Intelligent Stock Market Prediction

By

Rohan Maheshwari
17bit095

Rahul Gupta
17bit091

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Ahmedabad 382481

Intelligent Stock Market Prediction

Minor Project

Submitted in partial fulfillment of the requirements

For the degree of

Bachelor of Technology in Information Technology

By

Rohan Maheshwari
17bit095

Rahul Gupta
17bit091

Guided By
Dr. Ankit Thakkar
[DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING]

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Ahmedabad 382481

CERTIFICATE

This is to certify that the Minor Project entitled "Intelligent Stock Market Prediction" submitted by Rohan Maheshwari(17bit095) and Rahul Gupta(17bit091), towards the partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Engineering/ Information Technology of Nirma University is the record of work carried out by him/her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination.

Dr. Ankit Thakkar
Assistant Professor
Department of Computer Science & Engg.,
Institute of Technology,
Nirma University,
Ahmedabad

Dr. Madhuri Bhavsar (HOD)
Dept. of Computer Science & Engineering,
Institute of Technology,
Nirma University,
Ahmedabad

ACKNOWLEDGEMENT

We wish to acknowledge our guide Dr.Ankit Thakkar, who has been totally supportive throughout the whole project and helping us channel our resources into better paths. It is because of his kind and helpful supervision our tasks became a whole lot easier. Finally, we would like to give special thanks to our friends and family for their kind support and never letting us give up. Without their support this project would have not been possible.

ABSTRACT

The financial market is a very complex market, one of the primary aims of investment in such a market is to get higher and better benefits. It is influenced by a large number of factors wherein the estimation of future market conditions be it internal or external is very challenging. This paper aims to summarize and put forward some approaches and techniques to conduct the prediction of stock trends by taking fusion as an approach to integrate data and characteristics and improve the prediction basing the proposed solution on the combinational approach that can help and aid each other. We have conducted our review of the recent years, i.e. 2018 – 2020, by short-listing the articles and papers that have employed various fusion based techniques for stock market-based applications.

CONTENTS

1 Introduction

1.1 Introduction and motives

The financial market seems to be a very attractive field of study as it offers a wide variety of opportunities for investors, market analysts, as well as researchers from various disciplines. This paper aims to aid the study of various market analyzing techniques and also propose the best fit method for the same. Financial markets happen to be very complex in nature with many instruments such as stocks, bond, derivatives, currencies, etc. The economic market can be considered as a combination of financial investments, transactions, potential earning and/or losing, and several other actions that are performed at a massive level. Stock market is a financial market where the new issues of stocks, i.e., initial public offerings (IPOs), are created and sold at the primary market whereas the succeeding buying and selling are carried out at the secondary market. The primary motivation behind investing in a stock market is to gain potential benefits of the investment.

Research on Convoluted Neural Network models for video processing has considered learning 3D spatio-temporal filters over raw sequence data and learning of frame-to-frame representations which incorporates instantaneous optic flow or trajectory-based models aggregated over fixed windows or video shot segments. Such models explore two extrema of perceptual time-series representation learning: either learn a fully general time-varying weighting, or apply simple temporal pooling. The paper focuses on fusion based research and has been categorized into three basic fusion types: information fusion, feature fusion, and model fusion; also the report consist of a survey and comparison of various

methods involved in evaluating the models used for predicting the stock trends.

1.2 Related Works

It is observed that the majority of the existing review articles have not precisely focused on fusion in the stock market, however, they might have referred to fusion-based articles as a part of the survey. Therefore, we compare our survey with the existing surveys even if they are partially related to the central focus of our survey. In this we tend to briefly categorize fusion into data fusion, feature fusion, and model fusion. With every fusion class, we tend to explore the present analysis work having primary applications of stock value and/or trend prediction, portfolio management, risk/return statement, similarly as different stock exchange concepts; although the reviewed articles are primarily supported fusion techniques, a number of them may not have expressly referred their work as a fusion approach.

For collection analysis articles for this survey, we tend to apply a scientific strategy as follows. using Google Scholar program, we tend to conduct Associate in Nursing initial search with terms "stock market prediction" and "fusion"; for a serious concentration on the recent studies, we tend to restrict our search at intervals years 2011 – 2020. we tend to extended the search victimization extra terms like "information fusion", "feature fusion", and "model fusion" whereas the potential applications associated with stock markets were gathered using the terms "stock value prediction", "stock trend prediction", "portfolio management", "risk prediction", "return prediction", and commutation "prediction" by "forecasting" in every of those search operations, similarly as victimization different

monetary market terms; to confirm that these terms were closely associated with the particular work of the articles, we tend to restricted our search specified articles having such terms in their titles were place together. we tend to additionally set the exclusion criteria for articles ensuing with the thought of search operations;

In comparative analysis we tend to use Artificial Neural Networks strategies like forecasting, linear regression, and Moving averages.

• **In forecasting method:** The system is taking into consideration the last 3 days and also the current year stock portfolio, closing expected date, and performs calculations on the value from the data for predicting the stock portfolio value.

• **Moving averages method:** System takes the 10 days' stock portfolio closing price from the predicting date and calculates the stock price. During this technique, an alpha is estimated based on the previous mentioned parameters for the predicting price and it's a constant for this technique.

• **Regression method:** Its applied mathematics assessing the association between two variables. For pattern recognition, it will not realize the relationship between two variables.

• **Neural networks method:** It has effective, general purpose approach classification, and a bunch of particularly statistical predictions with an excellent degree of accuracy. Even so, their performance is not continuously satisfactory.

Back propagation rule is the best method to be utilized in Feed Forward Neural Network (FNN)[3]. As a result, it reduces error between the particular output and desired output in an exceedingly gradient descent manner.

- **SVM:** The SVMs were at first developed as Classification algorithms. Given a collection of training examples, each marked as happiness to at least one or the opposite of 2 categories, an SVM training rule builds a model that assigns new examples to at least one class or the opposite, creating it a non-probabilistic binary linear classifier. Decision of comparison presents itself within the following type -

a) Sale - Our system recommends client to sale a stock according to that client can create the profit

b) Buy - Our system recommends client to shop for a stock

therefore in an exceedingly means that client can create the profit.

c) Hold - Our system recommends client to carry a stock therefore it means that the client avoids his/her loss.

In case of getting the preceding terms in the main occurring within the reference section, the articles were examined to search out if the terms were being referred for the connected work and not for the particular planned approach; in such a situation, the articles were excluded from being thought of in our survey. to attenuate the redundancy, we tend to eliminate similar articles from being reviewed, if found. On the opposite hand, it had been found that though specific fusion ways were applied during a set of articles, "fusion" term had not been specifically accustomed describe the same; thus, we tend to closely studied an oversized range of articles

to spot whether or not they depicted fusion at any level; therefore, for associate complete survey, we tend to advanced our assortment by as well as such fusion-based articles similarly because the articles associated with those downloaded victimization the said search ways.

1.3 Proposals

• Data set - To predict any price, it needs a large amount of input, and processing this information along with applying algorithms on that, we are able to predict the long run price. So, during this system, we have a tendency that requires share market knowledge which can be taken from totally different firms.

• Data Processing: Applying 5 algorithms on input file set -

a) Forecasting algorithm-The method of constructing predictions on the basis of the long run supported past and present knowledge and analysis of trends is termed prediction. Example may be estimation of some variable of interest at some specified future date.[4] Each may ask formal applied mathematics strategies using statistics, cross-sectional or longitudinal knowledge, or class to less formal judgmental strategies.

b) Moving Averages - In statistics, a moving average (rolling average or running average) could be a calculation to research knowledge points by making a series of averages of various subsets of the complete knowledge set. It's additionally referred to as a millimeter or rolling means and could be a variety of finite impulse response filters. Variations include: straightforward, and additive, or weighted forms.

c) Neural Network Algorithm - Neural networks formed from the number of neurons. The neural network is made by input layer, hidden layer and output layer. Neural network is looking at

alternative parameters like a variety of layers in the network, number of neurons in input layer, number of neurons in hidden layer, Rate of momentum, Network learning rate. It offers correct output when processing input files.

d) Regression Algorithm - A regression could be an applied mathematics analysis assessing the association between two variables. It's accustomed to realize the connection between two variables. we are going to I st realize slope, intercept and use it to make regression of y on x. e)

e) Multitask learning: Multitask learning aims to improve all tasks simultaneously by combining the common knowledge from all tasks. Each task provides extra training data for the parameters that are shared or constrained,

Serving as a form of regularization for the other tasks [14].

f) Comparison of Algorithms - Using artificial intelligence these 5 algorithms are compared. Exploitation multilayer perception neural network, that rule offers the best prediction of exchange is known. The neural network is formed from number of neurons. The model contains a variety of neurons that contain bound information which information passes from one neuron to a connected neuron. The data is processed at neurons and the output is produced [5]. In our system, ANN is formed from layers of neurons. The first layer is connected to the input data, then in the multilayer network there are no hidden layers and final output layers are present. In SVM, this sliding window is the basis for how we are able to flip any statistical dataset into a supervised learning problem. It may be work to show a statistic into either a regression or a classification supervised learning drawback for real valued or labeled statistic values. This is the most effective rule for calculating expected price

and creating calls because it provides least errors. In this Comparison we found that using SVM is the best algorithm which provides best accuracy.

1.4 Contribution

Predictions need study of varied authoritative information yet as derivation of the previous patterns of price movements. Hence, data fusion may be integrated with the present techniques to create reliable information bases for rising exchange prediction.

The major concern around exchange analysis includes basic also as technical analyses; whereas the involved options of such analyses disagree, their fusion may be useful for a stronger info supply whereas predicting the exchange.

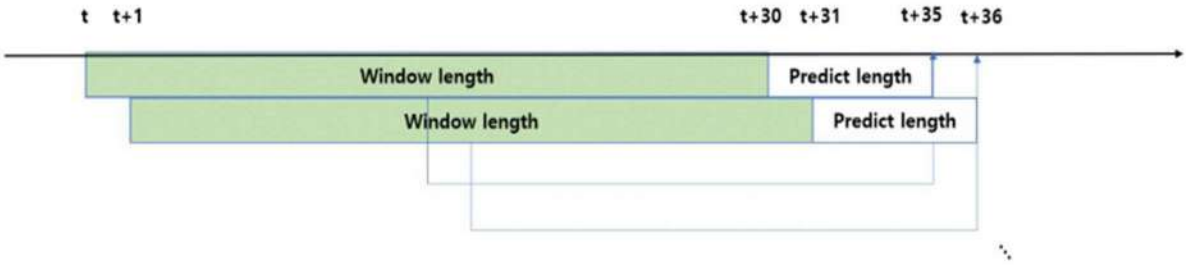
2 Methods and Dataset

2.1 Data representation

In this survey, we have used minute-by-minute SPY ticker data, it has the largest trading volume in ETF markets. We have also collected trade high, low, open, and close price and volume data, which cover 97,474 data points from the Thomson Reuter Database. By this financial time series data, we have created different representations to match input types of each model to extract features from the CNN and LSTM.

How we have set the training, validation, and testing dataset. Here we use the 68,800 data points for training data; 10,000 data points for validation data; and 19,474 data points for testing data.

Here we have set the window length to 30 minutes, rolling window to 1 minute, and predict term to 5 minutes. It means that we will predict the stock price after 5 minutes by looking at the data for the previous 30 minutes based on minute-by-minute. This setting is applicable for each representation since we combine the CNN with LSTM.



2.2 Fusion based research topics proposed

Information fusion

Quantitative Data — It consist with analysis of historical stock data.

Technical Analysis — It contains all the different kinds of indicators (Trend, Momentum, Volatility. etc.)

Fundamental Analysis — They are more concerned with the basic analysis of stocks (Profitability, Debt-Pay ability. etc)

Crowd-Sourced Data — They are governed by biological factors and more sentiment dependent.

Potential Research and Survey Areas Indian Stocks - Indian stock market tends to be very volatile and a different mix of information fusion parameters can be used to correctly identify and propose fusion parameters.

No inter-relation proposed as such - Although a lot of papers have successfully classified various categories for information fusion, a very less number of papers had integrated the best fit.

Feature Fusion

Derived Features - These features are made up using many other features.

Type - These can be further divided into Time Based, Pattern Based, and Chart Based, depending upon the feature taken into consideration.

Analysis — It includes various analysis factors such as Correlation, Diversity, and Informatics. [8]

Operation - Various operation-based factors.

Potential Research and Survey Areas Strategies proposing efficient feature selection - We could conduct our research to include various strategies to select appropriate mix of features.

Better feature derivation techniques - This will help us in selecting and potentially mixing and removing features with similar effect on the various models.

Model Fusion

The proposed models and the best predicted model have been described in the previous section.

2.3 Performance measure

We select three performance measures, the root mean square error (RMSE), the root mean absolute error (RMAE), and the MAPE, to evaluate the predictive power of our proposed models. When we train models, we use RMSE as a loss function, which means that the model is trained to reduce RMSE. RMSE is a good measure for revealing relatively large forecast errors , RMAE is useful for revealing the systematic bias of the model, and MAPE is a measure of the accuracy of predictions in statistics. These equations are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{1,i} - x_{2,i})^2}$$

$$RMAE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_{1,i} - x_{2,i}|}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{x_{2,i} - x_{1,i}}{x_{1,i}}\right|$$

where N is the number of data points, x1,i is a predicted value, and x2,i is a real value.

3 Results and Discussion

3.1 Results Achieved

We check three results of this experiment. First, we determine the optimal stock chart image to predict stock prices. Second, we show that fusion chart images perform better than stock chart images do. Last, we need to make sure that our proposed feature fusion LSTM-CNN model is meaningful. As we mentioned before, we use RMSE, RMAE, and MAPE as performance measures.

Comparison of out-of-sample results for stock chart images using the SC-CNN model.

Data	RMSE	RMAE	MAPE
Candlestick	0.1258	0.2896	0.0338
Line	0.1800	0.3756	0.0568
F-line	0.1549	0.3427	0.0473
Bar	0.1278	0.2933	0.0347

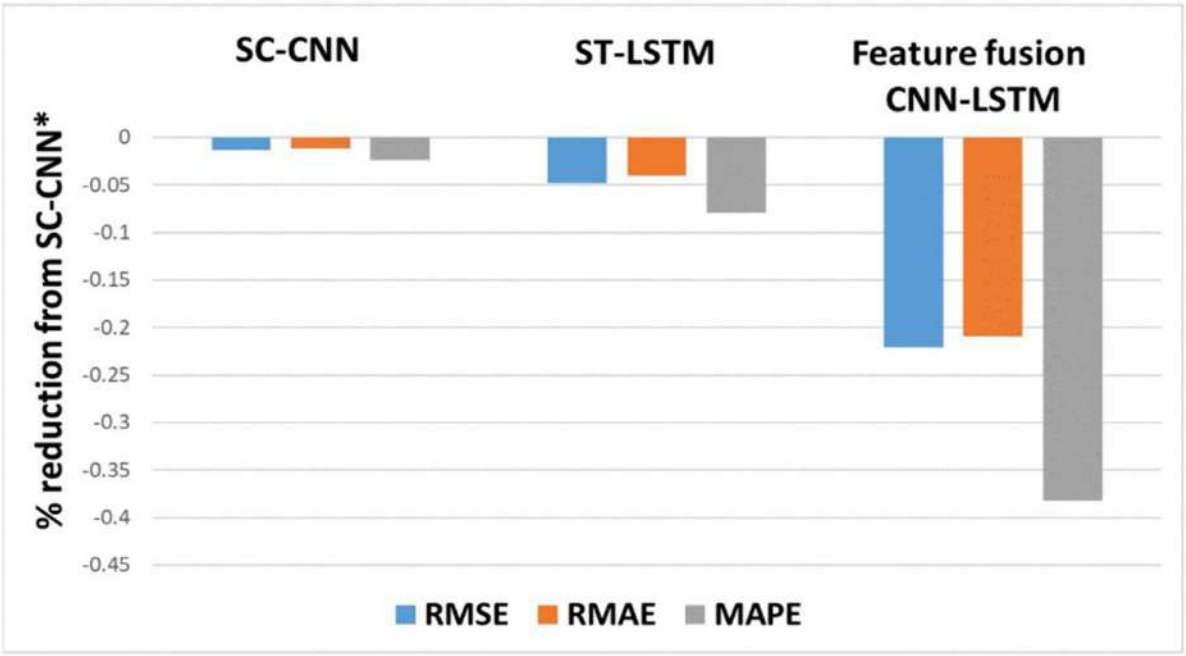
Comparison of out-of-sample results for fusion chart images using the SC-CNN model.

Data	RMSE	RMAE	MAPE
Candlebar	0.1241	0.2862	0.0330
Linebar	0.1685	0.3727	0.0559
F-linebar	0.1436	0.3287	0.0435

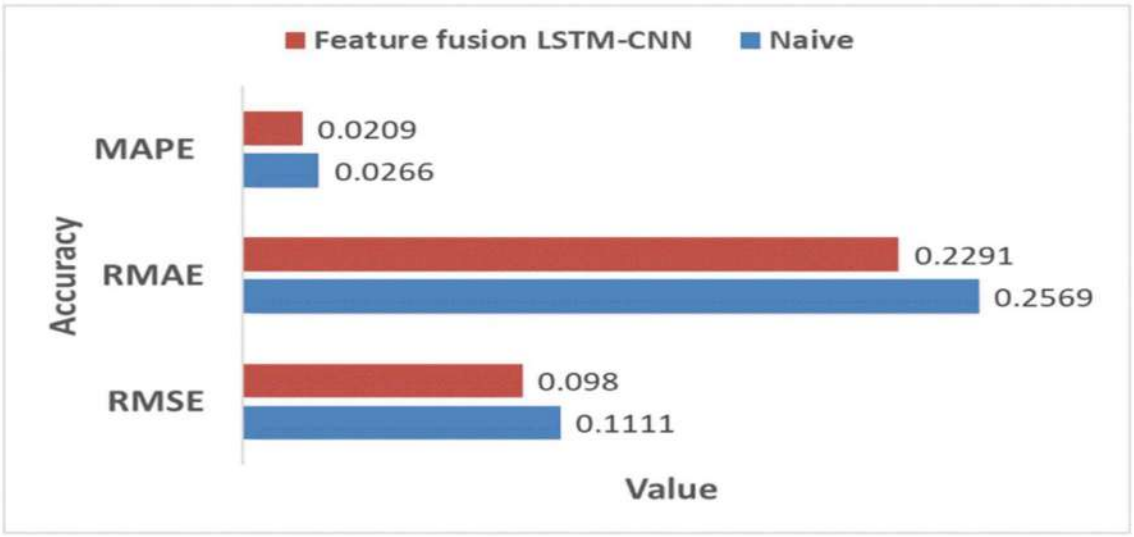
Comparison of out-of-sample results for stock chart images and stock time series data using the feature fusion LSTM-CNN model.

Data	RMSE	RMAE	MAPE
Candlebar and stock time series	0.0980	0.2291	0.0209
Linebar and stock time series	0.1081	0.2354	0.0231
F-linebar and stock time series	0.1063	0.2318	0.0222

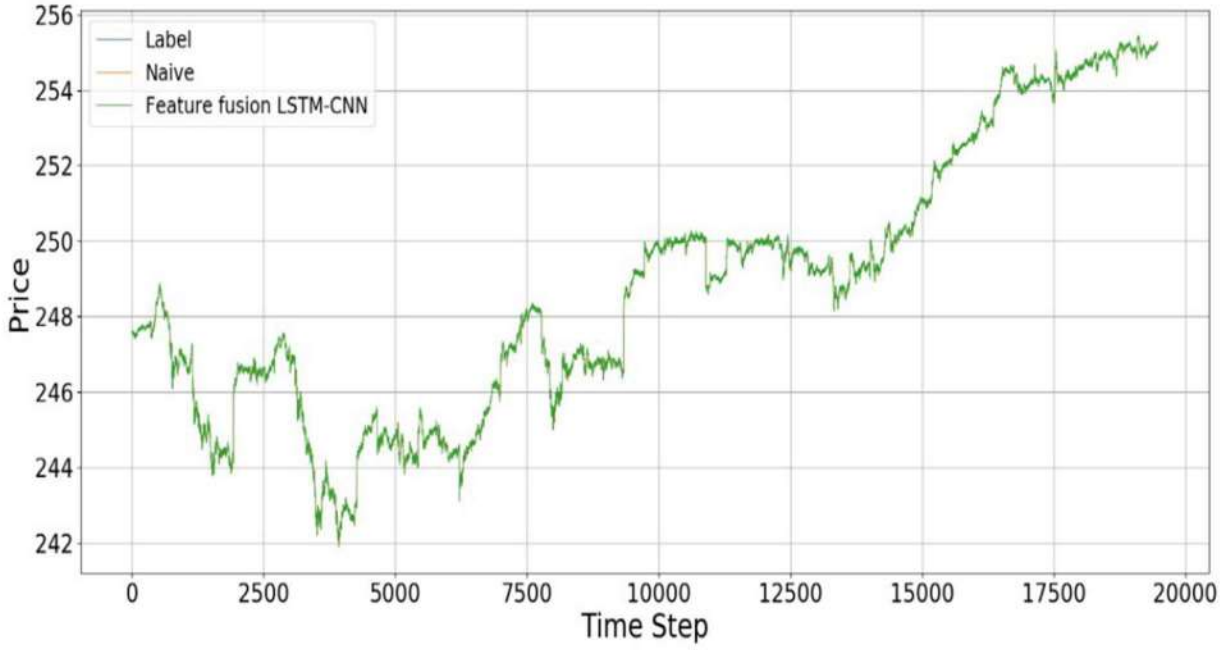
Compared with the results of the ST-LSTM model, the out-of-sample loss of the feature fusion LSTM-CNN model using candlebar charts and stock time series as inputs decreased by 18.18% (RMSE), 17.56% (RMAE), and 32.87% (MAPE). Experiments with other fusion chart images and stock time series showed better performances than that of the ST-LSTM model. We compared the accuracy with the naive model and feature fusion LSTM-CNN model. Naive model is based on the assumption that the value at the previous time point is the same as the value at the later time point. We found that feature fusion LSTM-CNN using fusion chart images and stock time series has outperformed the Naive model.



Comparison of prediction errors among three models based on SC-CNN. The input data are candle bar charts, which are the best performing fusion chart images, and stock time series data. Note: SC-CNN uses a candlestick chart, which is a stock chart image only.



Comparison of accuracy between feature fusion LSTM-CNN and naive model



An example of predicting stock prices using the feature fusion LSTM-CNN model and Naive model on the testing dataset. The input data are candle bar charts and stock time series.

Conclusion

In this paper, we have presented a better and well researched approach based on fusion parameters and have also identified various potential research areas and have provided a very detailed report for the same. The nonlinear stock market predictions have been attractive for investors as well as researchers; enormous amount of work has been carried out at various levels where different aspects of market dynamics are explored to make reliable predictions. Amongst such techniques, fusion plays an important role

in determining the usefulness of combinational information. To study the significance of fusion in stock.

References

[1] Dr. Ankit Thakkar and Kinjal Chaudhiri "Fusion In stock market prediction: A decade survey on the necessity, recent developments, and potential future directions."

[2] D. A. Kumar and S. Murugan, "Performance Analysis of MLPFF Neural Network Back Propagation Training Algorithms for Time Series Data", Congress tin Computing stock back propagation neural in 2014 World and Communication.

[3] Tay FEH, Cao LJ. A comparative study of saliency analysis and genetic algorithm for feature selection in support vector machines. Intelligent Data Analysis 2001 ; 5:191—209.

[4] G. Dong, K. Fataliyev and L. Wang, "One-step and multi-step Tech-nologies (WCCCT), 2014, pp. I 14—1 19. Ahead networks", prediction Communications and Signal Processing (ICICS) 2013 9th International Conference, 2013, pp. 1—5.
Stock Market Prediction using Linear Regression and Support Vector Machines.

[5]Hiransha M , Gopalakrishnan E. , Vijay Krishna Menonab, Soman K. "NSE Stock Market Prediction Using Deep-Learning Models "
[6] Stock Market Prediction Using LSTM Recurrent Neural Network Adil MOGHAR ,Mhamed HAMICHE.

[7] Hu G, Hu Y, Yang K, Yu Z, Sung F, Zhang Z, et a1. Deep Stock Representation Learning: From Candlestick Charts to Investment Decisions. arXi\ [Internet].(i). http://arxiv.org/abs/17(J9.(13803)

[8] Chen J-F, Chen W-L, Huang C-P, Huang S-H, Chen A-P. Financial Time-Series Data Analysis Using Deep Convolutional Neural Networks. In: 2016 7th International Conference on Cloud Computing and Big Data (CCBD) [Internet]. 2016. p. 87—92. http://ieeexp1ore.ieee.org/document/7979885/

[9] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. 2016. p. 770—8. http://ieeexplore.ieee.org/document/7780459/

[10] Stock Market Prediction Using LSTM Recurrent Neural Network Adil MOGHAR ,Mhamed HAMICHE.

[11] Li Z., and Hoiem D. (2018). Learning without Forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[12] q] Donahue .I, Hendricks LA, Rohrbach M, Venugopalan S, Guadarrama S, Sienko K, et a1. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. IEEE Trans Pattern Ana1 Mach 1nte11. 2017; 39(4):677—91.

[13] Bao W, Yue J, Rao Y. A deep learning framework for financial time series using stacked auto encoders and long-short term memory. PLoS One. 2017

[14] Siripurapu A. Convolutional Networks for Stock Trading. Stanford University Department Computer Sci. 2014;1—6.