Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 04, December 2022.

COMBINATION OF DEEP-LEARNING MODELS TO FORECAST STOCK PRICE OF AAPL AND TSLA

Zahra Berradi[1], Mohamed Lazaar[2], Oussama Mahboub[1], Halim Berradi[3] and Hicham Omara[4]

(Received: 20-Jun.-2022, Revised: 27-Aug.-2022, Accepted: 19-Sep.-2022)

ABSTRACT

Deep Learning is a promising domain. It has different applications in different areas of life and its application on the stock market is widely used due to its efficiency. Long Short Term Memory (LSTM) proved its efficiency in dealing with time-series data due to the unique hidden unit structure. This paper integrated LSTM with attention mechanism and sentiment analysis to forecast the closing price of two stocks; namely, APPL and TSLA, from the NASDAQ stock market. We compared our hybrid model with LSTM, LSTM with sentiment analysis and LSTM with Attention Mechanism. Three benchmarks were used to measure the performance of the models; the first one is Mean Square Error (MSE), the second one is Root Mean Square Error (RMSE) and the third one is Mean Absolute Error (MAE). The results show that the hybridization is more accurate than the LSTM model alone.

KEYWORDS

Deep learning, Hybrid model, LSTM, Attention mechanism, Sentiment analysis.

1. INTRODUCTION

Since the beginning of the stock market, forecasting the price of stocks is still one of the most challenging tasks for every investor. Searching for and developing new effective technologies are essential. The volatility and non-stationary aspect of the stock market push researchers to find the most rewarding AI technologies to predict its behavior. Based on the published research papers over the last years, we can easily watch the evolution of artificial intelligence and its applications in all aspects of life, especially in the development of stock-market forecasting. Its appearance and improvement go along with AI development. We mentioned in this paper the most significant waves that have affected the stock market in the 20th and the beginning of the 21st century.

We notice the statistical approach as a first wave used for the prediction problem. For example, in 1901, the mathematician Karl Pearson created the method of Principal Component Analysis to reduce the dimension of the input data. Then, other models appeared continuously, such as Auto-regressive Moving Average (ARMA) [1], Auto-regressive Integrated Moving Average (ARIMA) [2], the Generalized Auto-regressive Conditional Heteroskedasticity (GARCH) [3] and Quadratic Discriminant Analysis (QDA) [4].

The second wave is machine learning. It is an extension of AI that uses the input and the output of a specific system as features to be able to solve a prediction problem or a classification problem. It permits a machine to learn automatically without needing to program it explicitly. Many models exist in literature and have proved their efficiency, such as Support Vector Machine (SVM) [5], Genetic Algorithm (GA) and Multi-layer Perceptron (MLP) [6]. Besides those mentioned, there is a famous one: Neural Networks (NN). It is used to solve time-series data problems and has gained popularity among researchers, because it can perfectly handle non-stationary systems, including the stock market.

The development of NN leads to another branch of ML called Deep Learning (DL). It is characterized by a specific hidden-node calculation in the hidden layer. The power of DL remains in the ability to control time-series data with more precision than NN. It shows its efficiency throughout time in many areas, such as speech recognition, image classifications, face recognition [7], natural language processing, sentiment analysis, translation and health care [8]. Some of the well-known types of DL

1. Z. Berradi and O. Mahboub are with National School of Applied Sciences, Abdelmalek Esaadi University, Tetouan, Morocco. Emails: berradi.zahra@gmail.com and mahbouboussama@gmail.com
2. M. Lazaar is with ENSIAS, Mohammed V University in Rabat, Morocco. Email: mohamed.lazaar@ensias.um5.ac.ma
3. H. Berradi is with Telecommunication Systems Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco. Email: halimberradi@gmail.com
4. H. Omara is with Faculty of Sciences, Abdelmalek Esaadi University, Tetouan, Morocco. Email: Hichamomara@gmail.com

models are Deep Multi-layer Perceptron (DMLP), Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM) [9], Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNNs), Restricted Boltzmann Machines (RBMs), Deep Belief Networks (DBNs), Autoencoders (AEs) and Deep Reinforcement Learning (DRL).

DMLP is a model containing three layers: the first is an input layer, the second contains more than one hidden layer and the third is an output layer. RNN differentiates from DMLP by a recurrent unit, but it suffers from the vanishing-gradient problem. Long Short Term Memory (LSTM) was the solution to the vanishing problem. It has demonstrated its capability on speech recognition as mentioned in the works of [10], text classification [11], natural language processing and the stock market fluctuation [12] [13]-[14]. Gated Recurrent Unit (GRU) is a variation of RNN. It is similar to the LSTM unit with some differences in how to calculate the hidden-unit value [15]. CNN was first introduced in image processing, then its application expanded in other fields [16]. It contains several layers; the first is an input layer, the second is a convolutional layer, the next is a max-pooling layer, then a fully connected layer with dropout, while the last layer is the output layer.

The model RBM is used for unsupervised learning, which is applied in classification and dimension reduction. The particularity of RBM is to extract hidden patterns in the system [17]. Deep Belief Network (DBN) is also an unsupervised-learning model composed of RBMs. The model autoencoder (AE) is mainly used for feature extraction [18] and dimensionality reduction. It gets more accurate results when it is combined with LSTM [19]-[20]. In addition, Deep-learning models have proved their efficiency in areas beyond finance. Their efficiency expands into other non-stationary, volatile domains, such as education, robotics, smart cities [21] and health care.

Pattern recognition searches for repetitive patterns in the stock market over time. To extract these patterns, investors may rely on Attention Mechanism (AT), Convolution Neural Network (CNN) and Wavelet Transform (WT) to reach an accurate prediction. To make use of this type of prediction, traders will consider using volume, candlestick charts, as well as Simple and Exponential Moving Averages (SMA, EMA).

The third wave is sentiment analysis (SA). People's opinion on social media, such as Twitter, reflects the behavior and the new direction of the prices on the stock market. By the massive daily information, SA has become more desirable as mentioned by [22]-[23]. The core idea is to get from a sentence or an article the polarity; whether the sentence is positive or negative. CNN is an example of a deep-learning model that is used to get this polarity.

The last wave is hybridization. Data analysts frequently integrate the positive aspects of two, three or more models to create a reliable forecast. This type of model is a powerful tool for predicting the next stock-price movements. For example, LSTM works perfectly with historical prices, but the market is influenced by many other factors, such as the investors' opinion and the internal and external political factors of the country. So, limiting focus only on the historical data will not lead to a better prediction. That is the reason behind combining multiple models.

In other words, two primary business models are applied nowadays to predict stock prices which traders use to make good decisions.

1) Fundamental analysis: it is based on many factors, such as the company's evolution and performance, the politics and the news.
2) Technical analysis: technical analysis depends on the historical prices, the volume and other indicators that depend on the historical price.

In this paper, we are interested in making predictions of stock prices per day using both business models by using the historical price and the news from Twitter. By considering the huge effectiveness of deep learning, our research is about integrating the sentiment analysis, attention mechanism (AM) and LSTM are used to forecast the stock price of both AAPL and TSLA. The aim is to get a more accurate prediction by adding the sentiment of the tweets to the LSTM model with AM and using the results to support the investors' strategy.

This paper is arranged as follows: Section 2 is about related work. In Section 3, we explain the deep-learning model; namely, LSTM, SA and AM. Section 4 is about the experimental results and discussion. As for Section 5, it provides a summary of the findings.

2. RELATED WORK

Researchers always tried to find a developed model to predict the stock market. Many research papers proved the power of deep learning, sentiment analysis and other models to analyze the financial market.

The hybrid model has been widely used in recent years due to the beneficial results of getting more than one strength point of each model for better performance. Jin et al. [24] combined LSTM, sentiment analysis and Empirical Modal Decomposition EMD to forecast the closing price of AAPL. Berradi and Lazaar [25] integrated PCA and RNN to predict the daily closing price of some stocks on the Casablanca Stock Exchange. Ismail and Awjan [26] combined Empirical Modal Decomposition with the Exponential Smoothing Method to improve the forecasting results. Qiu et al. [27] used LSTM, Attention Mechanism and Wallet Transform to predict the stock prices. The proposed model outperforms LSTM, GRU and LSTM with Wallet Transform. Jin et al. [28] proposed a hybrid model to predict the stock price, using LSTM and sentiment analysis applied on Shanghai Stock Exchange (SSE).

On the other hand, the news is a factor that can affect the fluctuation in the stock market beyond other factors, like economy, natural phenomena and politics. The information that can impact the market is characterized by two factors: perfect timing and reliability. For example, on June 7, 2021, the Food and Drug Administration (FDA) admitted Biogen's Alzheimer's treatment. This information directly affected the Biogen (BIIB) stock price; the opening price was 295.35$ and the closing price was 395.85$ on the same day. So, the fact that the historical data of stock prices is the only reasonable way to have an accurate prediction is not valid. The correlation between the stock price and the news has been proven in many research papers.

Rakhi Batra and Sher Muhammad Daudpota [29] integrated StockTwits with sentiment analysis to predict the behavior of AAPL. They used the SVM model (a supervised machine-learning algorithm) to predict the next day's closing price. Mohan et al. [30] improved the predictions of S&P500 by gathering the news related to the mentioned index and the historical data for three years. They used LSTM and Facebook Prophet as models. Kirange et al. [31] emphasized the idea of the news effect on the stock price. They used SVM, Naïve Bayes and KNN to prove the correlation between stock-price movement and the news. Abraham et al. [32] presented an approach to predict the changes in Bitcoin and Ethereum based on Twitter and google trends' data.

Liu et al. [33] proposed a model containing CNN and LSTM to analyze the quantitative strategy in stock markets. In addition, Vargas et al. [34] proposed a hybrid model with CNN and LSTM for the intraday directional movements of S&P500 index using the news and seven technical indicators. Lee et al. [35] suggested a hybrid model called Recurrent Convolutional Neural Network (RCN) that combines CNN, sequence modeling and word embedding. The main work of this model is to extract the polarity from the news, then add technical indicators for stock-price forecasting.

Li et al. [36] stated a hybrid model which combined LSTM and Naïve Bayes. The Naïve Bayes was used for extracting the sentiment from the forum and LSTM for the prediction. Pang et al. [12] proposed an LSTM neural network with an embedded layer and LSTM with an Automatic Encoder neural network to forecast Shanghai A-share composite index and Sinopec. The accuracy of the two models was 57.2% and 56.9%, respectively. The input data is multi-stock high-dimensional historical data. Yan Hongju and Hongbing Ouyang [37] combined wavelet analysis with Long Short Term Memory (LSTM) to forecast the daily closing price of the Shanghai Composite Index. They compared the proposed models with other machine-learning techniques: Multi-layer Perceptron (MLP), Support Vector Machine (SVM) and K-nearest neighbors. They concluded that LSTM with wavelet analysis gives better accuracy. Chen et al. [38] used LSTM with Attention Mechanism to predict the daily return ratio of the HS300 index, while they used the embedding layer to extract the convenient features. Kim et al. [39] proposed a hybrid model with LSTM and GARCH-type models to forecast the volatility of the KOSPI 200 index.

Shen et al. [40] proposed a new model composed of Deep Belief Network (DBN) and Continuous Restricted Boltzmann Machines (CRBMs) to forecast currency exchange rates. They compared their model with Feed Forward Neural Network (FFNN) and found that the proposed model performs better than FFNN.

3. Proposed Model

The power of the hybrid model lies in the combination of different crucial points of each model, which produces a more reliable and robust model. In this section, we provide the mathematical background of each model.

3.1 LSTM Model

LSTM is a type of RNN. [41] invented LSTM to solve the problem of long-term dependencies caused by the traditional RNN. From recently published papers, LSTM is more widely used compared to other deep-learning models. It is applied in various domains, such as speech recognition, sentiment analysis and time-series problems. It is composed of one input layer, one hidden layer and one output layer. The hidden layer is formed of LSTM nodes (as shown in Figure 1).

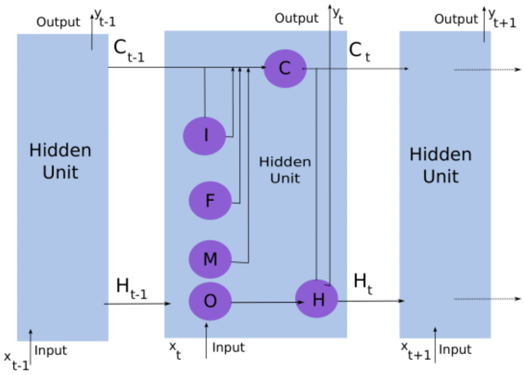

Figure 1. The composition of a hidden node of long short term memory (LSTM).

Each node value is calculated using Equations 1-6:

The forget gate is: $$F_t = \sigma\,(W_F x_t + U_F h_{t-1} + B_F) \quad (1)$$

The second equation, called the output gate: $$O_t = \sigma\,(W_o x_t + U_o h_{t-1} + b_o) \quad (2)$$

The input gate: $$i_t = \sigma\,(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

The state gate: $$S_t = \tanh\,(W_s x_t + U_s h_{t-1} + b_s) \quad (4)$$

The hidden state: $$h_t = O_t * \tanh c_t \quad (5)$$

The cell state: $$c_t = c_{t-1} * F_t + S_t * i_t \quad (6)$$

where x_t the input vector, the function *tanh* is the hyperbolic tangent function, $\sigma = \frac{1}{1+e^{-x}}$ is the sigmoid function, where $x \in$ R, $*$ is the element-wise product.

The other parameters in the hidden node F_t, O_t, i_t, S_t, h_t and c_t are forget gate, output gate, input gate, state gate, hidden state and cell state, respectively, at time *t*.

where B_F, b_o, b_i, b_s are bias. W_F , W_o, W_i, W_s, U_F , U_o, U_i, U_s are the weight matrix. The choice of hyper-parameters is crucial for better performance. LSTM has many hyper-parameters that affect its performance: the number of hidden layers, the number of nodes in each hidden layer, the weight initialization, the learning rate, the activation function, the number of epochs, the bias initialization, the optimization algorithms and the decay rate.

3.2 Sentiment Analysis

Sentiment Analysis is an approach used to extract the polarity from sentences. Many artificial models find the polarity from text, such as CNN (Conventional Neural Network). It is used in computer vision and was invented to deal with image treatments, such as image classification. CNN model is composed of many layers. The first one is the input layer, an input matrix of fixed dimensions. The second one is the convolution layer, the third one is the max-pooling layer, the fourth layer is the fully connected layer with dropout and the last one is the output layer.

Sentiment analysis is used to find whether the sentiment from the tweeter is positive (bullish) or negative (bearish). The first step is to get the tweets from the tweeter. The second step is the pre-processing step, which is about cleaning the text, because it contains a lot of noise, such as URL, emoji's, hashtags and numbers. The third step is word vector embedding. In this step, the method GloVe [42] was used for word representations. The architecture of the CNN is as follows.

3.3 Attention Mechanism

The attention mechanism was inspired by the natural-eye functions which focus only on a specific thing while looking at the whole image. It was first introduced in the Human Vision, then in Natural-language Processing (NLP) [43]-[44]. It also has proved its effectiveness in time-series data, such as speech recognition [45]. This fact led researchers to apply it to the stock-price prediction problem. There are two types of attention: hard attention and soft attention. The first one (hard attention) gives importance to one input element and does the training, which means that more training is required for better accuracy. The soft attention mechanism gives different attention weights to each of the input elements based on their importance (as shown in Figure 3). The work of [44] supposed that each output y_t is a probability of the previous output y_{t-1}, the hidden state s_t (see Equation 7) and the context vector c_t (as shown in Equation 8), where $t \in \{0; ...; T\}$ and f are nonlinear function.

$$s_t = f(c_t, s_{t-1}, y_{t-1}) \tag{7}$$

where s_{t-1} is the previous recurrent hidden state. The determination of the context vector is made by summation of $a_{t,i}$ and h_i.

$$c_t = \sum_{i=1}^{i=T} a_t, ih_i \tag{8}$$

The value of $a_{t,i}$ represents the weight of each hidden h_i, which means how important the h_i can be for our model (as shown in Equation 9).

$$a_{t,i} = \frac{\exp(e_{t,i})}{\sum_{k=1}^{T} \exp(e_{t,k})} \tag{9}$$

$e_{t,i}$ represents the *alignment model*, as shown in Equation 10.

$$e_{t,i} = g(s_{t-1}, h_i) \tag{10}$$

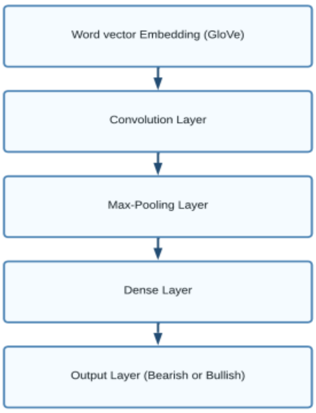

Figure 2. CNN architecture.

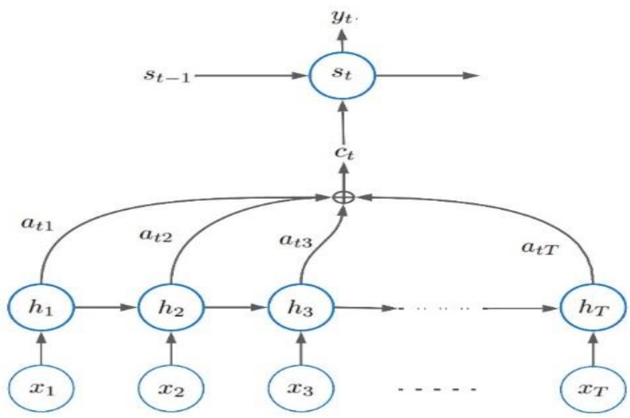

Figure 3. The pattern of attention mechanism taken from [44].

3.4 The Hybrid Model

Long Short Term Memory is widely considered the solution to long-term dependencies. On the other hand, the Attention Mechanism is capable of keeping the most reliving information from the input. Based on the work of Berradi et al. [46], hybridization of many basic models gives more accurate results. The work of Hollis et al. [47] explained that the combination of LSTM and AM models outperforms LSTM alone. These facts lead us to combine multiple deep-learning models to get a more robust one. The proposed model contains the input layer, the LSTM layer, the attention layer, the dense layer and the output layer. Figure 4 represents the different types of layers defined in our deep model.

In the following, the pseudo code used for the hybrid model is presented.

Algorithm 1 The algorithm of the hybrid model

H

1: Input $x = [x_1, \ldots, x_8]$, $x_i \in \mathbb{R}^{512}$
2: Define the attention mechanism model (AttentionDecoder)
3: Build the hybrid model
model = Sequential()
model.add(LSTM($20, activation = ' tanh', input_shape = (8, 1), return_sequences = True$))
model.add($AttentionDecoder(16, 20)$)
model.add(Dense($1, activation = ' tanh'$))
adam = keras.optimizers.Adam($lr = 0.001, beta_1 = 0.9, beta_2 = 0.999, epsilon = None, decay = 0.0$)
model.compile(loss='mse', optimizer='adam') 4:
Predict the closing price of the stock.

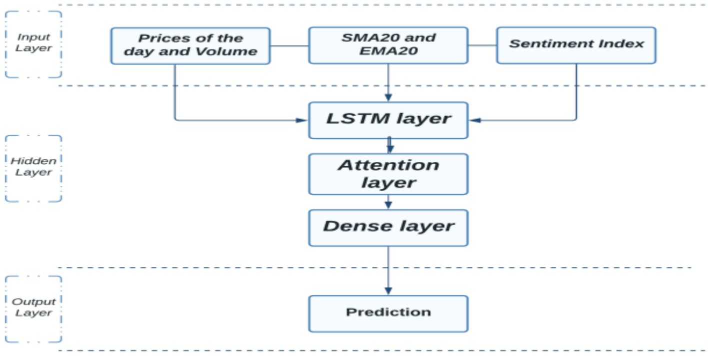

Figure 4. The deep-learning structure.

3.5 Examination Tools

To compare the models, we used benchmarks such as the mean square error (E_{MSE}), the root mean square error (E_{RMSE}) and the mean absolute error (E_{MAE}). The closer the error to zero, the better the model performance.

The mean square error:

$$E_{MSE} = \frac{1}{M}\sum_{i=1}^{M}\left\|\hat{Y}_i - Y_i\right\|^2 \tag{11}$$

The root mean square error:

$$E_{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left\|\hat{Y}_i - Y_i\right\|^2} \tag{12}$$

The mean absolute error:

$$E_{MAE} = \frac{1}{M}\sum_{i=1}^{M}\left|\hat{Y}_i - Y_i\right| \tag{13}$$

where Y_i and $\hat{Y}_i$ are the estimated output and the actual output, respectively, and M is the last time step.

4. Experimental Results and Discussion

The script was executed in Windows 10 (x64), Intel(R) Core(TM) i5-8250U processor running at 1.80 GHz (four CPUs), with 8 GB of RAM, 256 Hard Disk type SSD.

This paper aims to search for new factors and intelligent models that will enhance the prediction of AAPL and TSLA. The effect of social media on the fluctuation of prices is paramount. So, the first attempt was to get tweets from Twitter using the keywords AAPL and TSLA. After collecting the tweets from Twitter for five months using API, we found that the MSE of LSTM with sentiment analysis related to AAPL was close to 1. Therefore, instead of being a helping factor, it represents a perturbation. After several simulations, we found that MSE in train data was 0.0182478 and MSE in test data was 0.0269575. The relevance of the tweets that we collected plays a massive impact on our prediction model. This finding makes us search for a new way to extract only the most relevant tweets.

Dealing with the false tweets and differentiating between what is real and fake is very important. The

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 04, December 2022.

tweets should be from a confidential source. Otherwise, it can affect the real price in a wrong way and lead to a bad decision. We found that the website "www.stockstwit.com" has a policy of getting only the relevant information. We intend to combine LSTM with AM and add the sentiment analysis for better forecasting. Figure 5 represents the general steps used to forecast the closing prices of TSLA and APPL.

4.1 Data

We collected all the tweets from the website *www.stockstwit.com* from *07/03/2019* to *30/04/2021*. The keywords were “AAPL” and “TSLA”. It took almost three days for the script to get the needed data. For the historical data related to the price of AAPL, we collected it from Yahoo! Finance *www.yahoo.com*. Our previous work [48] was performed on APPL data; we applied LSTM, ANN and GRU to predict the closing stock prices of APPL. This work is an extension of it. We decided to compare the LSTM, LSTM+SI and the hybrid model. But, having only one stock is not enough; thus, we added the stock TSLA, because it’s in the same stock market (Nasdaq) to see how the model would work.

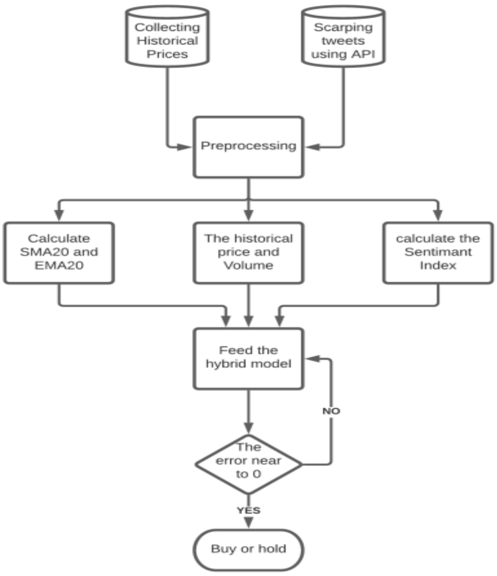

Figure 5. Flowchart of the forecasting steps.

The data needs a pre-processing step to remove all the noise. Table 1 and Table 2 show a sample of the technical indicators used. The data was divided into two semi-groups: 80% training data and 20% testing data.

Table 1. Sample of technical indicators of TSLA after pre-processing.

Date	Open	High	Low	Close	Adj Close	Volume
2019-04-18	54.24	54.96	53.95	54.65	54.65	29381500
..	..	..	..	..	..	..
2021-04-29	699.51	702.25	668.5	677	677	28845400

Table 2. Sample of technical indicators of AAPL after pre-processing.

Date	Open	High	Low	Close	Adj Close	Volume
2019-04-18	50.77	51.03	50.63	50.96	49.91	96783200
..	..	..	..	..	..	..
2021-04-29	136.47	137.07	132.44	133.47	133.47	151101000

First, we calculated the SMA20 and EMA20 based on [25], then we extracted the sentiment of tweets of each day (bullish or bearish). After that, we calculated the Sentiment Index (SI) based on the following equation which was taken from [24]:

$$SI = \ln\left(\frac{1+\alpha}{1+\beta}\right) \tag{14}$$

where α is the number of bullish tweets per day and β is the number of bearish tweets per day. The features are: opening price, highest price, lowest price, adjusted price, volume, SMA20, EMA20 and SI. All these features are different; so the normalization step was essential. We transformed the original data into numbers between 0 and 1 to improve the training. The implementation was developed using Python3. These hyper-parameters were chosen based on our previous work [48].

4.2 Results and Discussion

Three metrics (MSE, MAE and RMSE) were used to validate the accuracy of the models. The closer the value of the error to 0, the better the performance.

Table 3. The hyper-parameters of the models.

Hyper parameters	**Values**
Number of hidden layers	1
Number of nodes	20
Batch size	10
Epochs	100
The activation function	*tanh*
Optimizer	Adam
Learning rate	0.001

Table 4. Errors of the models related to TSLA stock data.

	LSTM	***LSTM with SI***	***LSTM with AM***	***Hybrid model***
E_{MSE} (train)	0.0027509	0.0026156	0.0003709	0.0007796
E_{MSE} (test)	0.0135314	0.0126432	0.0063939	**0.0058008**
E_{RMSE} (train)	0.0480684	0.0629436	0.0198917	0.0224225
E_{RMSE} (test)	0.1108812	0.1335756	0.0771000	**0.0765036**
E_{MAE} (train)	0.0449642	0.0465490	0.0154761	0.0183373
E_{MAE} (test)	0.1144399	0.11970902	0.07385018	**0.0693858**

For TSLA stock prices, the lowest error was of Mean Square Error (MSE) using the LSTM model. It was with a value of 0.0135314. For the model LSTM with SI, MSE was the lowest error too with 0.0126432. For the model LSTM with AM, the lowest error was 0.0063939. For the hybrid model, the lowest error was 0.0058008. We conclude that the hybrid model performs better than the other models (as shown in Table 4). In other words, the accuracy of the LSTM model is 98.64%, while the accuracy of the hybrid model is 99.42%. We notice the impact and the improvement of the model's performance using hybridization.

Table 5. The Error of the models related to APPL stock.

	LSTM	***LSTM with SI***	***LSTM with AM***	***Hybrid model***
E_{MSE} (train)	0.0031031	0.0025217	0.0005482	0.0007796
E_{MSE} (test)	0.0127536	0.0107397	0.0033089	**0.0032915**
E_{RMSE} (train)	0.0493602	0.0597210	0.0246819	0.0278273
E_{RMSE} (test)	0.1082771	0.1194630	0.0528034	**0.0548668**
E_{MAE} (train)	0.0538247	0.1077374	0.0199947	0.0216156
E_{MAE} (test)	0.1214208	0.1748030	0.0535530	**0.0506833**

For APPL stock prices (as shown in Table 5), the lowest Mean Square Error (MSE) was using the LSTM model. It was with a value of 0.0127536. For the model LSTM with SI, MSE was the lowest error with 0.0107397. The MSE for the model LSTM with AM was 0.0033089. For the hybrid model, it performs better than the other models, because the MSE was 0.0032915. We conclude that the hybrid model performs better for APPL Stock than other models.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 04, December 2022.

As shown in Figure 6 and Figure 7, it represents the prediction of the closing price of AAPL using the LSTM model and the hybrid model, respectively. The prediction of the closing price using the hybrid model between April 2019 and April 2020 proved to be close to the actual closing price. But, from April 2020 to April 2021, the gap between the two curves of the predicted price and the actual price of APPL extends. We can explain this due to other factors, such as politics and rumors. This fact leads us to believe that technical analysis alone is not enough; using fundamental analysis is also essential to eliminate the risk of losing money in long-term trading.

We noticed also that the SA didn't have a significant impact on the prediction, because the real problem is in the text itself, whether real or fake, a sarcasm or rumors, which makes the sentiment inaccurate. We get this conclusion based on our first attempts. After collecting the tweets from Twitter for five months using API, we found that the MSE of LSTM with sentiment analysis related to AAPL was close to 1. Therefore, instead of being a helping factor, it represents a perturbation. This finding leads us to conclude that SA has several limitations on the implementation side and needs more work in the future for obtaining better performance.

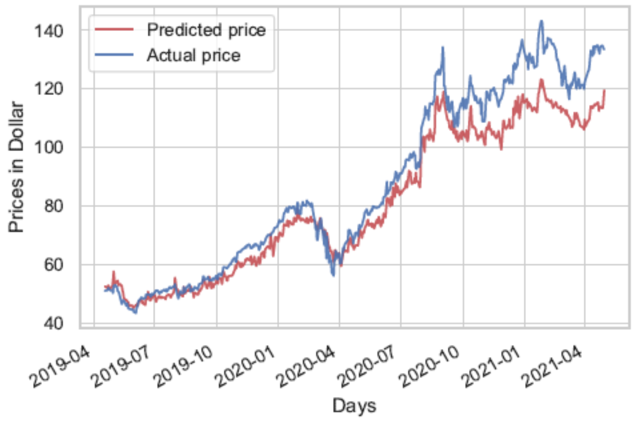

Figure 6. The prediction of AAPL using LSTM.

Figure 7. The prediction of AAPL using the hybrid model.

Figure 8 and Figure 9 represent TSLA and AAPL predictions of the closing price, respectively, using LSTM, LSTM with attention mechanism (LSTM+AM) and the hybrid model (LSTM+AM+SI). For both stocks, AAPL and TSLA, we can observe that between 2019 and 2020, the prediction was accurate compared with the period between 2020 and 2021. The movement of AAPL and TSLA closing prices in 2021 was chaotic. We can explain this perturbation by the beginning of the lockdown in 2021, which represents a very chaotic period in the whole world. People tended to spend more time on social media; thus, the impact of fake news and rumors is more powerful than in the period from April 2019 to April 2020.

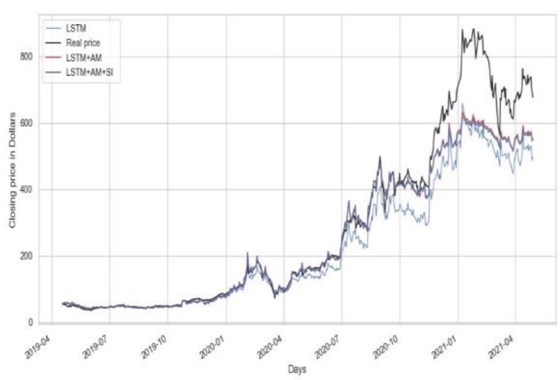

Figure 8. TSLA prediction using LSTM, LSTM+AM and the hybrid model.

Figure 9. APPL prediction using LSTM, LSTM with AM and the hybrid model.

However, if we compare the prediction of LSTM and LSTM+SI with that of the hybrid model, we conclude that the hybrid model surpasses the other models. Taking a decision based on the proposed model is beneficial, because it minimizes the risk of losing money while trading. The model is not perfect, but it will help traders make good decisions about buying or holding stocks.

5. CONCLUSION AND FUTURE WORKS

This paper is about using a hybrid model to forecast the stock prices of AAPL and TSLA. The chaotic movement of these stocks makes the prediction problem very challenging. We integrate LSTM, Attention Mechanism and Sentiment Analysis (the hybrid model) to solve this issue. The attention Mechanism with LSTM makes the prediction accuracy better than with the LSTM alone. In our case, the sentiment analysis with LSTM did not bring too much value to the prediction. Otherwise, the integration of LSTM with AM makes the prediction perform better than LSTM alone and LSTM with sentiment analysis. In conclusion, the hybrid model has higher accuracy than the other models. We also conclude that the news affects the results, but with minor improvement in this experiment. Therefore, this issue leads us to look for new methods to extract only the most relevant ones. Future work will be about integrating our proposed model into a simulator platform, such as Ninja trader or Meta trader to help traders predict the next move of the stock prices and make the right decision.

ACKNOWLEDGEMENTS

The authors would like to thank the editor and the three anonymous reviewers for their insightful comments, which helped improve the paper.

ETHICS AND IMPLICATIONS

The authors declare that they have no conflicts of interest to disclose. We are not responsible for anyone who uses these results in real-life trading without consulting professionals.

REFERENCES

[1] M. M. Rounaghi and F. N. Zadeh, "Investigation of Market Efficiency and Financial Stability between S&P 500 and London Stock Exchange: Monthly and Yearly Forecasting of Time Series Stock Returns Using ARMA Model," Physica A: Statistical Mechanics and its Applications, vol. 456, pp. 10–21, 2016.

[2] A. A. Ariyo, A. O. Adewumi and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," Proc. of the 16th IEEE International Conference on Computer Modeling and Simulation (UKSim-AMSS), pp. 106–112, Cambridge, UK 2014.

[3] P. H. Franses and D. V. Dijk, "Forecasting Stock Market Volatility Using (Non-linear) Garch Models," Journal of Forecasting, vol. 15, no. 3, pp. 229–235, 1996.

[4] R. A. K. Cox and G. W.-Y. Wang, "Predicting the US Bank Failure: A Discriminant Analysis," Economic Analysis and Policy, vol. 44, no. 2, pp. 202–211, 2014.

[5] W. Huang, Y. Nakamori and S.-Y. Wang, "Forecasting Stock Market Movement Direction with Support Vector Machine," Computers & Operations Research, vol. 32, no. 10, pp. 2513–2522, 2005.

[6] A. V. Devadoss and T. A. A. Ligori, "Forecasting of Stock Prices Using Multi Layer Perceptron," International Journal of Computing Algorithm, vol. 2, pp. 440–449, 2013.

[7] N. Singhal, V. Ganganwar, M. Yadav, A. Chauhan, M. Jakhar and K. Sharma, "Comparative Study of Machine Learning and Deep Learning Algorithm for Face Recognition," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 7, no. 3, pp. 313–325, 2021.

[8] F. Assiri and M. Alrehaili, "Development of Ensemble Machine Learning Model to Improve COVID-19 Outbreak Forecasting," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 08, no. 2, pp. 48 – 58, DOI: 10.5455/jjcit.71-1640174252, 2022.

[9] M. Hiransha, E. Ab Gopalakrishnan, V. K. Menon and K. P. Soman, "NSE Stock Market Prediction Using Deep-learning Models," Procedia Computer Science, vol. 132, pp. 1351–1362, 2018.

[10] A. Graves, N. Jaitly and A.-R. Mohamed, "Hybrid Speech Recognition with Deep Bidirectional LSTM," Proc. of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 273–278, Olomouc, Czech Republic, 2013.

[11] C. Zhou, C. Sun, Z. Liu and F. Lau, "A C-LSTM Neural Network for Text Classification," arXiv preprint, arXiv: 1511.08630, 2015.

[12] X. Pang, Y. Zhou, P. Wang, W. Lin and V. Chang, "An Innovative Neural Network Approach for Stock Market Prediction," The Journal of Supercomputing, vol. 76, no. 3, pp. 2098–2118, 2020.

[13] J. Huang, J. Chai and S. Cho, "Deep Learning in Finance and Banking: A Literature Review and Classification," Frontiers of Business Research in China, vol. 14, pp. 1–24, 2020.
[14] D. Shah, H. Isah and F. Zulkernine, "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques," International Journal of Financial Studies, vol. 7, no. 2, DOI: 10.3390/ijfs7020026, 2019.
[15] J. Chung, C. Gulcehre, K. H. Cho and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," arXiv preprint, arXiv: 1412.3555, 2014.
[16] A. Vidal and W. Kristjanpoller, "Gold Volatility Prediction Using a CNN-LSTM Approach," Expert Systems with Applications, vol. 157, DOI: 10.1016/j.eswa.2020.113481, 2020.
[17] Y. Bengio, "Deep Learning of Representations for Unsupervised and Transfer Learning," Proc. of ICML Workshop on Unsupervised and Transfer Learning, vol. 27, pp. 17–36, 2012.
[18] N. Bahadur and R. Paffenroth, "Dimension Estimation Using Autoencoders," arXiv preprint, arXiv: 1909.10702, 2019.
[19] A. Essien and C. Giannetti, "A Deep Learning Framework for Univariate Time Series Prediction Using Convolutional LSTM Stacked Autoencoders," Proc. of the 2019 IEEE Int. Symposium on Innovations in Intelligent Systems and Applications (INISTA), Sofia, Bulgaria, pp. 1–6, 2019.
[20] W. Bao, J. Yue and Y. Rao, "A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and Long-short Term Memory," PLOS One, vol. 12, no. 7, p. e0180944, 2017.
[21] G. H. Merabet, M. Essaaidi, M. Ben Haddou et al., "Intelligent Building Control Systems for Thermal Comfort and Energy Efficiency: A Systematic Review of Artificial Intelligence-assisted Techniques," Renewable and Sustainable Energy Reviews, vol. 144, DOI: 10.1016/j.rser.2021.110969, 2021.
[22] V. S. Pagolu, K. N. Reddy, G. Panda and B. Majhi, "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements," Proc. of the 2016 IEEE Int. Conf. on Signal Processing, Communication, Power and Embedded System (SCOPES), pp. 1345–1350, Paralakhemundi, India, 2016.
[23] J. Smailović, M. Grčar, N. Lavrač and M. Žnidaršič, "Predictive Sentiment Analysis of Tweets: A Stock Market Application," Proc. of the Int. Workshop on Human-computer Interaction and Knowledge Discovery in Complex Unstructured, Big Data (HCI-KDD 2013), vol. 7947 pp. 77–88, 2013.
[24] Z. Jin, Y. Yang and Y. Liu, "Stock Closing Price Prediction Based on Sentiment Analysis and LSTM," Neural Computing and Applications, vol. 32, pp. 9713–9729, 2020.
[25] Z. Berradi and M. Lazaar, "Integration of Principal Component Analysis and Recurrent Neural Network to Forecast the Stock Price of Casablanca Stock Exchange," Procedia Computer Science, vol. 148, pp. 55–61, 2019.
[26] M. Ismail and A. M. Awajan, "A New Hybrid Approach EMD-EXP for Short-term Forecasting of Daily Stock Market Time Series Data," Electronic Journal of Applied Statistical Analysis, vol. 10, no. 2, pp. 307–327, 2017.
[27] J. Qiu, B. Wang and C. Zhou, "Forecasting Stock Prices with Long-short Term Memory Neural Network Based on Attention Mechanism," PLOS One, vol. 15, no. 1, p. e0227222, 2020.
[28] N. Jing, Z. Wu and H. Wang, "A Hybrid Model Integrating Deep Learning with Investor Sentiment Analysis for Stock Price Prediction," Expert Systems with Applications, vol. 178, p. 115019, 2021.
[29] R. Batra and S. M. Daudpota, "Integrating Stock Twits with Sentiment Analysis for Better Prediction of Stock Price Movement," Proc. of the 2018 IEEE Int. Conf. on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1–5, Sukkur, Pakistan, 2018.
[30] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia and D. C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis," Proc. of the 2019 IEEE 5th Int. Conf. on Big Data Computing Service and Applications (BigDataService), pp. 205–208, Newark, CA, USA, 2019.
[31] D. K. Kirange, R. R. Deshmukh et al., "Sentiment Analysis of News Headlines for Stock Price Prediction," Composoft: An Int. J. of Advanced Computer Technol., vol. 5, no. 3, pp. 2080–2084, 2016.
[32] J. Abraham, D. Higdon, J. Nelson and J. Ibarra, "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis," SMU Data Science Review, vol. 1, no. 3, p.1, 2018.
[33] S. Liu, C. Zhang and J. Ma, "CNN-LSTM Neural Network Model for Quantitative Strategy Analysis in Stock Markets," Proc. of the Int. Conf. on Neural Information Processing (ICONIP 2017), vol. 10635, pp. 198–206, 2017.
[34] M. R. Vargas, B. De Lima and A. G. Evsukoff, "Deep Learning for Stock Market Prediction from Financial News Articles," Proc. of the 2017 IEEE Int. Conf. on Computational Intelligence and Virtual Environments for Measurement. Sys. and Appl. (CIVEMSA), pp. 60–65, Annecy, France, 2017.
[35] C.-Y. Lee and V.-W. Soo, "Predict Stock Price with Financial News Based on Recurrent Convolutional Neural Networks," Proc. of the 2017 IEEE Conf. on Technologies and Applications of Artificial Intelligence (TAAI), pp. 160–165, Taipei, Taiwan, 2017.
[36] J. Li, H. Bu and J. Wu, "Sentiment-aware Stock Market Prediction: A Deep Learning Method," Proc. of the 2017 IEEE Int. Conf. on Service Systems and Service Management, pp. 1–6, Dalian, 2017.
[37] H. Yan and H. Ouyang, "Financial Time Series Prediction Based on Deep Learning," Wireless Personal Communications, vol. 102, no. 2, pp. 683–700, 2018.
[38] Y. Chen, J. Wu and H. Bu, "Stock Market Embedding and Prediction: A Deep Learning Method," Proc.

of the IEEE 2018 15th Int. Conf. on Service Systems and Service Management (ICSSSM), pp. 1–6, Hangzhou, China, 2018.

[39] H. Y. Kim and C. H. Won, "Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple Garch-type Models," Expert Systems with Applications, vol. 103, pp. 25–37, 2018.

[40] F. Shen, J. Chao and J. Zhao, "Forecasting Exchange Rate Using Deep Belief Networks and Conjugate Gradient Method," Neurocomputing, vol. 167, pp. 243–253, 2015.

[41] S. Hochreiter and J. Schmidhuber, "LSTM Can Solve Hard Long Time Lag Problems," Proc. of the 9th International Conference on Neural Information Processing Systems (NIPS'96), pp. 473–479, 1997.

[42] J. Pennington, R. Socher and C. D. Manning, "GloVe: Global Vectors for Word Representation," Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543, Doha, Qatar, 2014.

[43] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation," arXiv preprint, arXiv: 1406.1078, 2014.

[44] D. Bahdanau, K. Cho and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," arXiv preprint, arXiv: 1409.0473, 2014.

[45] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho and Y. Bengio, "Attention-based Models for Speech Recognition," arXiv preprint, arXiv: 1506.07503, 2015.

[46] Z. Berradi, M. Lazaar, O. Mahboub and H. Omara, "A Comprehensive Review of Artificial Intelligence Techniques in Financial Market," Proc. of the 2020 6th IEEE Congress on Information Science and Technology (CiSt), pp. 367–371, DOI: 10.1109/CiSt49399.2021.9357175, 2020.

[47] T. Hollis, A. Viscardi and S. Eun Yi, "A Comparison of LSTMS and Attention Mechanisms for Forecasting Financial Time Series," arXiv preprint, arXiv: 1812.07699, 2018.

[48] Z. Berradi, M. Lazaar, H. Omara and O. Mahboub. "Effect of Architecture in Recurrent Neural Network Applied on the Prediction of Stock Price," IAENG International Journal of Computer Science, vol. 47, no. 3, pp. 436–441, 2020.

ملخص البحث:

يُعدّ التّعلُّم العميق ميداناً واعداً من ميادين البحث والتّطوير، وله تطبيقات عديدة في شتّى مجالات الحياة. والجدير بالذّكر أنّ تطبيقاته في أسواق الأسهم منتشرة على نطاق واسع نظراً لما يتمتّع به من فاعلية. وقد أثبتت تقنية ذاكرة المدى القصير والطويل (LSTM) فاعلية في التّعامل مع بيانات السّلاسل الزّمنية نظراً لبنيتها الفريدة ذات الوحدات المخفية.

وتجمع هذه الورقة بين تقنية ذاكرة المدى القصير والطويل (LSTM) المذكورة وبين كل من آلية النتباه (AM) وتحليل المشاعر (SA) للتنبّؤ بسعر الإغلاق في كلّ من بورصة AAPL وبورصة TSLA التّابعتين لسوق ناسداك المالي (NASDAQ). وقد تمت مقارنة النّموذج الهجين المقترح مع كلّ من نموذج (LSTM)، ونموذج (LSTM) مع تحليل المشاعر (SA)، ونموذج (LSTM) مع آلية الانتباه (AM). وقد استُعملت ثلاث علامات مرجعية لقياس أداء كلّ من تلك النّماذج هي: MSE، و RMSE، و MAE. وبينت النتائج أنّ النّظام الهجين تفوّق في الأداء على غيره من النّماذج.