence*. doi:https://doi.org/10.1111/coin.12519.

Huang, A.H., Wang, H. and Yang, Y. (2022). FinBERT: A Large Language Model for Extracting Information from Financial Text†. *Contemporary Accounting Research*. doi:https://doi.org/10.1111/1911-3846.12832.

Huang, W., Nakamori, Y. and Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10), pp.2513–2522. doi:https://doi.org/10.1016/j.cor.2004.03.016.

Illia, F., Eugenia, M.P. and Rutba, S.A. (2022). Sentiment Analysis on PeduliLindungi Application Using TextBlob and VADER Library. *Proceedings* *of The International Conference on Data Science and Official Statistics*, 2021(1), pp.278–288. doi:https://doi.org/10.34123/icdsos.v2021i1.236.

Janiesch, C., Zschech, P. and Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, [online] 31(31), pp.685–695. doi:https://doi.org/10.1007/s12525-021-00475-2.

Karlsson, N., Loewenstein, G. and Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, [online] 38(2), pp.95–115. doi:https://doi.org/10.1007/s11166-009-9060-6.

Kearney, C. and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, pp.171–185. doi:https://doi.org/10.1016/j.irfa.2014.02.006.

Khoa, B.T. and Huynh, T.T. (2022). Forecasting stock price movement direction by machine learning algorithm. *International Journal of Electrical and Computer Engineering (IJECE)*, 12(6), p.6625. doi:https://doi.org/10.11591/ijece.v12i6.pp6625-6634.

Kim, J., Kim, H.-S. and Choi, S.-Y. (2023). Forecasting the S&P 500 Index Using Mathematical-Based Sentiment Analysis and Deep Learning Models: A FinBERT Transformer Model and LSTM. *Axioms*, [online] 12(9), p.835. doi:https://doi.org/10.3390/axioms12090835.

Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2), pp.307–319. doi:https://doi.org/10.1016/s0925-2312(03)00372-2.

Ko, C.-R. and Chang, H.-T. (2021). LSTM-based sentiment analysis for stock price forecast. *PeerJ Computer Science*, 7, p.e408. doi:https://doi.org/10.7717/peerj-cs.408.

Kotelnikova, A., Paschenko, D., Bochenina, K. and Kotelnikov, E. (2022). Lexicon-Based Methods vs. BERT for Text Sentiment Analysis. *Lecture notes in computer science*, pp.71–83. doi:https://doi.org/10.1007/978-3-031-16500-9_7.

Kurani, A., Doshi, P., Vakharia, A. and Shah, M. (2021). A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting. *Annals of Data Science*, 10. doi:https://doi.org/10.1007/s40745-021-00344-x.

Lakshminarayanan, S.K. and McCrae, J.P. (2019). A Comparative Study of SVM and LSTM Deep Learning Algorithms for Stock Market Prediction. *AICS*, pp.446–457.

Lei, B., Zhang, B. and Song, Y. (2021). Volatility Forecasting for High-Frequency Financial Data Based on Web Search Index and Deep Learning Model. *Mathematics*, 9(4), p.320. doi:https://doi.org/10.3390/math9040320.

Li, X., Xie, H., Chen, L., Wang, J. and Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69, pp.14–23. doi:https://doi.org/10.1016/j.knosys.2014.04.022.

Liang, C., Tang, L., Li, Y. and Wei, Y. (2020). Which sentiment index is more informative to forecast stock market volatility? Evidence from China. *International Review of Financial Analysis*, 71, p.101552. doi:https://doi.org/10.1016/j.irfa.2020.101552.

Liapis, C.M., Karanikola, A. and Kotsiantis, S. (2023). Investigating Deep Stock Market Forecasting with Sentiment Analysis. *Entropy*, 25(2), p.219. doi:https://doi.org/10.3390/e25020219.

Liu, J.-X., Leu, J.-S. and Holst, S. (2023). Stock price movement prediction based on Stocktwits investor sentiment using FinBERT and ensemble SVM. 9, pp.e1403–e1403. doi:https://doi.org/10.7717/peerj-cs.1403.

Liu, Z., Huang, D., Huang, K., Li, Z. and Zhao, J. (2021). *FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining*. [online] Available at: https://www.ijcai.org/Proceedings/2020/0622.pdf.

LOUGHRAN, T. and MCDONALD, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), pp.35–65.

Loughran, T. and McDonald, B. (2014). The Use of Word Lists in Textual Analysis. *SSRN Electronic Journal*. doi:https://doi.org/10.2139/ssrn.2467519.

Maghyereh, A. and Abdoh, H. (2022). Global financial crisis versus COVID-19: Evidence from sentiment analysis. *International Finance*. doi:https://doi.org/10.1111/infi.12412.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L.T. and Trajanov, D. (2020). Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access*, [online] 8, pp.131662–131682. doi:https://doi.org/10.1109/ACCESS.2020.3009626.

Mittal, A. and Goel, A. (2011). *Stock Prediction Using Twitter Sentiment Analysis*. [online] Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4ecc55e1c3ff1cee41f21e5b0a3b22c58d04c9d6.

Pano, T. and Kashef, R. (2020). A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. *Big Data and Cognitive Computing*, 4(4), p.33. doi:https://doi.org/10.3390/bdcc4040033.

Prameswari Ekaputri, A. and Akbar, S. (2022). Financial News Sentiment Analysis using Modified VADER for Stock Price Prediction. doi:https://doi.org/10.1109/icaicta56449.2022.9932925.

Ranjan, S. (2019). Investor community sentiment analysis for predicting stock price trends. *International Journal of Management, Technology And Engineering*. [online] Available at: https://www.academia.edu/39810746/Investor_community_sentiment_analysis_for_predicting_stock_price_trends [Accessed 16 Jun. 2024].

Ren, R., Wu, D.D. and Liu, T. (2019). Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine. *IEEE Systems Journal*, [online] 13(1), pp.760–770. doi:https://doi.org/10.1109/JSYST.2018.2794462.

Sabherwal, S., Sarkar, S.K. and Zhang, Y. (2011). Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News. *Journal of Business Finance & Accounting*, 38(9-10), pp.1209–1237. doi:https://doi.org/10.1111/j.1468-5957.2011.02258.x.

Seo, S.W. and Kim, J.S. (2015). The information content of option-implied information for volatility forecasting with investor sentiment. *Journal of Banking & Finance*, [online] 50, pp.106–120. doi:https://doi.org/10.1016/j.jbankfin.2014.09.010.

Shapiro, A.H., Sudhof, M. and Wilson, D. (2017). Measuring News Sentiment. *Federal Reserve Bank of San Francisco, Working Paper Series*, pp.01-A2. doi:https://doi.org/10.24148/erwp2017-01.

Shi, Y. and Ho, K.-Y. (2021). News sentiment and states of stock return volatility: Evidence from long memory and discrete choice models. *Finance Research Letters*, 38, p.101446. doi:https://doi.org/10.1016/j.frl.2020.101446.

Siami-Namini, S. and Akbar Siami Namin (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. doi:https://doi.org/10.48550/arxiv.1803.06386.

Sohangir, S., Petty, N. and Wang, D. (2018a). *Financial Sentiment Lexicon Analysis*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICSC.2018.00052.

Sohangir, S., Petty, N. and Wang, D. (2018b). *Financial Sentiment Lexicon Analysis*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICSC.2018.00052.

Sohangir, S., Wang, D., Pomeranets, A. and Khoshgoftaar, T.M. (2018). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data*, 5(1). doi:https://doi.org/10.1186/s40537-017-0111-6.

Tay, F.E.H. and Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29(4), pp.309–317. doi:https://doi.org/10.1016/s0305-0483(01)00026-3.

Tetlock, P. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3), pp.1139–1168. doi:https://doi.org/10.1111/j.1540-6261.2007.01232.x.

Todd, A., Bowden, J. and Yashar Moshfeghi (2024). Text-based sentiment analysis in finance: Synthesising the existing literature and exploring future directions. *International journal of intelligent systems in accounting, finance & management*, 31(1). doi:https://doi.org/10.1002/isaf.1549.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017). *Attention is All you Need*. [online] Neural Information Processing Systems. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Wang, J., Hong, S., Dong, Y., Li, Z. and Hu, J. (2024). Predicting Stock Market Trends Using LSTM Networks: Overcoming RNN Limitations for Improved Financial Forecasting. *Journal of Computer Science and Software Applications*, [online] 4(3), pp.1–7. doi:https://doi.org/10.5281/zenodo.12200708.

Xu, Y., Zhou, Y., Sekula, P. and Ding, L. (2021). Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment*, [online] 6, p.100045. doi:https://doi.org/10.1016/j.dibe.2021.100045.

Zhang, J. and Wang, W. (2024). A Stock Trend Prediction Model Based on Wavelet Transform and TCN Combined with Market Sentiment. doi:https://doi.org/10.4108/eai.23-2-2024.2345895.