Stock Price Prediction

1. Ashish Kumar
21BCS11125
Computer science & Engineering
Apex Institute of Technology,
Chandigarh University
Mohali, Punjab, India.

2. Ayush Sharma
21BCS4010
Computer science & Engineering
Apex Institute of Technology,
Chandigarh University
Mohali, Punjab, India.

3. Yuvika
21BCS3690
Computer science & Engineering
Apex Institute of Technology,
Chandigarh University
Mohali, Punjab, India.

***Abstract*—At present, there are some problems in domestic stock market, such as difficulty in extracting effective features and inaccuracy in stock price forecast. This paper proposes a stock price prediction model based on Principal Component Analysis (PCA) and Long Short-Term Memory (LSTM). Firstly, PCA is used to extract the principal components of the technical indicators affecting stock prices, so as to reduce the data correlation and realize data dimensional reduction. Then, LSTM is used to model and predict the stock price. According to the experimental results of Pingan Bank, compared with the traditional stock price prediction models, the stock price prediction model based on PCA and LSTM can accurately predict the stock price fluctuation trend.**

Keywords: LSTM; Stock price prediction; Machine learning.

I. INTRODUCTION

Stock price time series prediction is the prediction behaviour of stock price fluctuation according to the historical data of stock price. Stock market is an important part of national econimic development. Forecasting stock price movements is important for governments, investors and investment institutions. Therefore, it attracts many scholars to conduct research. However, the price trend of the stock market may be influenced by political factors, macroeconomic factors, legal factors and etc., resulting in great uncertainty and volatility of the stock price, making it a major problem in research. This paper proposes a deep learning model based on PCA-LSTM to predict stock price fluctuation. Firstly, the model uses PCA to extract the main components from a number of technical indicators that affect stock prices, to achieve data dimensionality reduction, reduce network training time, and improve model performance. Secondly, the stock trading information is based on time series, and LSTM has the potential to learn long observation sequences. So, this paper uses LSTM to predict stock price time series. This paper uses Pingan Bank stock trading information as a data set from January 4, 2000 to December 28, 2021, using Convolution Neural Network (CNN) model, Multi-Layer Perceptron (MLP) model and Moving Average model as comparative experiments. The experimental results show that the combination of PCA method and LSTM model is better than the comparison models in prediction performance.

II. RELATED WORKS

Stock price time series prediction is the prediction behaviour of stock price fluctuation according to the historical data of stock price. The stock market is of great significance in the financial field. Therefore, the research on stock price prediction has attracted the attention of many researchers at home and abroad. Some researchers use the moving average analysis method to study the stock market price trend. For example, Aistis Raudys[1] proposed an optimal stock price smoothing weighting scheme based on the negative weight moving average. However, the moving average has a certain lag. In recent years, with the extensive application of deep learning technology, many domestic and foreign researchers began to use deep learning to conduct stock prediction research. For example, Volodymyr Turchenko[2] and others proposed to take the stock price of Fiat company as the research object and used MLP neural network to make short-term prediction of stock price. However, the number of parameters required to use the MLP neural network is too large and the scalability is poor. Therefore, Avraam Tsantekidis[3] and others proposed a stock price forecast based on the CNN model. CNN realizes the local connection and weight sharing of neurons, retains important parameters, and reduces a large number of unimportant parameters. Compared with the MLP model, it achieves better learning results. However, CNN also has certain limitations. CNN focuses on spatial mapping and has certain advantages in processing image data. It is not fully applicable to learning time series. The LSTM model is a special type of structure of the RNN model, in which three control units of the forgetting gate, the input gate and the output gate are added. As the information enters the model, the control unit in the model will make judgments on the information, leaving the conforming information and discarding the non-conforming information. Based on this principle, LSTM can solve the problem of long sequence dependence in neural networks. Therefore, this paper proposes a stock price time series prediction method based on PCA-LSTM model.

PCA-LSTM Based Stock Price Time Series Prediction Model This experiment adopts keras as the deep learning framework. LSTM is a deep learning model used to solve the problem of gradient disappearance in long sequences. In order to solve the problem of predicting the daily closing price of Pingan Bank according to the data of the first N days (the forecast range is 1), a deep learning model consisting of two layers of LSTM module is designed, and a dropout layer is set in the middle to avoid over-fitting. Finally, a Dense layer is set to output a specific number. The workflow of the model is shown below. Data PCA LSTM Dropout LSTM Dropout Dense Output Parameter updates loss Figure 1. The workflow of model. 3.1. Data Pre-processing In the research of stock price prediction, there are many factors that affect the stock price, and they may be correlated with each other and have different effects on the result. The PCA method can concentrate these factors affecting the stock price on several main components, so that these main components reflect as much information as possible in the original variables, eliminating data redundancy and achieving data dimensionality reduction. 3.2. Stock Price forecasting Most data of stock market are time series data, and LSTM neural network has obvious advantages in processing time series information. Based on the RNN, LSTM adds memory cells to each neural unit in the hidden layer, making the memory information in the time series controllable. The information passes through several controllable gates (forgotten gates, input gates, and output gates) when passing between the various elements of the hidden layer. It can control the memory and forgetting degree of the previous information and the current information, so that LSTM has long-term memory function. In addition, LSTM has been successfully applied to many fields, such as image processing[4,5], speech recognition[6,7], handwriting recognition[8] and so on. Therefore, this paper uses LSTM neural network for stock price forecasting. Figure 2 shows the internal structure of LSTM

III. LSTM Based Stock Price Time Series Prediction Model

This experiment adopts keras as the deep learning framework. LSTM is a deep learning model used to solvethe problem of gradient disappearance in long sequences. In order to solve the problem of predicting the daily closing price of Pingan Bank according to the data of the first N days (the forecast range is 1), a deep learning model consisting of two layers of LSTM module is designed, and a dropout layer is set in the middle to avoid over-fitting. Finally, a Dense layer is set to output a specific number.

3.1. Data Pre-processing

In the research of stock price prediction, there are many factors that affect the stock price, and they may be correlated with each other and have different effects on the result. The PCA method can concentrate these factors affecting the stock price on several main components, so that these main components reflect as much information as possible in the original variables, eliminating data redundancy and achieving data dimensionality reduction.

3.2. Stock Price forecasting

Most data of stock market are time series data, and LSTM neural network has obvious advantages in processing time series information. Based on the RNN, LSTM adds memory cells to each neural unit in the hidden layer, making the memory information in the time series controllable. The information passes through several controllable gates (forgotten gates, input gates, and output gates) when passing between the various elements of the hidden layer. It can control the memory and forgetting degree of the previous information and the current information, so that LSTM has long-term memory function. In addition, LSTM has been successfully applied to many fields, such as image processing[4,5], speech recognition[6,7], handwriting recognition[8] and so on. Therefore, this paper uses LSTM neural network for stock price forecasting.

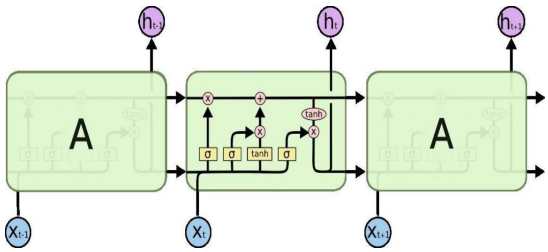

Fig: LSTM Arch

LSTM has three gates to control the storage state, including the forget gate, the input gate and the output gate.

(1) The Forgotten Gate is used to determine the information that is discarded. σ refers to the activation function of Sigmoid, and w refers to the weight, and b refers to the offset. Sigmoid outputs a value between 0 and 1, and f_t determines how much information about the state of the cell can pass at the previous moment. 0 means no information is allowed to pass, and 1 means all information is allowed to pass.

(2) The input gate is used to determine the information that needs to be updated. refers to how much information needs to be updated by Sigmoid, 0 means not updated, and 1 means completely updated. τ refers to the output of alternative content to be updated through tanh. The information to be transmitted is determined by the forgetting gate and the input gate.

(3)
The output gate determines the output information.

IV. EXPERIMENT

4.1. DATA SET

The data set adopted in this paper is the stock information of Pingan Bank, including the stock trading information of the three years from January 4, 2000 to December 28, 2021 (731 trading days). The data set is divided into 60% training set, 20% verification set and 20% test set. The model is trained using the training set, and the hyper parameter of the model is adjusted using the validation set, and finally the performance of the model is tested using the test set. The six technical indicators that affect the stock price of Pingan Bank are Close Price, Open Price, High Price, Low Price , Turnover and Trading Volume. The data comes from RESSET financial research platform.

Date	Date	Symbol	Series	Prev Close	Open	High	Low	Last	Close	VWAP	V
2021-04-26	2021-04-26	TATAMOTORS	EQ	294.0	297.0	299.0	294.60	295.85	295.40	298.49	30
2021-04-27	2021-04-27	TATAMOTORS	EQ	295.4	295.7	302.5	295.10	302.10	301.50	299.05	35
2021-04-28	2021-04-28	TATAMOTORS	EQ	301.5	303.5	309.5	303.00	305.15	305.90	307.22	44
2021-04-29	2021-04-29	TATAMOTORS	EQ	305.9	308.9	310.0	301.25	302.20	301.90	304.77	36
2021-04-30	2021-04-30	TATAMOTORS	EQ	301.9	298.2	301.3	292.55	293.10	293.85	297.01	36

FIG: DATA SETS OF TATA MOTERS

4.2. DATA PREPROCESSING

There may be correlation between various technical indicators that affect the stock price. In this paper, SPSS statistical analysis software is used to conduct principal component analysis on the six technical indicators, including Close Price, Open Price, High Price, Low Price, Turnover and Trading Volume. The analysis result is shown as table 1. According to the following table, the Close Price is the first principal component, its accumulative percentage accounts for 85.296%, and the cumulative percentage of general variance is greater than or equal to 85% to determine the principal component. Therefore, the first principal component is extracted and used to predict the stock price.

4.2.1 Feature Selection: In this step, data attributes are chosen that are going to be fed to the neural network.In this study Date & Close Price are chosen as selected features.

4.2.2 Train the NN model: The NN model is trained by feeding the training data set. The model is initiated using random weights and biases. Proposed LSTM model consists of a sequential input layer followed by 3 LSTM layers and then a dense layer with activation. The output layer again consists of a dense layer with a linear activation function.

4.2.3 Visualization: Using Keras and their function APIs the prediction is visualized.

4.2.4 Investigate different time interval: We repeated this process to predict the price at different time intervals. For our case, we took 2-month data set as training to predict 3-month, 6-month, 1 year & 3 years of close price of the share. In this different time span, we calculate the percentage of error in the future prediction. This would be different for different sectors. So, this will help to find a frame for the particular sector to predict future companies' net growth.

V. RESULTS

The proposed LSTM based model is implemented using Python. It has been observed from the result that for almost all the sectors the error level comes down drastically withthe test data for longer periods. So we suggest to apply this LSTM based model to predict the share price on long time historical data.

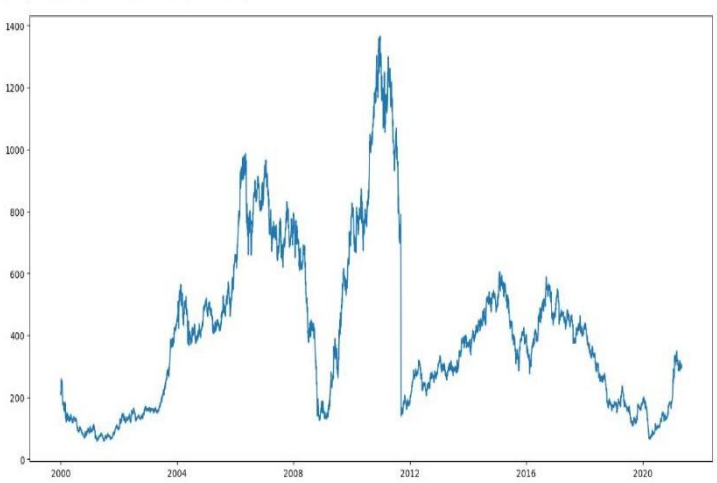

Fig.1: Normalization

```
15/15 [==============================] - 1s 97ms/step - loss: 8.3287e-05
Epoch 193/200
15/15 [==============================] - 2s 101ms/step - loss: 6.4112e-05
Epoch 194/200
15/15 [==============================] - 2s 102ms/step - loss: 6.4199e-05
Epoch 195/200
15/15 [==============================] - 2s 102ms/step - loss: 5.3731e-05
Epoch 196/200
15/15 [==============================] - 2s 105ms/step - loss: 5.5137e-05
Epoch 197/200
15/15 [==============================] - 2s 100ms/step - loss: 5.8930e-05
Epoch 198/200
15/15 [==============================] - 2s 100ms/step - loss: 6.2243e-05
Epoch 199/200
15/15 [==============================] - 2s 103ms/step - loss: 6.0862e-05
Epoch 200/200
15/15 [==============================] - 2s 101ms/step - loss: 6.2945e-05
Out[67]: <keras.callbacks.History at 0x1980bc53b50>
```

Fig.2: Model training

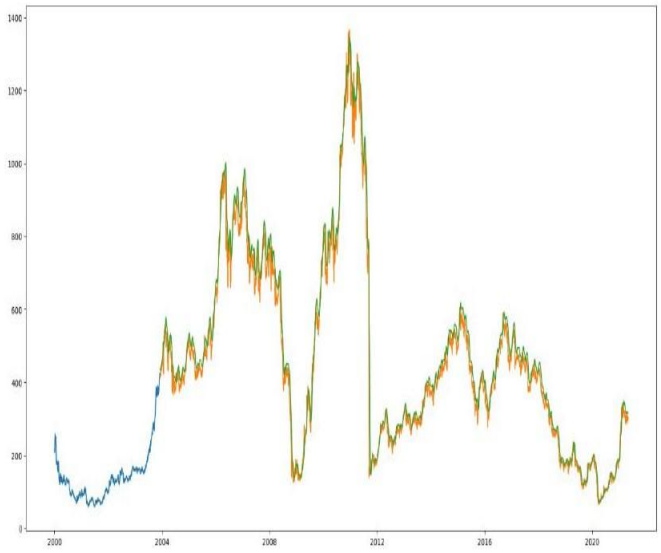

FIG.3: PREDICTION

VI .CONCLUSIONS

In this paper, we analyze the growth of the companies from different sector and try to find out which is the best time span for predicting the future price of the share. So, this draws an important conclusion that companies from a certain sector have the same dependencies as well as the same growth rate. The prediction can be more accurate if the model will train with a greater number of data set.
Moreover, in the case of prediction of various shares, there may be some scope of specific business analysis. We can study the different pattern of the share price of different sectors and can analyze a graph with more different time span to fine tune the accuracy. This framework broadly helps in market analysis and prediction of growth of different companies in different time spans. Incorporating other parameters (e.g. investor sentiment, election outcome, geopolitical stability) that are not directly correlated with the closing price may improve the prediction accuracy.

VII. REFERENCES

[1]. Stock Price Prediction Using LSTM on Indian Share Market by Achyut Ghosh, Soumik Bose1, GiridharMaji, Narayan C. Debnath, Soumya Sen

[2]. S. Selvin, R. Vinayakumar, E. A. Gopalkrishnan, V. K. Menon and K. P. Soman - Stock price predictionusing LSTM, RNN and CNN-sliding window model - 2017.

[3]. Murtaza Roondiwala, Harshal Patel, Shraddha Varma, “Predicting Stock Prices Using LSTM” in Undergraduate Engineering Students, Department of Information Technology, Mumbai University,2015.

[4]. Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, “An innovative neural network approachfor stock market prediction”, 2018

[5]. Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan Department of Computer Science and Engineering National Institute of Technology, Hamirpur – 177005, INDIA - Stock Market Prediction Using Machine Learning.

[6]. Pranav Bhat Electronics and Telecommunication Department, Maharashtra Institute of Technology, Pune. Savitribai Phule Pune University - A Machine Learning Model for Stock Market Prediction.

[7]. Anurag Sinha Department of computer science, Student, Amity University Jharkhand Ranchi, Jharkhand (India), 834001 - Stock Market Prediction Using Machine Learning.

[8]. Masoud, Najeb MH. "The impact of stock market performance upon economic growth." International Journal of Economics and Financial Issues 3.4 (2013): 788-798.

[9]. Schöneburg, Eberhard. "Stock price prediction using neural networks: A project report." Neurocomputing 2.1 (1990): 17-27.’

[10]. Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction: Algorithms and techniques.

[11]. Yu, P. and Yan, X., 2020. Stock price prediction based on deep neural networks. Neural Computing and Applications, 32, pp.1609-1628.

[12]. Adebiyi, Ayodele A., et al. "Stock price prediction using neural network with hybridized market indicators." Journal of Emerging Trends in Computing and Information Sciences 3.1 (2012).

[13]. Lu, Wenjie, Jiazheng Li, Jingyang Wang, and Lele Qin. "A CNN-BiLSTM-AM method for stock

price prediction." Neural Computing and Applications 33 (2021): 4741-4753.

[14]. Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014, March). Stock price prediction using the ARIMA model. In 2014 UKSim-AMSS 16th international conference on computer modelling and simulation (pp. 106-112). IEEE.

[15]. Adebiyi, Ayodele A., Charles K. Ayo, Marion Adebiyi, and Sunday O. Otokiti. "Stock price prediction using neural network with hybridized market indicators." Journal of Emerging Trends in Computing and Information Sciences 3, no. 1 (2012).