JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

ENHANCED DEEP LEARNING FRAMEWORK TO PREDICT VARIATONS IN STOCK MARKETS

Jasmine Sabeena[1], Pole Venkata Subba Reddy[2]

[1]Research Scholar, Dept. of CSE, S.V.U College of Engineering., S.V. University, Tirupati, India,
[2]Professor, Dept. of CSE, S.V.U College of Engineering., S.V. University, Tirupati, India,
E-mail: jasmine539@gmail.com[1], pvsreddy@hotmail.co.in[2]

Received: 14 March 2020 Revised and Accepted: 8 July 2020

ABSTRACT: This research study aims to develop an innovative approach to the neural network Realize accurate forecasts in financial prices. For in-house research and findings of visualisations and simulations, data from the live stock exchange is collected. Prove Digital Internet of Items to be evaluated in inventory. Traditional neural network algorithms analyze the effect of consumer characteristics on equity prices The stock market may be falsely forecast, since the original weight of the selection dilemma will simply be estimated incorrectly. On the basis of the. It is not a static index or price index, but high dimensional multi-market index. The intense simulation study demonstrates the performance of the proposed mechanism.

KEYWORDS: Stock market prediction, Sentiment analysis, Deep learning, Neural networks

I. INTRODUCTION

In recent years, the main technique for prediction of stock price movement is through machine learning or deep learning ways. In contempt of many studies that have presented for stock market, Financial market analysis is divided into two groups i) fundamental analysis ii) technical analysis. .in technical analysis prediction of data is done based on analysing the historical data. On the other way, fundamental analysis is done through balance sheets, cash flow statements and determining non-linear characteristics is a difficult task because of noise values[15]. Machine learning algorithms seems to be more promising approach for stock feature extraction problem, Convolution neural networks (CNN) have an advantage of predicting the features. This frame work is build based on CNN because of its proven capabilities that were mentioned successfully reported in stock prediction domain. Convolution neural network is implemented on stock market data with pre-trained model and applied 1 D- CNN with having convolution[4]. Parameters of stock market price like volume, open, close, adjacent price, high, low are considered for inputs In order to identify the stock market data online, a textual – based sentiment analysis is adopted technique called BERT model [14].this is a frame work based approach explained a better performance of BERT mode-based sentiment index for financial analysis. Finally we accomplish the stock prediction by powerful neural-network model CNN, which out performs by time-series analysis model by ARIMA model. ARIMA model provides investors to make a decision process by financial series analysis [13]. Many factors are taken in consideration for stock price predictions and relationship between the variables over a period of time that predict the trends in stock price. Technical analysis method used for future stock price prediction by providing the useful information to the traders .traders uses technical analysis tools for identifying the entry and exit point which reflects the price chart by applying the historical price of stocks like closing price ,opening price, volume traded, adjacent close values. Stock market price is dynamic and non-linear with the huge data set of variety of data. Machine learning techniques are proved to improve and compare to the past method.ANN are self learning process for identifying the hidden features. The correlation that exists among stock markets is another possible analysis of related variables. The contribution of the paper as follows: 1. Gathering stock price data and finding canonical correlation analysis is used among variables and it is aggregated with sentiment analysis BERT processing for exact extraction of stock related data. 2. After data is collected in order to predict the features with in a time period, we have choosen the ARIMA model.

The obtained features values is given to CNN 1-dimensional for analyzing diverse set of features which are provided as input are trained in such a way to predict stock market price. The two proposed algorithm collected the data and find out the canonical correlation among variables after that eliminates the negative values later applied ARIMA model to positive correlated values and generates the time series data. The CNN algorithm obtain the input from generated time series for prediction of stock price.CNN contains 7 layers: input layer, conv1d, averagePooling1d, conv1d ,flatten layer and dense layer. Stock's historical trading data and technical indicators an advanced NLP technique called BERT(Bidirectional embedding Representations from Transforms) is used for sentiment analysis for the source of fundamental analysis, Fourier transforms are used for fetching directions of overall trend, performing eginvalue decomposition for finding correlated assets. auto -

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

regression integrated moving average(ARIMA) to generate time series prediction for the stock function approximation and so on, to gather much information , patterns, dependencies, etc, from the stock data. CNN is used for the purpose of creating neural nets. Before applying the algorithm consider stock market are perfect without random.

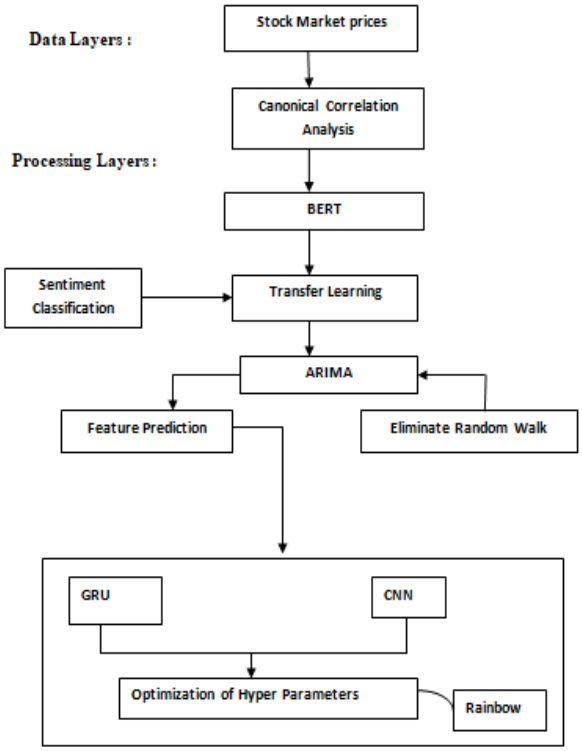
II. LITERATURE REVIEW

The goal of this segment is to address the analysis of stock market future predictions. Many research studies had mentioned the stock predictions using machine learning based but this study is lack of accuracy because of inconsistency of dataset [6]. In [7] it is shown the various deep learning techniques that has been evaluated on stock market data along with the encoder model (BERT model) is specified but it would be more efficient if sentiment analysis is used. Author saikat mandal[8] has proposed frame work for deep learning techniques like LSTM for stock prediction .in his proposed model author tries to explain about the combinational of machine learning and deep learning model gives the better results but this technique is more accurate if more deeper layers is added. In the context of optimize of network topology of CNN the author hyejung chung in his study[9] suggested the deep learning techniques to accomplish the optimization technique to improve the performance of CNN as a part of Feature extraction. It contains the fixed set of parameters with respect to the previous data for computing stock resource data the limitation of this model has fixed set of features and it can be improved by fully connected layers. In the analysis of ARIMA[10] an hybrid model of Artificial neural network is used the author Yakubu Musa[10] has focus on ANN with ARIMA model with aiming to capture various time series which enhances the forecasting over ARIMA model and ANN model , in this hybrid model author demonstrated with better results for financial time series forecasting .this model can be improves by adding optimization techniques.

With the implementation results the authors has shown the BERT[11] model stands as language interface for NLP with bidirectional encoding .This influences the pre-training of deep bi directional representations with state-of-art model where wide range of tasks such as language interface and question answering. This model can ne improved by hybridization of CNN model for Stock Predictions. wensi tang[12] has analysed theoretically about the performance of 1D-CNN be describing the importance of kernel size with a proposal of novel omni-scale 1D-CNN.this method can be improved with OS-CNN model with many hyper-parameters. Stock closing price prediction[16] author uses Artificial Neural Network and random forest techniques for predicting stock prices for next day along with closing price by considering financial data using strategic indicators RMSE and MAPE. This paper is about the comparative analysis of ANN algorithms.

III. PROPOSED MODEL

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020



A. Collection of data through online news database:

The objective of this section is to gather the data from real time news sources. The controlling parameters of stock market predictions are high price, lowest price, and closing price, opening price, adjusted closing price and volume. This factor`s of data set is downloaded from yahoo finance .Parsing the extracted dataset from online database will be suggested for required data upon which sentiment analysis could be implemented. Data is derived from online news sources database which shows the relevant classified output

B.BERT:

Bidirectional Encoders Representations for Transformers is used for capturing the stack market data from the document. This model is used for semantic information of stock prediction.[1].stock market prediction is done with parameters like low, high. Bert collects the parameters from the stock market sentence and generates the next sentence with different classes of embeddings with various time periods, then transforms the created embeddings of paired word which are close occurrence to next sentence generator[1]. It is a Bidirectional context transform with pre-trained data. next sentence prediction is accomplished through a function by taking inputs as two sentences and decide through logically which follows the first sentence for illusion consider the below example:

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

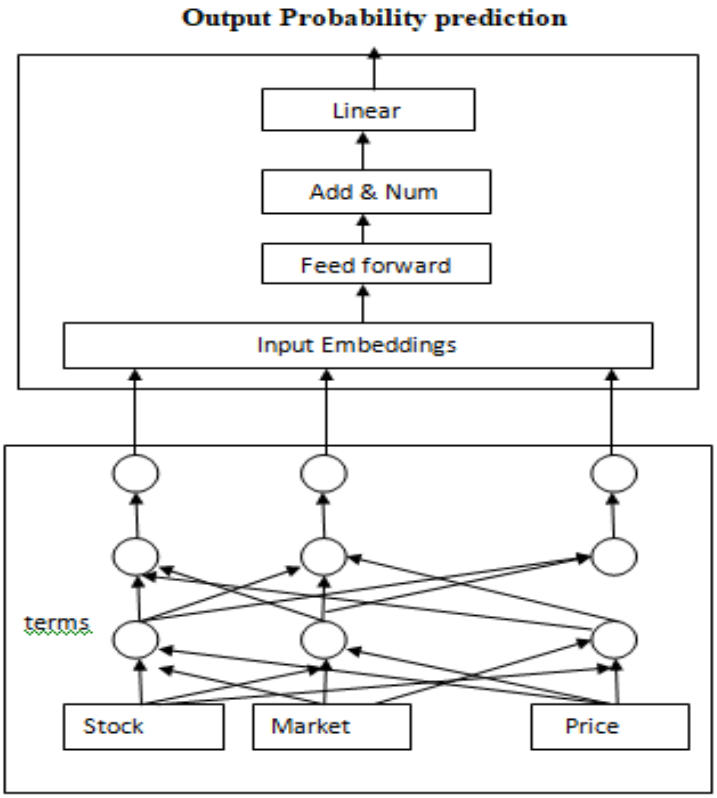


Figure 2. BERT

Search terms are collected and this terms are bidirectionally compared with next sentence by undergoing through different layers of embedding with fully connected layers of encoders and decoders.[2]

C. Canonical Correlation analysis:

The Technique of Canonical correlation analysis is to generate the linear relations between various set of variables which is the best way of expressing the correlation in such a way to predict the highest related variables. The main objective of canonical correlation analysis is data reduction which explain co-variation between two sets of variables. Let us consider a set variables p_1, p_2, p_3,..p_n as 'P' and the other set of variables q_1,q_2,q_3,….q_n as 'Q' and the correlation among these variables are as follows:

function CoM[u][v] with order of (A+B).

1.$CoM_{u,v} = cov(u,v)/t_{uv}$ --- (1)

2.$Cov(u,v) = \sum_{i=1}^{k}[(u_i - u) * (v_i - v)]/k$ --- (2)

3. $t_{uv} = \sum_{i=0}^{k}[(u_i - u) * (v_i - v)]^2$ ---- (3)

Where u and v taken as mean values obtained from the variable set 'A' and 'B' respectively. The correlated set of variable are inputed for semantic analysis.

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

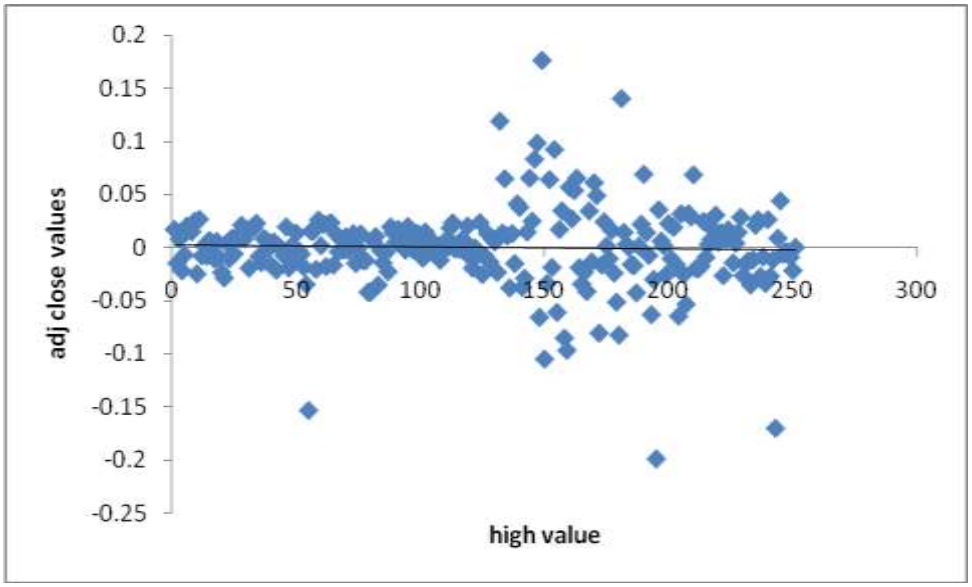


Figure 3:Corelation values

D.ARIMA MODEL:

Auto Regression Integrated Moving Average(ARIMA)

It is a time series prediction model which describes the connection among pre-defined number of Lagged observations and current values. ARIMA stands for Auto-Regressive Integrated Moving ".the term "autoregressive" is referred to lags of stationeries series, "moving average" terms represented as lags of forecast errors, "integrated " version of stationary series are differenced time series.

The ARIMA model uses :

- Number of lag observations values
- The degree of the differencing.
- Size of moving average window.

Model selects the parameters and identified the appropriate automatic algorithm for further comparisons of diagnosis testing of the selected model. Finally the forecasting is done through selected model.

The Generalized AR(p,q)model:

$adp_t = a+\sum_{i=0}^{p} b_i adp_{t-i} + d_0 u_t + \sum_{j=1}^{q} d_j u_{t-j}$ -------(4)

here adp in time t is the previous value(in time t-1) and 'a' is white noise error term u. model contains p lags of the dependent variables and q lags of the error terms.

Algorithm predictstockprice(data_set)
Inputs: data_set is stock data which contains columns of close price, low price, high price, adjusted price.
Output: obtain time series data stationary with canonical correlation and ARIMA Model which is used to model prediction.

1. collect the stock data set.
2. //apply canonical correlation analysis to measure how stock price move in relation to each other.
3. for stock_data in two years
4. U=log ($\frac{adjclose}{adjclose+1}$)
5. V= log (clode_price – open_price)
6. $t_{uv} = \sum_{i=0}^{k}[(u_i - u) * (v_i - v)]^2$
7. Cov(u,v) = $\sum_{i=1}^{k}[(u_i - u) * (v_i - v)]/k$
8. $CoM_{u,v}$ = cov(u,v)/t_{uv}
9. r_i= $CoM_{u,v}$ assign correlation matrix to r.

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

10. if(r_i >0.5) then
11. r_i.label=positive_correlation
12. else
13. r_i.label=negative_correlation
14. remove the negative correlated data.
15. Apply the **ARIMA model** for obtained positive correlated data.
16. To check whether the result is stationary or not .calculate Dickey fuller test
17. $P_t = \alpha + (\beta-1)P_{t-1}+e_t$ // where P is variable of interest, t is the time index, α, β are coefficients and e_t is error term.
18. While(i==TIME_Range) then
19. {
20. If ($r_t <= P_t$) then
21. Write "The strong evidence against null hypothesis and reject the null hypothesis where Data has no unit root is stationary".
22. Stock_price=r_i
23. else
24. Write "The weak evidence against null hypothesis and time series has a unit root by indicating it is non-stationary".
25. //Reject the data
26. }
27. //apply Auto Regression Model for obtained stationary data
28. $x_t^1 = \mu+Ø_1 x_{t-1}+\ldots+Ø_u x_{t-u} –Ø_1 e_{t-1}-\ldots.-Ø_v e_{t-v}$ // where e_{t-1} =xt-1 $–x_{t-1}^1$ and Ø is moving average parameters.

E. Convolution Neural Network:

Convolution neural network (CNN) has higher advantage for classification as well as prediction. one dimensional with shallow architecture computation complexity is less than two dimensional [4.The design of 1D – CNN accomplished as follows:

- Configure the number of neurons and hidden layers.
- Define filter size and make convolution
- Mention the sub sampling factors in each CNN layers.
- Choose the pooling and activation functions.

The convolution operation is calculating the changes made by the filters are assigned with weights. this weights are modified and updated based on training .if the training data is over fitting then the process of pooling layer in CNN then it lower the level of over fitting. The final layer of CNN is fully connected layer in which converts the extracted features into final output. The CNN as follows the vector input data i.e historical stock market data where the convolution is classified as 1D CNN with linear function. The model is trained with vectors of stock market data to predict the future value.[5] consider the vector input data as

$F=(f^1,f^2,f^3\ldots f^m)$ (5)

Here m is taken as sample inputs. f is mentioned as vector matrix with multiple features if there are 8 features then feature matrix formatted as Lx8 output sequence is as follows:

$G=(g^1,g^2,g^3\ldots g^m)$ (6)

Here G is Boolean value which represents 0 or 1. To enhance convergence speed and to recognize the optimal sequence of variable gradient descent technique has been implemented. The feature of input data should be normalized if there is large variance by using min-max normalization

$$H^* = \frac{h-\min(h)}{\max(h)-\min(h)} \quad (7)$$

Stock prediction model is linear where convid function is poor to solve nonlinear problem, so in this context activation functionuses Relu and Lrelu function in order to solve gradient vanishing.

$$\text{Relu}(h)=\begin{cases}0 & (h \le 0)\\ h & (h > 0)\end{cases} \quad (8)$$

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

$$\text{Lrelu(h)} = \begin{cases} a_i h_i \ (h \leq 0) \\ h_i (h > 0) \end{cases} \qquad (9)$$

In the above equation h considered as outputs obtained from layer a_i of convolution layer that measures the hyper parameters of threshold value of layer (0.01 as initial layer value). The process of Relu[11] and Lelu is used for next convolution layer as features which are considered as input to the pooling layers where classification is obtained .

$$J_i = \frac{p^{c_i}}{\sum_{k=1}^{n} p^{c_k}} \quad i=2 \qquad (10)$$

Where i represents the count of classification 'c'.n is number of c and P is natural algorithm. Error function as follows:

$$Q(S^l) = -\sum_{k=1}^{n} \sum_{i=1}^{g} l_{ki} \log (e_i) \qquad (11)$$

S^i value is gradient descent . N is learning rate set is 0.001.

To analyzing stock price prediction five parameters are calculated: on balance estimate(OBE) ,stock momentum oscillator(SMO) ,Relative strength Index(RSI),stochastic(%S) and moving average value(MV).TV: Today's closing price value. YV: Yesterday's closing price, volume: today's Volume.

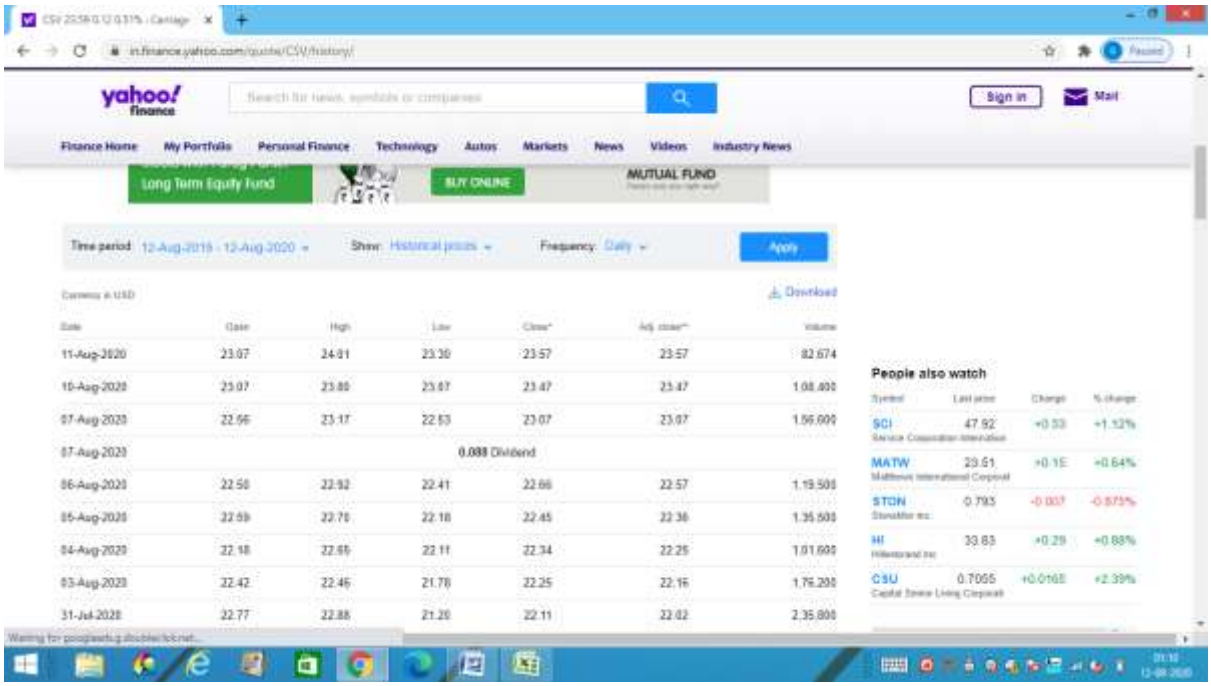
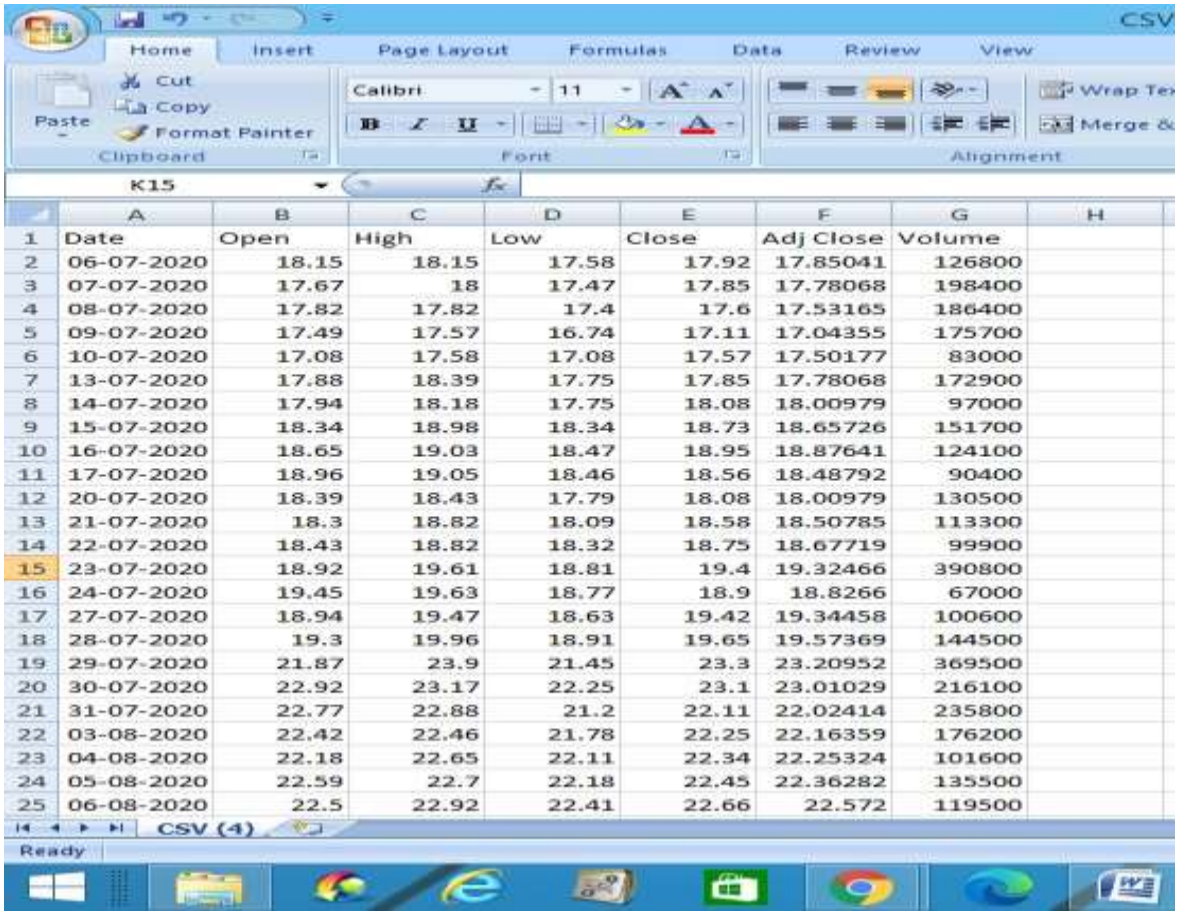


Figure 4: Data collection

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020



	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	06-07-2020	18.15	18.15	17.58	17.92	17.85041	126800
3	07-07-2020	17.67	18	17.47	17.85	17.78068	198400
4	08-07-2020	17.82	17.82	17.4	17.6	17.53165	186400
5	09-07-2020	17.49	17.57	16.74	17.11	17.04355	175700
6	10-07-2020	17.08	17.58	17.08	17.57	17.50177	83000
7	13-07-2020	17.88	18.39	17.75	17.85	17.78068	172900
8	14-07-2020	17.94	18.18	17.75	18.08	18.00979	97000
9	15-07-2020	18.34	18.98	18.34	18.73	18.65726	151700
10	16-07-2020	18.65	19.03	18.47	18.95	18.87641	124100
11	17-07-2020	18.96	19.05	18.46	18.56	18.48792	90400
12	20-07-2020	18.39	18.43	17.79	18.08	18.00979	130500
13	21-07-2020	18.3	18.82	18.09	18.58	18.50785	113300
14	22-07-2020	18.43	18.82	18.32	18.75	18.67719	99900
15	23-07-2020	18.92	19.61	18.81	19.4	19.32466	390800
16	24-07-2020	19.45	19.63	18.77	18.9	18.8266	67000
17	27-07-2020	18.94	19.47	18.63	19.42	19.34458	100600
18	28-07-2020	19.3	19.96	18.91	19.65	19.57369	144500
19	29-07-2020	21.87	23.9	21.45	23.3	23.20952	369500
20	30-07-2020	22.92	23.17	22.25	23.1	23.01029	216100
21	31-07-2020	22.77	22.88	21.2	22.11	22.02414	235800
22	03-08-2020	22.42	22.46	21.78	22.25	22.16359	176200
23	04-08-2020	22.18	22.65	22.11	22.34	22.25324	101600
24	05-08-2020	22.59	22.7	22.18	22.45	22.36282	135500
25	06-08-2020	22.5	22.92	22.41	22.66	22.572	119500

Figure 5: Stock price data in CSV file

Stock price prediction.

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

IV. CONCLUSION

The proposed neural network with a built-in layer and the long-term in this article. The proposed semantic memory network with automated encoder Neuronal network. Neural network. First, we review the efficiency of the Shanghai A-share models Sinopec and composite score. Then, for others, we use an updated Deep learning model Stock. The total precision of the 3 stocks is 58.2 percent and the composition of the A-share Index is 44% greater than the stochastic estimate. In brief, the approaches better predictive output for the Shanghai A-share listed in this article Index of composite. We also address the importance of our job for financial reporting. For financial research, this is a breakthrough study since we have incorporated multiple The cross-disciplinary methodology involves methods and algorithms.

V. REFERENCES:

[1] "Leveraging BERT to Improve the FEARS Index for Stock Forecasting",Linyi Yang , Yang Xu , Tin Lok James Ng , Ruihai Dong,2019.

[2] "BERT:Pre-trainingofDeepBidirectionalTransformersfor LanguageUnderstanding", JacobDevlin Ming-WeiChang KentonLee KristinaToutanova Google AI Language,2019.

[3] Rodrigo Loureiro Malacarne, Canonical Correlation Analysis, The Mathematica Journal 2014.

[4] "Research on Stock Price Prediction Method Based on Convolutional Neural Network ",Sayavong Lounnapha ,Wu Zhongdong,Chalita Sookasame, 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS).

[5] "A Modified Deep Learning Enthused Adversarial Network Model to Predict Financial Fluctuations in Stock Marke"t ,Jasmine Sabeena, P.Venkata Subba Reddy,2019.

[6] 5R.K. Dase , D. D. Pawar and D.S. Dasput,"Methodologies for Prediction of Stock Market: An Artificial Neural Network", E-ISSN:2249-8605, Volume 1, issue 1, 2011 pp 08-15.

[7] "Financial Sentiment Analysis for Predicting Direction of Stocks using Bidirectional Encoder Representations from Transformers (BERT) and Deep Learning Models", Derya Othan, Zeynep Hilal Kilimci, Mitat Uysal,Dec-2019.

[8] "Application of Deep Learning Techniques for Precise Stock Market Prediction", Saikat Mondal,feb-2020.

[9] "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction", Hyejung Chung1 ,Kyung-shik Shin1, Published 2019,Computer Science, Mathematics, Neural Computing and Applications.

[10] "Analysis of ARIMA-Artificial Neural Network Hybrid Model in Forecasting of Stock Market Returns" ,Yakubu Musa1* and Stephen Joshua1 , Asian Journal of Probability and Statistics 6(2): 42-53, 2020; Article no.AJPAS.53807 ISSN: 2582-0230.

[11] "BERT:Pre-trainingofDeepBidirectionalTransformersfor LanguageUnderstanding",

[12] JacobDevlin Ming-WeiChang KentonLee KristinaToutanova Google AI Language,may-2019.

[13] "Rethinking1D-CNNforTimeSeriesClassification: ",AStrongerBaseline,WensiTang , GuodongLong, LuLiu, TianyiZhou , JingJiang , MichaelBlumenstein, **2020-02-24** , *DOI: arxiv-2002.10061*.

[14] "Stock Price Prediction Using the ARIMA Model", Ayodele A. Adebiyi., Aderemi O. Adewumi ,march-2014.

[15] "BERT-based Financial Sentiment Index and LSTM-based Stock Return Predictability ", Joshua Zoen Git Hiew,2019.

[16] " Stock Prediction Using Convolutional Neural Network ",Sheng Chen and Hongxiang He 2018 IOP Conf. Ser.: Mater. Sci. Eng. 435 012026.

[17] " Stock Closing Price Prediction using Machine Learning Techniques", Mehar Vijh, , Deeksha Chandola, Vinay Anand Tikkiwal, Arun Kumar. ICCIDS,2019.