ZKG INTERNATIONAL

ISSN: 2366-1313

ADVANCED ASPECTS STOCK MARKET TREND ANALYSIS USING LSTM

D. KALYAN KUMAR[1], K. NIKHIL[2], CH. SRIKANTH[3], G. TIRUPATHI RAO[4], A. SAI PRAKASH[5].

[1] Associate Professor, CSE,Chalapathi Institute of Technology,Guntur, India

[2]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

[3]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

[4]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

[5]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

ABSTRACT: The rapid advancement in artificial intelligence and machine learning techniques, availability of large-scale data, and increased computational capabilities of the machine opens the door to develop sophisticated methods in predicting stock price. In the meantime,easy access to investment opportunities has made the stock market more complex and volatile than ever. The world is looking for an accurate and reliable predictive model whichcan capture the market's highly volatile and nonlinear behavior in a holistic framework. This study uses a long short-term memory (LSTM), a particular neural network architecture, to predict the next-day closing price of the S&P 500 index. A well-balanced combination of nine predictors is carefully constructed under the umbrella of the fundamental market data, macroeconomic data, and technical indicators to capture the behavior of the stock market in a broader sense. Single layer and multilayer LSTM modelsare developed using the chosen input variables, and their performances are compared usingstandard assessment metrics–Root Mean Square Error (RMSE), Mean Absolute PercentageError (MAPE), and Correlation Coefficient (R). The experimental results show that the single layer LSTM model provides a superior fit and high prediction accuracy compared tomultilayer LSTM models.

1. INTRODUCTION

Stock market movement has always been ambiguous for investors because of various influential factors. This study aims to significantly reduce the risk of trend prediction with machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals from Tehran stock exchange, are chosen for experimental evaluations. Stock Market prediction remains a secretive and empirical art. Few people, if any, are willing to share what successful strategies they have. A chief goal of this project is to add to the academic understanding of stock market prediction. The hope is that with a greater understanding of how the market moves, investors will be better equipped to prevent another financial

crisis. The project will evaluate some existing strategies from a rigorous scientific perspective and provide a quantitative evaluation of new strategies. There are several data mining algorithms that can be used for prediction purposes in the field of finance.

The stock price fluctuations are uncertain, and there are many interconnected reasons behind the scene for such behavior. The possible cause could be the global economic data, changes in the unemployment rate, monetary policies of influencing countries, immigrationpolicies, natural disasters, public health conditions, and several others. All the stock market

ISSN: 2366-1313

stakeholders aim to make higher profits and reduce the risks from the thorough market evaluation. The major challenge is gathering the multifaceted information, putting them together into one basket, and constructing a reliable model for accurate predictions.

Things were getting more interesting from the eighties because of the development in data analysis tools and techniques. For instance, the spreadsheet was invented to model financial performance, automated data collection became a reality, and improvements in computing power helped predictive models to analyze the data quickly and efficiently. Because of the availability of large-scale data, advancement in technology, and inherent problem associated with the classical time series models, researchers started to build models by unlocking the power of artificial neural networks and deep learning techniques in the area of sequential data modeling and forecasting. These methods are capable of learning complex and non-linear relationships compared to traditional methods. They are more efficient in extracting the most important information from the given input variables.

Several deep learning architectures have been developed to deal with various problems and the intrinsic structure of datasets. Information flows only in the forward direction in a basic feed forward neural network architecture. Since each input is processed independently, it does not retain information from the previous step. Thus, these models are ineffective in dealing with sequential data where series of prior events are essential in predicting future events. Recurrent neural networks (RNN) are designed to perform such tasks. The RNN architecture consists of loops, allowing relevant information to persist over time. Information is being passed from one time step to the next internally within the network. Therefore, the RNN is more suitable for sequential data modeling and time series applications such as stock market predictions, language translations, auto-completion in messages/emails, and signal processing. During the training process of the RNN, the cost or error is calculated between the predicted values and the actual values from a labeled training dataset. The error is minimized by repeatedly updating the networks' parameters (weights and biases) until the lowest possible value is obtained. The training process utilizes a gradient, the rate at which cost changes with respect to each parameter. The gradient provides a direction to move in the error surface by adjusting the parameters iteratively. This strategy is called back propagation, where the error is propagated backward from the output layer all the way up to the input layer. One of the challenges of this technique is that parameters can be anywhere in the networks, and finding a gradient involves calculations of partial derivatives with respect to all the parameters. This process sometimes needs a long chain rule, especially for the parameters in earlier layers of the networks.

2. LITERATURE SURVEY

TITLE:" Deep learning-based feature engineering for stock price movement prediction". By Wenjie Long, Zhenzhen Lu, and Lizhen Cui - Year-2019

CONTENT:

Stock price prediction is a critical and challenging task in financial analysis.

Traditional technical indicators used for stock price prediction are limited in their ability to capture complex market dynamics. In this paper, we propose a deep learning-based feature engineering approach for predicting stock price movements. Specifically, we use a convolutional neural network (CNN) to automatically learn high-level features from raw stock price data and then use these features to predict stock price movements. We compare our approach with traditional technical indicators and

ISSN: 2366-1313

other machine learning algorithms, such as support vector machines and random forests, on a range of benchmark datasets. Our experimental results show that the proposed approach outperforms traditional methods in predicting stock price movements. We also conduct feature importance analysis to identify the most important features for predicting stock price movements. Our findings demonstrate the potential of deep learning-based feature engineering approaches for improving the accuracy of stock price prediction.

"Efficacy of News Sentiment for Stock Market Prediction." AUTHORS: Shivam Kalra and Jai Shankar Prasad – Year - 2019. CONTENT:

The use of sentiment analysis in financial prediction has become increasingly popular over the years. The correlation between news sentiment and stock prices has been explored and exploited by many researchers. In this paper, we investigate the efficacy of news sentiment in predicting stock market trends. We first develop a sentiment analysis model using a dataset of news articles related to the stock market. We then use the output of this model to predict stock prices using various machine learning algorithms. Our experiments show that news sentiment can indeed be an effective predictor of stock market trends, with our sentiment-based models outperforming traditional time-series models in terms of accuracy. We also investigate the impact of different news sources and show thatsentiment from certain sources has a stronger correlation with stock prices than others. Overall, our results suggest that sentiment analysis can be a valuable tool for stock market prediction and can help investors make better decisions.

3. EXISTING SYSTEM

The methods used to predict the stock market includes a time series forecasting along with technical analysis, machine learning modelling and predicting the variable stock market.

4. PROPOSED SYSTEM

This study considers the computational framework to predict the stock index price using the LSTM model, the improved version of neural networks architecture for time series data. Then input sequence for the LSTM model is created using a specific time step. The hyper parameters such as number of neurons, epochs, learning rate, batch size, and time step have been incorporated in the model. The regularization techniques have beenutilized to overcome the over-fitting problems. Once the hyper parameters are tuned, the input data is fed into the LSTM model to predict the closing price of the stock market index.The quality of the proposed model is assessed through RMSE, MAPE, and R.

In a nutshell, plenty of research has been done in predicting the stock market. Some research focuses on complex statistical or without focusing on the type of attributable variables. Others use only the fundamental data without exploring additional factors that could influence the stock market prediction. There is a need to develop a model with a good combination of features of the stock market variables and simplicity in model architecture. Thus, our contribution is to create a model without adding any complexity in model architecture and maintaining well-balanced set of variables to capture the behavior of the stock market from multiple dimensions.

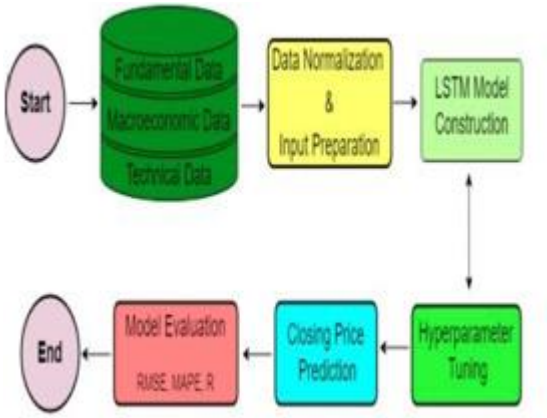

FIG-1: Schematic Diagram Of The

ISSN: 2366-1313

Proposed Research Framework.

MODULES

Data Collection:

In this study, S&P 500, a popular US stock market index, is used for the model prediction. The process of feature selection includes identifying the core factors that contribute to the index value fluctuations. Some features, such as fundamental data and technical indicators, are directly extracted from the underlying index. Other factors, namely macroeconomic variables, are selected based on their potential impact on the overall economy and broader markets. A complete 15 years of data have been collected from 2006 to 2020. The time frame selection incorporates two major bear markets, the financial crisis in 2008 and the COVID-19 pandemic in 2020. Thus, the construction of the model, including both bear and bull market, resembles the overall market scenario and may lead to a better prediction.

We start with a brief description of the features used in the proposed model. The closing price is predicted based on the fundamental trading data, macroeconomic data, and technical indicators of the underlying index. A combination of all the features from three different categories presented in Table 1 is input features. All the available features except Civilian Unemployment Rate and Consumer Sentiment Index are by default daily data. We have converted the monthly data to daily through the forward filling method to have uniformity among the variables.

Data	Source	Frequency	Abbreviation
Fundamental			
Open price	Yahoo	Daily	···
Close price	Yahoo	Daily	···
Macroeconomic			
Cboe volatility index	Yahoo	Daily	VIX
Interest rate	FRED	Daily	EFFR
Civilian unemployment rate	FRED	Monthly	UNRATE
Consumer sentiment index	FRED	Monthly	UMCSENT
US dollar index	Yahoo	Daily	USDX
Technical indicator			
Moving average convergence divergence	···	Daily	MACD
Average true range	···	Daily	ATR
Relative strength index	···	Daily	RSI

Fig-2: List of potential features for the model.

Pre-Processing Data

pre-processing is a part of data mining, which involves transforming raw data into a more coherent format. Raw data is usually, inconsistent or incomplete and usually contains many errors. The data pre-processing involves checking out for missing values, looking for categorical values, splitting the data-set into training and test set and finally do a feature scaling to limit the range of variables so that they can be compared on common environs.

Feature selection strategy

All the input variables explained above have some level of contribution in predicting the closing price. A correlation heat map of all input features explained above is presented. The numerical value in the heat map represents the correlation between the variables on the horizontal and the vertical axes. For instance, the diagonal value of the matrix is 1, the correlation between the variable to itself. Thus, the values on the diagonal are unimportant in our analysis. The entries on the off-diagonal are used for the feature selection process. These values are presented based on the intensity of the color, which is also the indicator of the level of relationship between the given variables. The vertical bar next to the graph shows the intensity of the color on the

ISSN: 2366-1313

scale from 0 to 1. The correlation between the closing price and the remaining variables may be high or low. It indicates the intensity of the relationship. For example, the correlation between closing price to consumer sentiment index is 0.67, which is moderately positive. Thus, the consumer sentiment index may playa vital role in the closing price prediction. The same explanation applies to the rest of the entries presented in the graph. Most importantly, the high correlation (positive or negative) between the predictors can behave as a duplicate feature. In such case, either one of the highly correlated variables can be dropped from the analysis. For instance, open price and closing price have a strong correlation, indicating duplicate features. One of the features can be discarded because it will not provide significant additional information for the prediction. In this study, the open price has been dropped from further analysis. The correlation coefficient of 0.80 is considered as a threshold for removing the duplicate features. After removing the open price, none of the pair-wise correlations between the remaining variables exceeds the threshold. The heat map validates the statistical significance of the variables chosen for the prediction.

After careful analysis of heat map in the feature selection process, the 11 variables are considered as input variables. The partial snapshot of the dataset used in the study is presented in Table 2.

Fig-3 Correlation heat map among the attributable variables

SYSTEM ARCHITECTURE

LSTM cell takes 3 different pieces of information: the current input sequence , the short-term memory from the previous cell h_{t-1}, and the long-term memory from the previous cell state c_{t-1} at time . The forget gate takes the information from x_t and h_{t-1} and produces the output between 0 and 1 through the sigmoid layer and then it identifies which information to discard from the previous cell state c_{t-1}. When the valueis 1, it stores all the information into the cell while with a value of 0, it forgets all the information from the previous cell state. Similarly, the input gate identifies which information to be updated from the change gate. The output gate decides which information to be taken as an output from the present cell state.

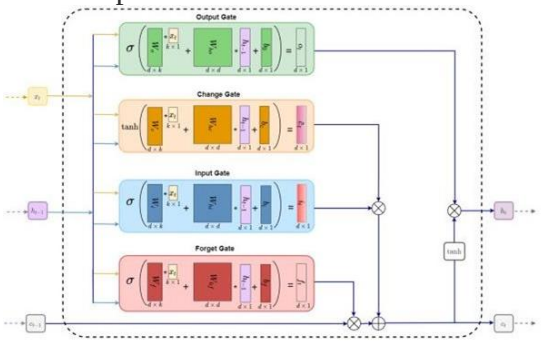

Fig-4 Long short-term memory (LSTM) architecture.

CONCLUSION

Stock price prediction is the area of high interest for equity traders, individual investors, and portfolio managers. However, precise and consistent stock price prediction is a difficulttask due to its noisy and nonlinear behavior. There are several factors that can impact the prediction such as fundamental market data, macroeconomic data, technical indicators, and others. This study focuses on developing LSTM based models to predict S&P 500 index's closing price by extracting a well-balanced combination of input variables capturing the multiple aspects of the economy and broader markets. Both single and multilayer LSTM architectures have been implemented and

ISSN: 2366-1313

their performances are analyzed by using various evaluation metrics to identify the best model. The experimental results show that single layer LSTM model with around 150 hidden neurons can provide a superior fit and high prediction accuracy compared to multilayer LSTM. The proposed model can be easily customized to apply in other broad market indexes where the data exhibits a similar behavior. Interested stakeholders can use the proposed model to better inform the market situation before making their investment decisions.

FUTURE SCOPE

In the near future, we plan to explore the possibility of incorporating unstructured textual information in the model such as investor's sentiment from social media, earning reports of underlying companies, the immediate policy-related news, and research reports from market analysts. Another potential direction of the future work can be developing hybrid predictive models by combining the LSTM with some other neural networks architectures. To improve the prediction accuracy even further, we also plan to implement hybrid optimization algorithms to train the model parameters by combining the existing local optimizers with the global optimizers such as genetic algorithms and particle swarm optimization algorithm.

REFERENCE

1. J. J. Murphy, Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. Penguin, 1999.
2. T. Turner, A Beginner's Guide to Day Trading Online, 2nd ed. New York, NY, USA: Simonand Schuster, 2007.
3. H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Adil, M. M. Selim, and K. Muhammad, ''A local and global event sentiment based efficient stock exchange forecasting using deep learning,'' Int. J. Inf. Manage., vol. 50, pp. 432–451, Feb. 2020.
4. W. Long, Z. Lu, and L. Cui, ''Deep learning-based feature engineering for stock price movement prediction,'' Know. -Based Syst., vol. s164, pp. 163–173, Jan. 2019.
5. J. B. Duarte Duarte, L. H. Talero Sarmiento, and K. J. Sierra Juárez, ''Evaluation of the effect of investor psychology on an artificial stock market through its degree of efficiency,'' Contaduría y Administración, vol. 62, no. 4, pp. 1361–1376, Oct. 2017.
6. Lu, Ning, A Machine Learning Approach to Automated Trading. Boston, MA, USA: BostonCollege Computer Science Senior, 2016.
7. M. R. Hassan, B. Nath, and M. Kinley, ''A fusion model of HMM, ANN and GA for stock market forecasting,'' Expert Syst. Appl., vol. 33, no. 1, pp. 171–180, Jul. 2007.
8. W. Huang, Y. Nakamori, and S.-Y. Wang, ''Forecasting stock market movement direction with support vector machine,'' Compute. Opler. Res., vol. 32, no. 10, pp. 2513–2522, Oct.2005.
9. Kalra S, Prasad JS. Efficacy of News Sentiment for Stock Market Prediction. Proc Int ConfMach Learn Big Data, Cloud Parallel Compute Trends, Perspectives Prospect Com 2019.Published online 2019:491-496. doi:10.1109/COMITCon.2019.8862265
10. Menon A, Singh S, Parekh H. A review of stock market prediction using neural networks. 2019 IEEE Int Conf Syst Comput Atom Networking, ICSCAN 2019. Published online 2019.
11. Fama, E.F. (1965). The behavior of stock market prices.