**Abstract**

In recent years, stock price forecasting has been recognized as one of the most challenging activities in finance. The price of financial assets is non-linear, dynamic and irregular. Therefore, it is very difficult to form financial time series and predict them. The share price is affected by many factors such as the company's fundamental parameters, company events, political and social situations and human psychology, which complicates the analysis of financial markets. In the meantime, deep learning models are one of the latest available techniques for predicting stock prices, which have a high ability to recognize complex patterns in different fields. The purpose of this paper is to present a hybrid approach by considering different deep learning techniques to predict the stock market by examining the impact of news. In this regard, news data related to the shares of Tesla Motors Company have been collected from reliable news websites for 11 months since the beginning of 2020, and using the FinBERT model (an extension of the BERT model), the effect of this news on the movement of the stock price of this company has been investigated. Furthermore, using this model, a mechanism for predicting the price of Tesla shares has been presented. The results of this research showed that the accuracy of this model, which is based on deep learning concepts, is higher than the traditional models.

**Keywords: stock price prediction, deep learning, natural language processing, text mining, FinBERT model**

**1- Introduction**

The prediction of stock exchange is considered one of the most important and challenging aspects related to time series (Chen and Hao, 2017). Despite the introduction of the efficient market hypothesis by Malkiel and Fama (1970), later revisited by Fama (1991), suggesting that financial markets follow random paths and are thus unpredictable, the quest for profitable models and systems continues to captivate the attention of many experts in the research field (Weng et al., 2017). Furthermore, the literature reveals evidence of financial market inefficiency (e.g., Fama,1991; Malkiel, 2003; Atsalakis and Valavanis, 2009).

In order to achieve an accurate prediction in the stock market, the identification of influential features is of utmost importance. In other words, the precise selection of features plays a fundamental role in the efficiency of prediction (Barak et al., 2017). Among the classical techniques of financial market prediction, special attention is given to the following: technical analysis, which utilizes support and resistance standards and indicators calculated from past prices to indicate upward or downward trends (Chen et al., 2014; Lahmiri, 2014), and fundamental analysis, which focuses on economic factors that impact market trends (Cavalcante et al., 2016). Fundamental analysis can be employed to evaluate the performance and financial condition of a company over a specific time period by conducting a detailed analysis of its financial statements (Huang, 2012). Conversely, technical analysis assesses securities by using statistical measures such as past price and volume of transactions, which are produced by market activities (Barak et al., 2015).

Furthermore, the examination of stock prices and financial indices can be conducted through the utilization of time series analysis tools. Primary techniques for prediction encompass moving averages, discriminant analysis (DA), and regression analysis (Wang et al., 2012; Kumar and Thenmozhi, 2014). On the other hand, one of the encouraging domains in the realm of time series prediction research is artificial intelligence, which has recently been the subject of discussion (Wang et al., 2012; Yan et al., 2017).

Technological advancements have facilitated the analysis of historical price databases through computational systems (Chiang et al., 2016). One notable application of data mining, which has gained significant attention in recent years, is the utilization of text mining tools to examine texts and extract hidden knowledge embedded within them. Text mining, or knowledge discovery from text (Karanikas and Theodoulidis, 2002), refers to the process of uncovering non-trivial, interesting, and high-quality patterns, as well as information and knowledge, from unstructured textual documents. Unlike data mining, which focuses on extracting knowledge from structured databases, text mining primarily involves searching through textual data to extract valuable insights, typically from unstructured sources (Alwidian et al., 2015).

According to studies conducted, machine learning algorithms have shown remarkable performance in this field, highlighting their significance. As a result, most research in the area of stock price prediction nowadays focuses on intelligent methods. Consequently, further exploration in this domain is deemed necessary. Machine learning methods have gained recognition as beneficial approaches for predicting stock prices due to their high capability in modeling complex problems and nonlinear systems. Examinations have indicated that in efficient capital markets, analyzing important variables can yield good results in understanding price trends and prediction company stock prices. Additionally, markets that exhibit high efficiency demonstrate a good response to published news surrounding stocks. In this regard, the aim of this research is to present a stock price prediction model by examining textual news data using various machine learning approaches.

The research is structured as follows. The second section examines the relevant literature related to the research topic. The third section introduces the research methodology and the collected data. Subsequently, the fourth section analyzes the obtained results and evaluates the performance of the proposed model. Finally, the last section provides a comprehensive conclusion and suggests areas for future research.

**2- Literature Review**

The stock exchange is a complex and volatile system. The prediction of stock exchange trend characterized by a data intensity, noise, non-stationary, high degree of uncertainty, and hidden relationships (Yu et al., 2008). Consequently, it is of utmost importance to develop an appropriate model that can effectively address these challenges and provide accurate predictions of stock market price trend. Extensive research has been conducted in this field, with one of the pioneering studies being conducted by Malkiel and Fama (1970), who introduced the efficient market hypothesis. This hypothesis has been examined in three forms: weak, semi-strong, and strong. Fama later revisited and evaluated this theory (Fama, 1991). In other researches, Engle (1982) and Bollerslev (1986) introduced significant econometric models commonly used in financial market prediction, namely autoregressive conditional heteroskedasticity (ARCH) and Generalized ARCH. Furthermore, financial markets are characterized by non-linear dynamics and their interaction with political and economic news, as well as the expectations held by market participants. Hence, there is a need to employ alternative methodologies, such as the approach proposed by Elman (Elman, 1990), which introduces a memory-based prediction network that surpasses some conventional neural network models. On the other hand, Campbell (1987) aimed to categorize all variables that predict stock returns into two distinct categories. Nevertheless, he concluded that no simple model could accurately predict all fluctuations in stock price returns.

In another study conducted by Kim (2003), the application of Support Vector machine (SVM) was explored for the classification of daily movements in the Korean stock market index (KOSPI). Kim employed technical analysis indicators as predictive variables in this research. Furthermore, the results obtained from Kim's approach were compared to neural network methods and case-based reasoning (CBR), leading to the conclusion that Kim's proposed method exhibited superior performance in terms of accuracy. Similarly, Huang et al. (2005) utilized SVM to predict stock movements in their study. They conducted a comparison between this method and other approaches such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and elman backpropagation neural network (EBNN). They established that the SVM method, either on its own or in conjunction with one of the mentioned approaches, exhibited higher accuracy. In another related study, Pai and Lin (2005) employed SVM as a predictive method. However, they utilized this method not solely for predicting stock prices, but for other purposes as well.

The SVM classification model is capable of being adapted to regression tasks for predicting values in financial time series, known as Support Vector Regression (SVR). Huang and Tsai (2009) demonstrated an example of utilizing this model in their work. In another study conducted by Yu et al. (2008), they employed a variant of SVM called Least Squares Support Vector Machine (LSSVM), which offered lower computational cost and higher generalization capability compared to the original model. They further enhanced the performance by combining LSSVM with a genetic algorithm to optimize the values for each generation. It is evident that the SVM method, either in isolation or in conjunction with other models, has found extensive application in numerous researches.

On the other hand, artificial neural networks (ANNs) have been repeatedly mentioned in the literature and extensively utilized by researchers. In a study by Kara et al. (2011), they compared these two widely used models, namely artificial neural networks and support vector machines, based on their predictive abilities for stock market movements in the Turkish market. The researchers incorporated historical stock price data from the past ten years into their modeling process, aiming to enhance the model's generalizability as much as possible. Yoon et al. (1993) conducted influential research that effectively addressed stock price prediction. In this study, the authors utilized neural networks to predict stock performance in financial markets. They demonstrated that employing neural networks yielded superior outcomes compared to traditional analytical methods. In another research conducted by Abu-Mostafa and Atiya (1996), they elucidated the fundamental approaches to stock price prediction and elucidated various techniques for selecting inputs, outputs, and error performance by leveraging neural networks as a learning framework. In addition, it is worth mentioning the research by Zhang et al. (1998), which serves as a comprehensive review on the application of neural networks in predicting financial markets. The authors believe that neural networks, given their adaptability and compatibility in handling nonlinear time series, can be employed for predicting financial markets. The literature on financial market prediction encompasses a wide range of studies utilizing neural networks (e.g., Fernandez-Rodrıguez et al., 2000; Leung et al., 2000; Chen et al., 2003).

Some recent works have also addressed the utilization of hybrid and innovative models. Krauss et al. (2017) conducted research on combining classification models for predicting financial markets. The results of their study demonstrate improved performance in the hybrid model. In another work, Naik and Mohan (2019) extracted 33 technical indicators based on daily stock prices in the Indian stock exchange, such as opening, high, low, and closing prices. They proposed machine learning techniques and deep learning models for predicting stock price movements. Based on their findings, the employed deep learning model exhibited superior performance compared to machine learning techniques.

Continuing on, research has been conducted to explore the prediction of stock markets by considering news using both traditional machine learning models and novel deep learning models. In the domain of financial market prediction, particularly in relation to the influence of news, there have been limited studies that have introduced innovative deep learning concepts. The majority of studies have predominantly relied on traditional machine learning models. One noteworthy study in this regard is the distinct research conducted by Schumaker and Chen (2009). They employed a support vector machine along with text analysis to investigate the impact of news on stock prices. They utilized a machine learning approach to predict and analyze financial news, employing various text representations such as bag of words, noun phrases, and named entities. In another research study, Timmons and Lee (2007) introduced a text classification system to categorize news containing references to company lists, with the aim of predicting the influence of news on stock prices. Another investigation conducted by Xu (2014) focused on stock price prediction using Google Trends and Yahoo Finance data, combining time series analysis techniques with information gathered from these websites to predict weekly changes in stock prices. In their research, Nguyen et al. (2015) focused on assessing the effectiveness of sentiment analysis in predicting stock performance through a large-scale experiment. Comparing it to a model that solely relied on historical prices, their proposed approach exhibited a 2.07% better performance in predicting the movement of 18 stocks over a year.

The term "deep learning" was first introduced to the field of machine learning by Rina Dechter in 1986. In recent years, there has been a significant interest in the innovative concepts of deep learning, particularly the Transformer models. One of the most recent Transformer models is the Bidirectional Encoder Representations from Transformers (BERT) model, which was first introduced by Devlin et al. (2018). Among the studies that have utilized this model, the research by Jang et al. (2020) stands out. In their research, they proposed an effective combination of macroeconomic indicators and news using a deep learning model to enhance the prediction of the Dow Jones index (DJI). They employed the NLTK VADER algorithm and a BERT model for the analysis of news. They categorized the news into positive, negative, and neutral classes. In another study, Sousa et al. (2019) tackled the issue of sudden stock price changes upon the release of important news, which may take minutes for human analysis while investors in financial markets need to make prompt decisions. To address this, they proposed a BERT model to analyze sentiment in news and provide decision-making insights for the stock market. They manually labeled 582 news into positive, negative, and neutral categories to precisely fine-tune the proposed model. As a result, they achieved an accuracy of 72.5% after meticulous model calibration. In another research paper, Hiew et al. (2019) introduced a textual financial sentiment index based on a BERT model. They utilized this index to investigate three stocks in the Hong Kong market and demonstrated that their model outperformed traditional deep learning models in terms of accuracy. In a recent study conducted by Warner et al. (2020), they presented and examined a combined model of Genetic/Support Vector Regression algorithm and BERT model. They applied this model to predict the closing price of the Dow Jones Industrial Average by analyzing news. Their findings revealed that incorporating both the research headlines and their content improved the mean squared error by up to 36.5%.

In the study conducted by Jing et al. (2021), they merged a hybrid model based on deep learning with a sentiment analysis model to predict stock prices. They utilized a CNN model for classifying investors' hidden sentiments. Subsequently, they employed a combined research model using the LSTM approach to analyze technical indicators from the stock market and incorporate the results of sentiment analysis. In another study, Rezaei et al. (2021) presented a novel combination algorithm that integrates LSTM, CNN, Empirical Mode Decomposition (EMD), and Complete Ensemble Empirical Mode Decomposition (CEEMD) methods. The research findings suggest that CNN, alongside LSTM and CEEMD or EMD, can enhance prediction accuracy and outperform other counterparts.

A study conducted by Zhao and Chen (2022) presents a combined deep learning model for stock price prediction. In their study, they employed a fusion of Autoregressive Integrated Moving Average (ARIMA), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) models. The findings indicate that the hybrid model can effectively predict both the linear and non-linear components of stock dataset time series.

In a study conducted by Yadav et al. (2022), a deep learning model was developed to predict stock value. Their proposed model combines Recurrent Neural Network (RNN) and LSTM error method. Another study by Zaheer et al. (2023) introduces a deep learning model that combines LSTM and CNN methods to predict stock prices, specifically the next day's close price and high price. The research findings reveal that CNN performs poorly compared to other models. In the study conducted by Patel Krishne Gowda (2023), they present hybrid neural network and deep learning models for stock price prediction. The results demonstrate that the proposed model achieves favorable performance in estimating stock prices. Table 1 summarizes the key studies reviewed in the field of financial market prediction using machine learning techniques.

**Table 1- Reviewed articles based on the use of machine learning techniques**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Author | Publish year | Case study | Asset type | Predictor variable | Predictive variable | Technique in use |
| Chen, Y., & Hao, Y | 2017 | China | Index | Technical analysis | Direction | SVM, KNN |
| Weng et al | 2017 | USA | Index | Technical analysis, text | Direction | Neural Network, SVM, DT |
| Barak et al. | 2017 | Iran | Stock | Fundamental analysis | Return and risk | Neural Network, SVM, DT |
| Chen et al. | 2014 | Taiwan and Hong Kong | Index | Technical analysis | Price | Fuzzy logic |
| Lahmiri | 2014 | USA | Index | Technical analysis | Direction | SVM |
| Huang | 2012 | Taiwan | Stock | Fundamental analysis | Price | SVM, GA |
| Barak, S., & Modarres, M. | 2015 | USA-NYSE | Stock | Technical analysis | Price | ICA,ANFIS |
| Wang et al. | 2012 | China, USA | Index | Price | Price | Neural Network, GA |
| Kumar, M., & Thenmozhi, M. | 2014 | India | Index | Return | Return | Neural Network, SVM, DT, ARIMA |
| Yan et al. | 2017 | China | Index | Price | Price | Neural Network |
| Chiang et al. | 2016 | Hybrid | Index | Technical analysis | Direction | Neural Network |
| Xiao et al. | 2014 | China | Stock | Technical analysis | Index | WNN,PSO,GRNN,SVM,GA |
| Kara et al. | 2011 | Turkey | Index | Technical analysis | Direction | ANN, SVM |
| Patel, N. S., & Krishne Gowda, Y. T. | 2015 | India | Stock | Technical analysis | Direction | Neural network, SVM,RF,NB |
| Krauss et al. | 2017 | USA | Stock | Price | Return | Neural network, RF, DT |
| Huang, C.-L., & Tsai, C.-Y. | 2009 | Taiwan | Index | Technical analysis | Price | SVM |
| Naik, N., & Mohan, B. R. | 2019 | India | Stock | Technical analysis | Price | Neural network, SVM, Deep learning |
| Jang et al | 2020 | USA- Dow Jones | Index | Macro economy index, news | Direction | BERT, NLTK, VADER |
| Sousa et al. | 2019 | USA-Dow Jones | Index | Text news | Direction | BERT |
| Hiew et al. | 2019 | Hong Kong | Stock | Text news | Return | BERT, LSTM |
| Warner et al. | 2020 | USA- Dow Jones | Index | Text news | Price | BERT, GA, SVR |
| Jing et al. | 2021 | China | Stock | Technical analysis | Price | LSTM و CNN |
| Rezaei et al. | 2021 | Sample dataset | Stock | Technical analysis | Price | LSTM, CNN, CEEMD, EMD |
| Zhao, Y., & Chen, Z. | 2022 | S&P 500 | Stock | Price | Price | ARIMA, CNN, LSTM |
| Yadav et al. | 2022 | Apple, Facebook, Nike and Uber's stock | Stock | Technical analysis | Price | LSTM, RNN |
| Zaheer et al. | 2023 | China | Stock | Fundamental analysis | Price | LSTM,CNN |
| Patel, N. S., & Krishne Gowda, Y. T. | 2023 | India | Stock | Technical analysis | Price | Neural Network, Deep learning |
| Current study | | USA-NASDAQ | Stock | Technical analysis, fundamental analysis, text news | Price | BERT, FinBERT |

Based on the conducted studies, it has been found that the majority of research conducted in the field of stock market trend prediction has utilized technical and fundamental indicators as input variables. However, it is important to note that the stock market is not solely influenced by these indicators, as there exist numerous factors that can significantly impact stock price movements. This aspect becomes particularly evident in an efficient market, where the trends of stocks are influenced not only by technical and fundamental indicators but also by the dissemination of news. In this research, instead of relying on traditional machine learning models, novel deep learning models developed in recent years have been employed to incorporate the influence of published news and trader sentiments. Furthermore, as an innovative approach, the proposed model in this study, known as the multi-label model, has been introduced alongside other commonly used models in previous research, allowing for a comprehensive comparison of their respective performance. Therefore, the innovations of the present research can be outlined as follows:

1- Simultaneous prediction of stock market trends using a combination of technical and fundamental indicators, along with the incorporation of published news.

2- Integration of advanced deep learning models to capture the influence of published news and trader sentiments on a daily basis.

3- Introduction of a novel multi-label model and conducting a comparative analysis with other prevalent models in this field.

Furthermore, the assumptions of the current research are as follows:

1- In order to capture a representative value of the day's trading transactions, the closing price is utilized.

2- Due to the unavailability of real-time price data, it is assumed that all transactions occur at the closing price.

3- To examine the extent of the impact of news release timing on stock price fluctuations, all models are trained twice: once under normal conditions, without considering any time lag, and once with a one-day time lag taken into account. The results obtained from these models are then thoroughly examined. The concept of "time lag" implies that the influence of news released on a particular day is observed in the subsequent day.

**3- Data and Software**

**3-1- Data**

This research focuses on predicting stock prices by incorporating news related to the given stocks. The study examines Tesla's stock price data from the American Nasdaq market, covering the time period from January 1, 2020, to November 13, 2020. Additionally, credible news sources pertaining to Tesla's stock are analyzed to determine the expected movement of its closing price in the future. The news has been categorized using four different labeling methods for this purpose. Historical data regarding Tesla stock prices has been collected from January 16, 2020, to November 13, 2020 (a total of 318 days), including opening, closing, high, and low prices. The data was obtained from the website www.yahoofinance.com (a portion of the database is provided in Appendix 1). Additionally, news data relevant to these stocks, in textual format, was gathered from reputable news websites such as cnbc.com, Zacks.com, Benzinga, StockNews, InvestorsObserver.com, etc. Among the examined days, some had a higher number of news, while others had fewer or no news coverage. As a result, the news section consists of 1,931 rows.

**3-1- Software**

In the Python software, the identification of appropriate features and the analysis of news are conducted using a BERT deep learning model. The accuracy of the model's results is evaluated using F1 and Accuracy metrics. Python is a high-level, object-oriented programming language with dynamic semantics, widely used for web development and creating functional software applications. Python's rapid application development capabilities make it highly appealing in various domains.

**4- Methodology**

The primary approach in this paper involves around utilizing an unsupervised natural language processing model known as BERT. This algorithm, when properly fine-tuned, can accurately implement 11 common natural language processing tasks and serve as an effective tool for processing and comprehending natural language. BERT is a deeply bidirectional algorithm, carefully considering both the left and right context of a phrase or sentence.

This implies that the algorithm accurately examines the words preceding and following each word, matching them with the information provided in Wikipedia. This enables the algorithm to obtain a precise comprehension of the intended concept that the user is seeking.

**4-1- Initial Analysis of the News Data**

The gathered data for creating the database in question consists of several columns, including the news date and publication day, the title and full description of the news article, the opening, high, low, and closing prices, as well as four distinct labels for article classification. These four classification methods are outlined as follows:

1- This label categorizes the news data into three categories based on the percentage change in the closing price compared to the opening price on the day of news publication.

2- This label classifies the news data into three categories based on the percentage change in the closing price on the day of news publication compared to the closing price on the previous day.

3- This label classifies the news data into three categories by calculating the delta between the closing price on the day of news publication and the closing price on the previous day, divided by the average true range indicator on the day prior to news publication.

4- This label is created using sentiment analysis and the FinBert sentiment analysis model. All news and publications are categorized into three labels using this analysis model. Additionally, to determine the final label for each day, a weighted average of the labels for all news is calculated. News labeled as "Bad" have a weight of zero, news labeled as "Neutral" have a weight of 0.5, and news labeled as "Good" have a weight of 1. Afterward, the average weighted labels for all published news of that day are calculated. If the weighted average falls within the range of 0 to 0.33, the overall label for that day is considered "Bad". If it falls within the range of 0.33 to 0.66, the overall label is classified as "Neutral". Finally, if it falls within the range of 0.66 to 1, the overall label is determined as "Good".

The categorization method and the frequency of news data for each category, based on their respective labels, are illustrated in Table 2.

Table 2- Classification method and number of news data of each category based on the first tag

|  |  |  |  |
| --- | --- | --- | --- |
| Label's name | Class name | The equation for choosing classes | Number of observations |
| Label 1 | Bad |  | 51 |
| Neutral |  | 178 |
| Good |  | 52 |
| Label 2 | Bad |  | 42 |
| Neutral |  | 187 |
| Good |  | 52 |
| Label 3 | Bad |  | 23 |
| Neutral |  | 218 |
| Good |  | 40 |
| Label 4 | Bad | The weighted average of news should be in the range of 0 to 0.33 | 58 |
| Neutral | The weighted average of news should be in the range of 0.33 to 0.66 | 158 |
| Good | The weighted average of news should be in the range of 0.66 to 1 | 65 |

**4-2- The FinBERT Model**

Due to the lack of labeled training data in the field of financial text mining, the utilization of deep learning and pre-trained models is often considered unsuccessful. As a result, the FinBERT model, an extension of the BERT approach, was introduced by Zhang et al. (2020) and it was trained and deployed on a substantial amount of financial data provided by large and specialized financial companies. The FinBERT model has demonstrated better results than the original BERT model in the domain of financial text mining, making it an essential component of this study alongside the original BERT model.

To investigate the influence of news on stock prices and classify news into different categories, the current article employs the FinBERT model using two distinct methods:

1- Multi-class Approach: In this approach, the model is trained on the first three labels (associated with stock prices) separately, and the resulting accuracy and outcomes are presented individually for each label.

2- Multi-label: In this approach, all the existing labels related to each news are combined into a single binary vector. In general, the objective of this method is to find a model that can map the input values, x, to binary vectors, y, where each element (label) in y is assigned a value of either 0 or 1. For instance, if the values of the first to fourth labels for a given news are "Good", "Neutral", "Good", and "Bad", respectively, the resulting vector in the multi-label model would be:

(1,0,0,0,1,0,1,0,0,0,0,1)

**4-3- Evaluation** **Metrics for the Model**

In this research, the evaluation of the proposed models and the comparison of their performance under various conditions have been carried out using the criteria of accuracy and F1 score.

The accuracy metric was chosen for its ease of comprehension and use. However, it is essential to note that this metric is generally not applied to multi-label models. The reason behind this lies in the presence of multiple diverse labels and possible label displacements in multi-label models, making accuracy a relative and subjective measure. In other words, when assessing the accuracy of a multi-label task, the question arises: should the accuracy be calculated for each label separately, focus solely on the overall correctness of all labels, or even consider a relative measure of label correctness? Consequently, this metric is less commonly utilized in multi-label models.

**5- Results**

In this section, the influence of news release timing on stock price changes is investigated. All models were trained twice: once in the normal state, without considering any time lag, and again with the inclusion of a one-day time lag. The results from these models were examined and analyzed.

**5-1- Models Training**

To train the described models, the input and output data need to be separated. The input data consists of a list containing the concatenation of each element from the Headline column with each element from the Article column. The output of the models is represented by a list consisting of three classes: "good," "bad," and "neutral," each assigned a respective subscript. For instance, the label "good" is assigned a subscript of 0, "bad" is assigned a subscript of 1, and "neutral" is assigned a subscript of 2. Subsequently, the data is split into training and testing sets using a Shuffle, with an 80% to 20% ratio for training and testing, respectively. Tokenization is then applied to the training and testing data, followed by the initiation of the model training process through the invocation of the training function.

**5-2- Multi-class Model Outputs**

In this section, the results obtained from training the introduced models on different labels individually are presented. All FinBERT models in this section are trained to evaluate their performance and monitor the accuracy changes during 10 training epochs. However, it should be noted that the results of training the model on the fourth label, which is based on sentiment analysis of news, are not included here. This is because this particular label does not contribute to predicting stock prices effectively. Even if the model achieves high accuracy and performs well on this label, it does not provide valuable insights for predicting future stock prices, as it solely reflects the sentiment derived from the news rather than a direct price correlation.

**5-2-1- Results based on the First Label without Considering Time Lag**

The model's output, which is calculated without considering a one-day time lag, is based on the percentage change between the closing and opening prices on the news release day. For the FinBERT model, the results are presented in Table 3. It can be observed that the model's loss decreases during the validation phase until the third epoch, but starts increasing thereafter. This suggests over-fitting of the model and a misleading increase in accuracy from the third epoch onwards. Hence, the third epoch is used as the benchmark for evaluating this model. The accuracy of this model is reported as 79.06%, with an F1 score of 0.688.

**Table 3- The results of FinBERT model training on the first label without considering the time lag**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Epoch | Train Loss | Validation Loss | Train Accuracy | Validation Accuracy | Train F1 | Validation F1 |
| 1 | 0.396679 | 0.338377 | 75.123217 | 78.80829 | 0.642538 | 0.666613 |
| 2 | 0.360494 | 0.338036 | 78.171206 | 79.326425 | 0.672803 | 0.698258 |
| 3 | 0.350466 | 0.335994 | 79.079118 | **79.067358** | 0.697719 | **0.688967** |
| 4 | 0.340997 | 0.350019 | 80.505837 | 78.80829 | 0.727215 | 0.685371 |
| 5 | 0.335302 | 0.361398 | 80.376135 | 78.80829 | 0.73773 | 0.690029 |
| 6 | 0.330712 | 0.371598 | 81.54345 | 79.067358 | 0.761405 | 0.698856 |
| 7 | 0.324801 | 0.388799 | 82.191958 | 80.103627 | 0.772527 | 0.713495 |
| 8 | 0.318089 | 0.406881 | 82.905318 | 80.621762 | 0.781688 | 0.722321 |
| 9 | 0.311521 | 0.416292 | 83.294423 | 80.362694 | 0.786531 | 0.725447 |
| 10 | 0.303818 | 0.426971 | 84.785992 | 79.585492 | 0.807435 | 0.717956 |

**5-2-2- Results Based on the Second Label without Considering Time Lag**

The model's output, calculated without considering a one-day time lag, is based on the percentage change between the closing price on the news release day and the closing price on the day prior to the news release. For the FinBERT model, the results are in accordance with Table 4. It can be observed that the loss of this model increases during the validation phase starting from the second epoch. Therefore, the first epoch is considered as the evaluation benchmark for this model. The accuracy of this model is reported as 83.21%, with an F1 score of 0.775.

**Table 4- The results of FinBERT model training on the second label without considering the time lag**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Epoch | Train Loss | Validation Loss | Train Accuracy | Validation Accuracy | Train F1 | Validation F1 |
| 1 | 0.3838332 | 0.35015372 | 76.2905318 | **83.2124352** | 0.6609231 | **0.7754914** |
| 2 | 0.3461848 | 0.36261114 | 79.2088197 | 80.10362694 | 0.6948343 | 0.74549173 |
| 3 | 0.3350237 | 0.3584517 | 80.6355383 | 78.5492228 | 0.7250324 | 0.74161548 |
| 4 | 0.325636 | 0.35111427 | 81.6083009 | 79.32642487 | 0.7466308 | 0.74357803 |
| 5 | 0.3187976 | 0.35366659 | 82.3216602 | 79.84455959 | 0.7629181 | 0.7461751 |
| 6 | 0.3151649 | 0.35884066 | 82.6459144 | 78.80829016 | 0.7727859 | 0.74116395 |
| 7 | 0.3098916 | 0.36188346 | 83.4241245 | 80.3626943 | 0.7862847 | 0.74977459 |
| 8 | 0.3047847 | 0.37348839 | 83.0998703 | 78.03108808 | 0.7875875 | 0.73796648 |
| 9 | 0.2993594 | 0.38960819 | 83.2944228 | 75.95854922 | 0.7930631 | 0.72675628 |
| 10 | 0.2922463 | 0.40197894 | 84.3968872 | 75.44041451 | 0.8084087 | 0.7226851 |

**5-2-3- Results Based on the Third Label without Considering Time Lag**

The output of this model, calculated without considering a one-day time lag, is based on the delta between the closing price on the news release day and the closing price on the day prior to the news release, divided by the average true range indicator on the day prior to the news release. For the FinBERT model, the results align with Table 5. It is evident that the model's loss increases during the validation phase starting from the second epoch. Therefore, the evaluation benchmark for this model is set as the first epoch. The accuracy of the model is reported as 91.72%, with an F1 score of 0.822. It should be noted that, due to their poor performance, the results obtained from training the BERT model have been omitted from this section onwards.

**Table 5- The results of model training on the third label without considering the time lag**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Epoch | Train Loss | Validation Loss | Train Accuracy | Validation Accuracy | Train F1 | Validation F1 |
| 1 | 0.297258 | 0.2718267 | 86.92607 | **91.7253886** | 0.7676493 | **0.822831** |
| 2 | 0.2617956 | 0.2766144 | 89.714656 | 88.39378238 | 0.8109056 | 0.8223436 |
| 3 | 0.2520991 | 0.2804755 | 90.103761 | 88.65284974 | 0.8262115 | 0.8277789 |
| 4 | 0.2434021 | 0.2853489 | 90.817121 | 86.83937824 | 0.8396939 | 0.8157715 |
| 5 | 0.2366425 | 0.293588 | 90.817121 | 87.0984456 | 0.8456256 | 0.819269 |
| 6 | 0.231297 | 0.3099501 | 91.595331 | 86.83937824 | 0.8561951 | 0.8232611 |
| 7 | 0.22545 | 0.3251992 | 92.30869 | 85.54404145 | 0.8710541 | 0.8227298 |
| 8 | 0.2172837 | 0.3348161 | 92.892348 | 83.98963731 | 0.8834828 | 0.8118056 |
| 9 | 0.2091004 | 0.3459997 | 93.670558 | 81.91709845 | 0.8928977 | 0.8006229 |
| 10 | 0.2010751 | 0.3585061 | 93.929961 | 81.65803109 | 0.8989017 | 0.8037375 |

**5-2-4- Results Based on the First Label with a One-Day Time Lag**

The results of this model, considering a one-day time lag, are presented in Table 6. It is observed that the model's loss increases during the validation phase from the fourth epoch onwards. Therefore, the third epoch is chosen as the evaluation benchmark for this model. The accuracy of the model is recorded as 75.96%, with an F1 score of 0.6814.

**Table 6- The results of model training on the first label considering one day time lag**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Epoch | Train Loss | Validation Loss | Train Accuracy | Validation Accuracy | Train F1 | Validation F1 |
| 1 | 0.3676341 | 0.3585845 | 80.830091 | 76.735751 | 0.6907875 | 0.6703282 |
| 2 | 0.3364252 | 0.3568041 | 81.608301 | 75.958549 | 0.7011396 | 0.6521973 |
| 3 | 0.3269819 | 0.3536559 | 82.840467 | **75.96321** | 0.7273983 | **0.681404** |
| 4 | 0.318584 | 0.371321 | 82.581064 | 74.404145 | 0.7391335 | 0.6825197 |
| 5 | 0.3152119 | 0.388744 | 82.775616 | 73.88601 | 0.7530258 | 0.6735074 |
| 6 | 0.3113212 | 0.4079304 | 82.645914 | 74.663212 | 0.7621104 | 0.6769119 |
| 7 | 0.3081656 | 0.4145566 | 83.81323 | 74.145078 | 0.7829373 | 0.6760955 |
| 8 | 0.304963 | 0.4276242 | 83.942931 | 74.663212 | 0.7883828 | 0.6876913 |
| 9 | 0.2992634 | 0.4271822 | 84.461738 | 73.367876 | 0.7990494 | 0.6813625 |
| 10 | 0.2925999 | 0.4272819 | 85.175097 | 71.813472 | 0.7085788 | 0.6754202 |

**5-2-5- Results Based on the Second Label with a One-Day Time Lag**

The output of this model, taking into account a one-day time lag, is displayed in Table 7. As evident, the model's loss increases during the validation phase starting from the third epoch. Consequently, the second epoch is selected as the evaluation reference for this model. The accuracy of the model is reported as 78.29%, with an F1 score of 0.635.

**Table 7- The results of model training on the second label considering one day time lag**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Epoch | Train Loss | Validation Loss | Train Accuracy | Validation Accuracy | Train F1 | Validation F1 |
| 1 | 0.3784268 | 0.3694396 | 78.560311 | 77.772021 | 0.6859886 | 0.6258021 |
| 2 | 0.3447171 | 0.3687085 | 81.738003 | **78.290155** | 0.7070768 | **0.6352464** |
| 3 | 0.3338749 | 0.3714403 | 81.997406 | 78.03109 | 0.718904 | 0.654899 |
| 4 | 0.325996 | 0.3745548 | 82.256809 | 77.253886 | 0.7301781 | 0.6491474 |
| 5 | 0.3199449 | 0.3783104 | 82.710765 | 77.772021 | 0.7444451 | 0.6634237 |
| 6 | 0.3121434 | 0.3829518 | 83.424125 | 78.549223 | 0.7660735 | 0.6742015 |
| 7 | 0.3017951 | 0.3908426 | 84.007782 | 76.994819 | 0.7804115 | 0.6638768 |
| 8 | 0.2923263 | 0.4037426 | 85.499351 | 76.994819 | 0.8026458 | 0.6605105 |
| 9 | 0.2850283 | 0.4200127 | 85.564202 | 76.217617 | 0.8069982 | 0.6562378 |
| 10 | 0.2787136 | 0.4356667 | 85.953307 | 75.699482 | 0.8137111 | 0.6554298 |

**5-2-6- Results Based on the Third Label with a One-Day Time Lag**

The results of this model, taking into account a one-day time lag, are shown in Table 8. As observed, the model's loss increases during the validation phase from the second epoch onwards. Consequently, the initial epoch is selected as the evaluation benchmark for this model. The accuracy of the model is recorded as 91.53%, with an F1 score of 0.830.

**Table 8- The results of model training on the third label considering one day of time lag**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Epoch | Train Loss | Validation Loss | Train Accuracy | Validation Accuracy | Train F1 | Validation F1 |
| 1 | 0.2487409 | 0.2675466 | 93.281453 | **91.53886** | 0.8483793 | **0.83063** |
| 2 | 0.2165519 | 0.2823489 | 95.032425 | 91.502591 | 0.8636007 | 0.8444658 |
| 3 | 0.207459 | 0.2916725 | 94.773022 | 89.170984 | 0.8680023 | 0.8289359 |
| 4 | 0.1968077 | 0.2954153 | 95.162127 | 89.170984 | 0.8797535 | 0.8334721 |
| 5 | 0.18836 | 0.2965869 | 95.551232 | 89.689119 | 0.8923452 | 0.8444621 |
| 6 | 0.1832566 | 0.3073274 | 95.291829 | 88.134715 | 0.896081 | 0.8380232 |
| 7 | 0.1760055 | 0.3229505 | 95.810636 | 88.134715 | 0.9033059 | 0.836588 |
| 8 | 0.1688942 | 0.348344 | 96.199741 | 89.170984 | 0.9107937 | 0.8391093 |
| 9 | 0.1613287 | 0.3766214 | 96.977951 | 87.357513 | 0.9224503 | 0.8245823 |
| 10 | 0.1552799 | 0.3847452 | 97.885863 | 86.321244 | 0.9337444 | 0.8183235 |

**5-3- Results of Multi-label Model Outputs**

In this section, the results obtained from the multi-label model are presented under two scenarios: one without considering a one-day time lag and the other with considering a one-day time lag. The models have been trained for up to 20 epochs to evaluate their performance and observe changes in accuracy. It should be noted that the accuracy metric, as explained in the previous chapter, is not suitable for evaluating multi-label models. Therefore, in this section, the focus is solely on presenting the results based on the F1 score metric.

**5-3-1- Results based on the Multi-label Model without Considering Time Lag**

The results obtained from this model, as shown in Table 9, indicate an increasing loss during the validation phase starting from the second epoch. Therefore, the initial epoch is considered as the benchmark for evaluating this model. The F1 score for this model is observed to be 0.846.

**Table 9-The results of training the Multi-label model without considering the time lag**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Epoch | Train Loss | Validation Loss | Train F1 | Validation F1 |
| 1 | 0.408131159 | 0.409063196 | 0.844353624 | **0.846252158** |
| 2 | 0.376546806 | 0.41039062 | 0.871530469 | 0.847348489 |
| 3 | 0.369528339 | 0.413517418 | 0.881213688 | 0.846236378 |
| 4 | 0.364134793 | 0.423024662 | 0.889346123 | 0.85688391 |
| 5 | 0.359348158 | 0.43799111 | 0.896992855 | 0.850999754 |
| 6 | 0.355012413 | 0.449839145 | 0.908967587 | 0.858442879 |
| 7 | 0.350747954 | 0.456306552 | 0.925996833 | 0.856483922 |
| 8 | 0.347014601 | 0.462359748 | 0.927265827 | 0.859064956 |
| 9 | 0.343471355 | 0.46522078 | 0.935542001 | 0.858801275 |
| 10 | 0.338750809 | 0.468949928 | 0.941365111 | 0.860017629 |
| 11 | 0.333890834 | 0.472407165 | 0.946639504 | 0.856597842 |
| 12 | 0.329141486 | 0.473942442 | 0.949241117 | 0.85645512 |
| 13 | 0.324418659 | 0.47468871 | 0.957165187 | 0.861331717 |
| 14 | 0.319736015 | 0.47575175 | 0.955000743 | 0.877704603 |
| 15 | 0.314941681 | 0.480051829 | 0.954601073 | 0.898999027 |
| 16 | 0.310250621 | 0.486596401 | 0.958085336 | 0.897746674 |
| 17 | 0.30637083 | 0.4896629 | 0.962289527 | 0.897402775 |
| 18 | 0.303485547 | 0.491891158 | 0.96470784 | 0.888861916 |
| 19 | 0.298483017 | 0.49807474 | 0.970479796 | 0.891353207 |
| 20 | 0.292499553 | 0.503586464 | 0.977102822 | 0.896898095 |

**5-3-2- Results based on the Multi-label Model with Consideration of Time Lag**

The output of this model, as presented in Table 10, demonstrates an increase in loss during the validation phase from the second epoch onwards. As a result, the initial epoch is taken as the benchmark for evaluating this model. It is noted that the F1 score for this model is 0.787.

**Table 10- The results of Multi-label model training considering the time lag**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Epoch | Train Loss | Validation Loss | Train F1 | Validation F1 |
| 1 | 0.476222604 | **0.465180589** | 0.77293761 | **0.787372742** |
| 2 | 0.448352097 | 0.467086565 | 0.780292891 | 0.790772253 |
| 3 | 0.441936767 | 0.472902167 | 0.800309658 | 0.789092658 |
| 4 | 0.437280626 | 0.481499741 | 0.808247148 | 0.783298937 |
| 5 | 0.433197944 | 0.495247636 | 0.826812551 | 0.779778449 |
| 6 | 0.429993182 | 0.510746306 | 0.830874087 | 0.784495394 |
| 7 | 0.425941943 | 0.526218076 | 0.835283619 | 0.781837897 |
| 8 | 0.422470952 | 0.54231405 | 0.835499925 | 0.778597654 |
| 9 | 0.41924737 | 0.558362939 | 0.839779951 | 0.791640425 |
| 10 | 0.416174996 | 0.572874572 | 0.843163374 | 0.804174342 |
| 11 | 0.412981656 | 0.583431389 | 0.84593094 | 0.815498749 |
| 12 | 0.409293984 | 0.591076787 | 0.853087738 | 0.813488477 |
| 13 | 0.405211609 | 0.595724967 | 0.856158892 | 0.813114582 |
| 14 | 0.400918837 | 0.598614858 | 0.858519904 | 0.808634165 |
| 15 | 0.396453762 | 0.601682543 | 0.86256665 | 0.808239631 |
| 16 | 0.391658188 | 0.603695679 | 0.861784752 | 0.810380864 |
| 17 | 0.385860409 | 0.599181179 | 0.870026532 | 0.808452556 |
| 18 | 0.379958685 | 0.59009136 | 0.878406744 | 0.816064374 |
| 19 | 0.375096385 | 0.57722574 | 0.887100134 | 0.809309801 |
| 20 | 0.37080359 | 0.569141594 | 0.879795275 | 0.824666599 |

**5-4- Comparison Study**

Considering the presented results and the performance evaluation of the different models, the analysis of the findings is outlined below. Referring to the results in Table 11, which pertain to the outcomes of the multi-class models, it is evident that the multi-class model without considering the time lag exhibits the highest accuracy and F1 score. This indicates that the third label outperforms the first and second labels. The reason for this can be attributed to technical analysts' increased utilization of the average true range indicator for predicting stock price movements and setting profit targets. As mentioned earlier, the average true range indicator, based on historical and current market fluctuations, accurately forecasts potential future price movements, surpassing the accuracy achieved solely by considering price movements.

**Table 11- Examining the results of Multi-class models**

|  |  |  |
| --- | --- | --- |
| Model name | Accuracy | F1 |
| Multiclass model on the first label without considering the time lag | **79.06736** | **0.688967** |
| Multiclass model on the second label without considering the time lag | **83.21244** | **0.775491** |
| Multiclass model on the third label without considering the time lag | **91.72539** | **0.822831** |
| Multiclass model on the first label considering the time lag | **75.96321** | **0.681404** |
| Multiclass model on the second label considering the time lag | **78.29016** | **0.635246** |
| Multiclass model on the third label considering the time lag | **91.53886** | **0.82063** |

Additionally, based on Table 11, it is evident that the accuracy and F1 score metrics for all multi-class models are higher when a one-day time lag is not considered compared to when it is considered. This can be attributed to the efficient-market hypothesis. The efficient-market is characterized by the rapid incorporation of publicly available information into stock prices, causing prices to adjust accordingly. According to this hypothesis, when specific news about a stock is released in news agencies, it is expected that the stock price will react rapidly to the news, resulting in either an upward or downward movement depending on the nature of the news. Therefore, as anticipated, the performance of all models without any interruption or one-day time lag is superior.

Furthermore, based on the results provided in Table 12, which represent the outcomes of Multi-label models, it is evident that the multi-label model, similar to the multi-class models without considering a time lag, demonstrates a higher level of accuracy, which is also relatively significant.

**Table 12- Examining the results of Multi-label models**

|  |  |  |
| --- | --- | --- |
| Model name | Accuracy | F1 |
| Multi-label model without considering time lag | - | **0.846252158** |
| Multi-label model considering time lag | - | **0.787372742** |

Comparing the results from these models based on Tables 11 and 12, it is evident that the proposed model in this study, the multi-label model, performs better than the best result obtained by the multi-class model. The multi-label model achieves an impressive F1 score of 0.846, surpassing the best F1 score of 0.822 achieved by the multi-class model.

Finally, to validate the results, the performance of the proposed model in this study, namely the multi-class model on the third label (without considering the time lag and with considering the time lag), was examined and compared to findings from previous research studies.

**Table 13- Comparison of the results of the proposed Multi-class model with other previous studies**

|  |  |  |  |
| --- | --- | --- | --- |
| Author/Year | Model name | Accuracy | F1 |
| Barak et al.(2017) | Return Forecasting in TSE-Iran (Fusion of Multiple Diverse Predictors in Stock Market) | 83.65 | - |
| Sousa et al.(2019) | BERT for Stock Market Sentiment Analysis | 82.5 | 0.725 |
| Wang et al.(2018) | Financial Market Prediction using Deep Random Subspace Ensembles (DRSE) | 91.0 | 0.918 |
| Barak et al.(2015) | Return Forecasting in TSE-Iran | 80.24 | - |
| Current study | Multi-class model on the third label without considering the time lag | 91.72539 | 0.822831 |
| Current study | Multi-class model on the third label considering the time lag | 91.53886 | 0.82063 |

According to the results presented in Table (13), the proposed model achieved an accuracy of 91.725 and 91.538 without and with considering the time lag, respectively. These values outperformed the proposed models in studies conducted by Barak et al. (2015), Sousaet al. (2019), Wang et al. (2018), and Barak and Modarres (2015). Furthermore, among the previous studies, the proposed model by Wang et al. (2018) known as DRSE achieved the highest accuracy of 91.0.

**6- Conclusion and Future Research**

In this research, the focus was on developing a combined model using deep learning models to examine the impact of published news on the price movement and stock prediction of Tesla in the American Nasdaq market. To achieve this, two algorithms were utilized: BERT, a pre-trained model, and FinBERT, a model specifically trained on financial data. These algorithms were trained on news data related to Tesla stocks. To gather the necessary data, all news articles pertaining to Tesla stocks from January 16, 2020, to November 13, 2020, were collected, covering a total of 318 days. The data included opening and closing prices, as well as the highest and lowest prices, sourced from reputable news websites and were labeled using four different labels. Subsequently, the proposed models were trained to classify the news based on these assigned labels. This process was conducted using both the multi-class and multi-label models. The results indicate that the introduced multi-label model in this research performed better than the commonly used multi-class models in other studies, achieving an F1 score of 0.846. This suggests its suitability for other research and practical applications.

In conclusion, future research recommendations include label modifications, considering high or low prices instead of closing prices, incorporating alternative news sources such as tweets instead of financial news, adjusting the number of classes, and investigating stocks of other companies and financial markets.

**6. References**

Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied intelligence*, *6*(3), 205-213.

Alwidian, S. a. A., Bani-Salameh, H. A., & Alslaity, A. a. N. (2015). Text data mining: a proposed framework and future perspectives. *International Journal of Business Information Systems*, *18*(2), 127-140.

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques–Part II: Soft computing methods. *Expert Systems with Applications*, *36*(3), 5932-5941.

Barak, S., Arjmand, A., & Ortobelli, S. (2017). Fusion of multiple diverse predictors in stock market. *Information Fusion*, *36*, 90-102.

Barak, S., Dahooie, J. H., & Tichý, T. (2015). Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese Candlestick. Expert Systems with Applications, 42(23), 9221-9235.

Barak, S., & Modarres, M. (2015). Developing an approach to evaluate stocks by forecasting effective features with data mining methods. Expert Systems with Applications, 42(3), 1325-1339.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307-327.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of financial economics*, *18*(2), 373-399.

Cavalcante, R. C., Brasileiro, R. C., Souza, V. L., Nobrega, J. P., & Oliveira, A. L. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, *55*, 194-211.

Chen, A.-S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & operations research*, *30*(6), 901-923.

Chen, Y.-S., Cheng, C.-H., & Tsai, W.-L. (2014). Modeling fitting-function-based fuzzy time series patterns for evolving stock index forecasting. *Applied intelligence*, *41*(2), 327-347.

Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*, *80*, 340-355.

Chiang, W.-C., Enke, D., Wu, T., & Wang, R. (2016). An adaptive stock index trading decision support system. *Expert Systems with Applications*, *59*, 195-207.

Dechter, R. (1986). Learning while searching in constraint-satisfaction problems.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179-211.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Fama, E. F. (1991). Efficient capital markets: II. *The journal of Finance*, *46*(5), 1575-1617.

Fernandez-Rodrıguez, F., Gonzalez-Martel, C., & Sosvilla-Rivero, S. (2000). On the profitability of technical trading rules based on artificial neural networks:: Evidence from the Madrid stock market. *Economics letters*, *69*(1), 89-94.

Hiew, J. Z. G., Huang, X., Mou, H., Li, D., Wu, Q., & Xu, Y. (2019). BERT-based Financial Sentiment Index and LSTM-based Stock Return Predictability. *arXiv preprint arXiv:1906.09024*.

Huang, C.-F. (2012). A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing*, *12*(2), 807-818.

Huang, C.-L., & Tsai, C.-Y. (2009). A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. *Expert Systems with Applications*, *36*(2), 1529-1539.

Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & operations research*, *32*(10), 2513-2522.

Jang, E., Choi, H., & Lee, H. (2020). Stock prediction using combination of BERT sentiment Analysis and Macro economy index. *Journal of the Korea Society of Computer and Information*, *25*(5), 47-56.

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, *38*(5), 5311-5319.

Karanikas, H., & Theodoulidis, B. (2002). Knowledge discovery in text and text mining software. *Centre for Research in Information Management, Department of Computation*.

Kim, K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1-2), 307-319.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689-702.

Kumar, M., & Thenmozhi, M. (2014). Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *International Journal of Banking, Accounting and Finance*, *5*(3), 284-308.

Lahmiri, S. (2014). Entropy-based technical analysis indicators selection for international stock markets fluctuations prediction using support vector machines. *Fluctuation and Noise Letters*, *13*(02), 1450013.

Leung, M. T., Daouk, H., & Chen, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International journal of forecasting*, *16*(2), 173-190.

Liu, Z., Huang, D., Huang, K., Li, Z., & Zhao, J. (2020). Finbert: A pre-trained financial language representation model for financial text mining. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI,

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of economic perspectives*, *17*(1), 59-82.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383-417.

Naik, N., & Mohan, B. R. (2019). Stock price movements classification using machine and deep learning techniques-the case study of indian stock market. International Conference on Engineering Applications of Neural Networks,

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603-9611.

Pai, P.-F., & Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*(6), 497-505.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS)*, *27*(2), 1-19.

Sousa, M. G., Sakiyama, K., de Souza Rodrigues, L., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019). BERT for Stock Market Sentiment Analysis. 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI),

Timmons, R., & Lee, K. (2007). Predicting the stock market with news articles. *CS224n Final Report*.

Wang, J.-J., Wang, J.-Z., Zhang, Z.-G., & Guo, S.-P. (2012). Stock index forecasting based on a hybrid model. *Omega*, *40*(6), 758-766.

Wang, Q., Xu, W., & Zheng, H. (2018). Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. Neurocomputing, 299, 51-61.

Warner, B., Crook, A., & Cao, R. (2020). Predicting the DJIA with News Headlines and Historic Data Using Hybrid Genetic Algorithm/Support Vector Regression and BERT. International Conference on Big Data,

Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications*, *79*, 153-163.

Xiao, Y., et al., *Ensemble ANNs-PSO-GA approach for day-ahead stock e-exchange prices forecasting.* International Journal of Computational Intelligence Systems, 2014. **7**(2): p. 272-290.

Xu, S. Y., & Berkely, C. (2014). Stock price forecasting using information from Yahoo finance and Google trend. *UC Brekley*.

Yan, D., Zhou, Q., Wang, J., & Zhang, N. (2017). Bayesian regularisation neural network based on artificial intelligence optimisation. *International Journal of Production Research*, *55*(8), 2266-2287.

Yoon, Y., Swales Jr, G., & Margavio, T. M. (1993). A comparison of discriminant analysis versus artificial neural networks. *Journal of the Operational Research Society*, *44*(1), 51-60.

Yu, L., Chen, H., Wang, S., & Lai, K. K. (2008). Evolving least squares support vector machines for stock market trend mining. *IEEE transactions on evolutionary computation*, *13*(1), 87-102.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, *14*(1), 35-62.

Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. Expert Systems with Applications, 178, 115019.

Rezaei, H., Faaljou, H., & Mansourfar, G. (2021). Stock price prediction using deep learning and frequency decomposition. Expert Systems with Applications, 169, 114332.

Zhao, Y., & Chen, Z. (2022). Forecasting stock price movement: New evidence from a novel hybrid deep learning model. Journal of Asian Business and Economic Studies, 29(2), 91-104.

Yadav, K., Yadav, M., & Saini, S. (2022). Stock values predictions using deep learning based hybrid models. CAAI Transactions on Intelligence Technology, 7(1), 107-116.

Zaheer, S., Anjum, N., Hussain, S., Algarni, A. D., Iqbal, J., Bourouis, S., & Ullah, S. S. (2023). A Multi Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model. Mathematics, 11(3), 590.

Patel, N. S., & Krishne Gowda, Y. T. (2023, January). Hybrid Deep Learning Models for Improving Stock Index Prediction. In Cognition and Recognition: 8th International Conference, ICCR 2021, Mandya, India, December 30–31, 2021, Revised Selected Papers (pp. 282-295). Cham: Springer Nature Switzerland.

**Appendix**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | Day | Headline | Article | Open | High | Low | Close | Label 1 | Label 2 | Label 3 | Label 4 |
| 1/2/2020 | Thursday | Canaccord raises price target for Tesla to $515 | Canaccord raises price target for Tesla to $515 | 84.9 | 86.14 | 84.34 | 86.05 | neutral | neutral | neutral | good |
| 1/2/2020 | Thursday | The Ratings Game: Tesla’s stock rises after analyst hikes price target above $500 | The Ratings Game: Tesla’s stock rises after analyst hikes price target above $500 | 84.9 | 86.14 | 84.34 | 86.05 | neutral | neutral | neutral | good |
| 1/2/2020 | Thursday | Stocks making the biggest moves midday: Tesla, AMD, Peloton Interactive, United Airlines & more | Tesla rose 2.9% after Canaccord Genuity reiterated its buy rating of the electric car maker’s equity and told investors to expect another 23% rally for the stock. The brokerage’s new, $515 price target makes it the biggest Tesla bull of the major brokerages on Wall Street. Tesla is up 89% over the last six months. | 84.9 | 86.14 | 84.34 | 86.05 | neutral | neutral | neutral | good |
| 1/3/2020 | Friday | Tesla delivers more than 367K vehicles in 2019 | Tesla’s vehicle deliveries were up 50 percent over 2018. | 88.1 | 90.8 | 87.38 | 88.6 | neutral | neutral | good | good |