Optimizing Stock market prediction using Long Short Term Memory

R. Kavitha, Sonal Singh

Abstract: Stock market prediction has been an important issue in the field of finance, engineering and mathematics due to its potential financial gain. Stock market prediction is a process of predicting the future value of a company stock or other financial instrument traded in financial market. Stock market prediction brings with it the challenge of proving whether the financial market is predictable or not, since stock market data is of high velocity. This project proposes a machine learning model to predict stock market price based on the data set available by using LSTM model for performing prediction by de-noising the data using wavelet transform and performing auto-encoding on the data. The process includes removal of noise, preprocessing, feature selection, data mining, analysis and derivations. This project focuses mainly on the use of LSTM algorithm along with a layer of neural network to forecast stock prices and allocate stocks to maximize the profit within the risk factor range of the stock buyers and sellers.

***Keywords*: prediction; LSTM; de-noise; auto encoders; feature selection**

I. INTRODUCTION

Stock market also known as share or equity market is accumulation of investor and a purveyor. In simple word it is a network of relationship formed over stocks between its seller and buyer. Stock market act as an important source for fund raising for a company. Stock value prediction is modus to intuit the ensuing of future value of the stock. Stock prediction is given its due importance with the advent of time as efficient and successful prediction of stock could capitulate high profits. It enables the decision making process more easier and relevant. Stock market prediction has become centre of interest for many researches from various background due to delinquent mercantile and enterprise adjure . It is quite evident that the stock data prediction is highly exigent and enthralling at the same time for its qualities of being precarious that is nonlinear , unpredictability, riskiness and scruple. It consist of lots of unwanted values . Many of the machine learning algorithms are being used to make prediction over the stock data but the success rate achieved is less , also to be noted the fact that only movement of stock data can be predicted not the exact values. The prediction sometimes are acceptable being fairly close enough but may deviate due to various factor like political policies or internal policy of the company.

II. RELATED THEORETICAL CONCEPT

A. Stock Market Data

When any industry is established the only parties involved as the shareholder are the founder and the cofounder and the investors. But with the advent of time the company want more capital, join hands with more capitalist. at the same time many initial investor come to sell their share and exchange monetary benefit from it. Thus these parts of the share is termed as stock which is broadly classified into two major categories common stock and preferred stock. Common stock are the stocks in the form of dividends who can claim on monetary benefit from the company while preferred stocks are deprived of their voting ability and function mostly like a bind. A stock data comprises of various rows and columns pertaining to its time domain nature. The attributes include Date, high ,low ,open price ,close price ,turnover and total number of trading.

B. Time Sequence Data

It is a kind of analysis of trend over a period of time. A time series data is a data in its sequential form it is a cluster of data that a variable takes over a monitored period of time. In simple terms, the data which changes its value over a period of time and is recorded over a consecutive time duration for some set period of time is considered as time sequence data.

C. RNN and LSTM

RNN consist of a special layer commonly known as the hidden layer which is responsible for retaining the memory in the neural network. It basically helps the RNN to retain and remember some of the data from the entire input sequence that one feeds in the neural network. Even after having the feature of memory retention and ability to understand sequences. Recurrent neural network offers many challenges to the user such as it is difficult for RNN to remember very long data inputs or variable for long period of time. The issue is vanishing gradient and exploding tension. These issue makes it very difficult for the RNN to remember the data. As the issue is explained that as the memory reads the input further simultaneously it forgets the previous data input in the memorAnother most important issue is that training of the dataset for RNN is a very time consuming and difficult work. To eliminate such issues of the RNN a new neural network was introduced. LSTM are the more improvised version of RNN .It is quite different from the normal neural network for its feedback connection that makes it capable enough to perform all the operations that can be performed on turing machines also.

Revised Manuscript Received on November 10, 2019.

R.Kavitha, Department of CSE, Vel Tech RangarajanDr.Sagunthala R & D Institute of Science and Technology
Sonal Singh, Student, Department of CSE, Vel Tech Rangarajan Dr.Sagunthala R & D Institute of Science and Technology

Long short term memory unit can retain data in the memory for long period of time. This property of LSTM accounts for its architecture that is composed mainly of gates . The three gates that helps to form the LSTM model are output ,input and forget gate . Amongst these forget gate is more prominent as it is responsible for memory. As the LSTM has three gates the forget gate has cell state the input to forget gate is multiplicative to the cell input and this value is treated upon by the sigmoid function for activation. The resulted value decides the future of the cell state that is whether to retain or remove the cell state value . this is how the memory management function in the LSTM.

D. Data Cleaning

Stock market is very dynamic in nature and so to prepare the data for prediction we have to remove rare and high frequency data to improve the chances for better prediction. We need to clean the data for fluctuating signals. We have tried to achieve this using wavelet transform. The main motive to use wavelet transform is the ability to capture both location and frequency. Data cleaning process ensures that the resultant analysis is free of any ambiguity. De noise in simple terms is a process of reconstructing a signal from a noisy one. We cannot make use of spline estimator as they do not go well with local structures that is they do not resolve local structure well enough which is a must in neurophysiologic signals. Fourier transform is also avoided as it is more suitable for static signal rather than dynamic signal like stock market. The ‘haar’ wavelet family is used on the data but ‘symlet’ and ‘db’ wavelet family were also tested.

E. Feature Selection

Feature selection is one of the core concept of machine learning which contributes largely on the performance of the model. As we all are aware that features which are irrelevant or partial data can impact the model performance negatively. So these features which we select for the training should reduce training time and over fitting and simultaneously it should improvise the accuracy of the model. The above qualities in stock data is achieved by using auto encoders. Auto encoder is also a neural network which is self sufficient to learn and understand the compressed data. It is typically made of layers where encoding and decoding layers are used.

III. ARCHITETURE OF THE PROPOSED MODEL

Figure 1, represents the architecture diagram of our proposed model. It shows all the modules and basic steps of our model. Some of the prominent steps include Data collection, its cleaning i.e. De noise, Feature extraction, LSTM training of the model and its prediction. Lets have laconic description of all below.

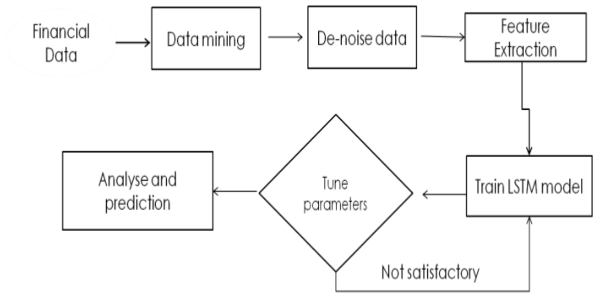
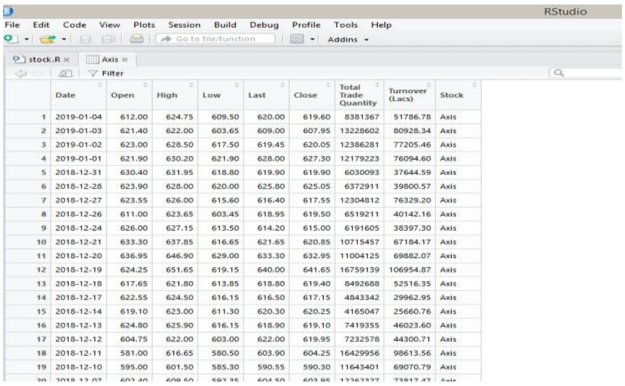


Figure 1Architectur diagram

A. Data Collection

The data for this paper is collected from national stock exchange. The historical data is present on the Quandl website from where it is transferred into Rstudio for calculation. The data obtained has 8 attributes Date, Open, High, Low, Last , Close, Traded quantity and turnover.


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
stock.R
Axis
Filter

	Date	Open	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)	Stock
1	2019-01-04	612.00	624.75	609.50	620.00	619.60	8381367	51786.78	Axis
2	2019-01-03	621.40	622.00	603.65	609.00	607.95	13228602	80928.34	Axis
3	2019-01-02	623.00	628.50	617.50	619.45	620.05	12386281	77205.46	Axis
4	2019-01-01	621.90	630.20	621.90	628.00	627.30	12179223	76094.60	Axis
5	2018-12-31	630.40	631.95	618.80	619.90	619.90	6030093	37644.59	Axis
6	2018-12-28	623.90	628.00	620.00	625.80	625.05	6372911	39800.57	Axis
7	2018-12-27	623.55	626.00	615.60	616.40	617.55	12304812	76329.20	Axis
8	2018-12-26	611.00	623.65	603.45	618.95	619.50	6519211	40142.16	Axis
9	2018-12-24	626.00	627.15	613.50	614.20	615.00	6191605	38397.30	Axis
10	2018-12-21	633.30	637.85	616.65	621.65	620.85	10715457	67184.17	Axis
11	2018-12-20	636.95	646.90	629.00	633.30	632.95	11004125	69882.07	Axis
12	2018-12-19	624.25	651.65	619.15	640.00	641.65	16759139	106954.87	Axis
13	2018-12-18	617.65	621.80	613.85	618.80	619.40	8492688	52516.35	Axis
14	2018-12-17	622.55	624.50	616.15	616.50	617.15	4843342	29962.95	Axis
15	2018-12-14	619.10	623.00	611.30	620.30	620.25	4165047	25660.76	Axis
16	2018-12-13	624.80	625.90	616.15	618.90	619.10	7419355	46023.60	Axis
17	2018-12-12	604.75	622.00	603.00	622.00	619.95	7232578	44300.71	Axis
18	2018-12-11	581.00	616.65	580.50	603.90	604.25	16429956	98613.56	Axis
19	2018-12-10	595.00	601.50	585.30	590.55	590.30	11643401	69070.79	Axis
20	2018-12-07	602.40	609.50	592.35	604.50	603.95	12262327	73817.47	Axis

Figure 2 Data set collected from NSE

It is a time series numeric data. The data set consist of nine columns namely date, open, low, high, last, close traded quantity and stock. The stock column is the ticker column which is used to identify the stock, date column represent the particular day, open represents the opening price of the stock, high represent the highest price of that particular stock on a particular day, last represent the last stock price while the close represent the closing price of the stock, total trade quantity represents the total number of trades that is the exchange of stocks by buying and selling done on the particular date. Our focus of interest will be the closing price and by considering that we will be predicting or estimating a closing price of the stock for the a period of time. Our data holds lot of value as it is a historical data which not only gives the present value of stock but also helps in predicting the future value by providing the past values of the stock, thereby helping us in determining the trends.

B. Some Common Mistakes

In cleaning the data, we perform de noising using wavelet transform. Wavelet transform functions well with time series data and gives a smooth curve as its resultant signal. An array of data is passed to the wavelet function where ‘haar ’ wavelet family is used.

In return we get the approximation and details coefficient on which threshold is applied to remove and reduce noise from the data. After applying threshold inverse wavelet function is used which combine the two coefficients to produce the original array without noise. Wavelet transform removes outliers rapid fluctuating signals and white noise from the data. The thresholding mechanism chosen here is soft threshold which not only smoothens the data but gives a continuous result.

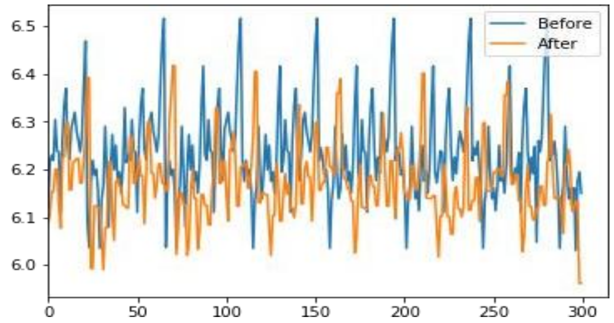


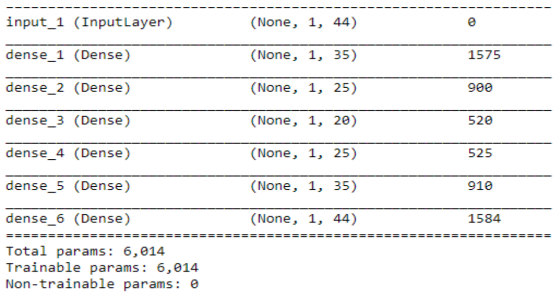
Figure 3 De noise using wavelet transform

This journal uses double-blind review process, which means that both the reviewer (s) and author (s) identities concealed from the reviewers, and vice versa, throughout the review process. All submitted manuscripts are reviewed by three reviewer one from India and rest two from overseas. There should be proper comments of the reviewers for the purpose of acceptance/ rejection. There should be minimum 01 to 02 week time window for it. The graph above represents the data before and after data cleaning process. The signals in blue represents the data value before performing the wavelet transformation while the orange signals represents the data after removing noise. Orange signals are de noised data. The wavelet transforms sets the threshold and any signals falling beyond the range is removed. The transform function removes all the noise and outliers from the signal. As we can see the four fluctuating signals were removed and included only the part which falls under the range. Hence, the new de noised data is less clumsy and easily interpretable. The resultant signal is free from noise and other fluctuating signals which may affect the performance of prediction model.

C. Feature Selection by using Auto Encoders

It plays vital role in prediction of stock data as it ensures high or improved accuracy by using more relevant and less redundant data features from the feature space. We made use of auto encoders for performing feature selection because the model has the ability to self-learn and provide feature by encoding the relevant ones. Here six hidden layers are used which has three encoders and three decoders in it. Auto encoders are different in the sense that the output expected is same as the input for which we can test the encoder on different sets of data and find the best features. First the structure of the model is defined and then create layers using dense function form the keras library. The model is then compiled and the data is trained on it. The encoded data efficiency is tested on the test data by measuring the error rate.

We can observe here that there are six layers of encoding and decoding are formed using the function dense. The first three encoding layers are given the input of their own previous layer and in the same way we structure the decoding layer. Later a compile function is used to intact these layers together and to model them to work together.

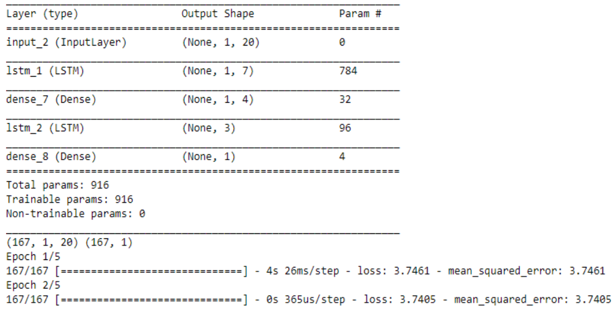


```
_________________________________________________________________
input_1 (InputLayer)         (None, 1, 44)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1, 35)             1575
_________________________________________________________________
dense_2 (Dense)              (None, 1, 25)             900
_________________________________________________________________
dense_3 (Dense)              (None, 1, 20)             520
_________________________________________________________________
dense_4 (Dense)              (None, 1, 25)             525
_________________________________________________________________
dense_5 (Dense)              (None, 1, 35)             910
_________________________________________________________________
dense_6 (Dense)              (None, 1, 44)             1584
=================================================================
Total params: 6,014
Trainable params: 6,014
Non-trainable params: 0
```

Figure 4 Structure of auto encoders

D. Training the LSTM Model

In this method we feed the auto encoded data into the neural network for prediction which consist of four hidden layers. The activation function used is ‘relu ‘ which we find suitable for our data. For analyzing efficiency we have used root mean square error.



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         (None, 1, 20)             0
_________________________________________________________________
lstm_1 (LSTM)                (None, 1, 7)              784
_________________________________________________________________
dense_7 (Dense)              (None, 1, 4)              32
_________________________________________________________________
lstm_2 (LSTM)                (None, 3)                 96
_________________________________________________________________
dense_8 (Dense)              (None, 1)                 4
=================================================================
Total params: 916
Trainable params: 916
Non-trainable params: 0
_________________________________________________________________
(167, 1, 20) (167, 1)
Epoch 1/5
167/167 [==============================] - 4s 26ms/step - loss: 3.7461 - mean_squared_error: 3.7461
Epoch 2/5
167/167 [==============================] - 0s 365us/step - loss: 3.7405 - mean_squared_error: 3.7405
```

Figure 5 LSTM layers

We train our train data set by feeding it into the LSTM model and check its error rate and later we test the remaining test data with the help of training data sets.

IV. RESULT AND ANALYSIS

The three layer architecture was trained and tested on data from Jan 2017 to Dec 2018. The dataset was divided into train and test data in appropriate ratio.

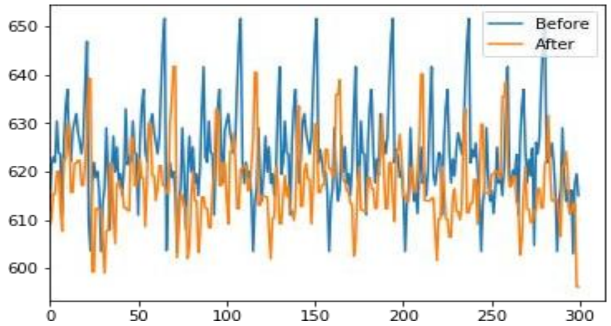


Figure 6 haar wavelet

Figure 6 represent the reduction in noise of the original data by using ‘haar’ wavelet family and Fig 7 represent by using ‘db2’ wavelet family. We can further check it through Table 1 which shows the standard deviation of the data before and after de-noising.

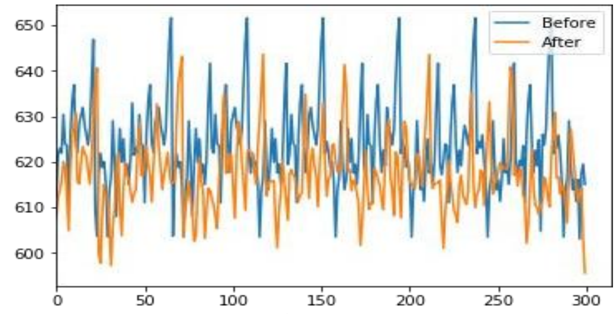


Figure 7

The figures are the resultant figures after removing noise from the data. The below graph shows the overall effect of de noising the data of axis bank. As we can see before de noising, the data values are very high which is represented through blue line, but once after successful completion of data cleaning, the value of data is decreased which clearly indicates that the signals are free from any additional unwanted signals i.e. noise. On following the pattern it is clearly observed that the data signals have reduced its value on an overall range. This depicts that the orange line is indicative of the data values which are free from noise, fluctuations, and outliers. So de noise, helped us in a positive way by eliminating unwanted signals which can alter the performance of our prediction model.

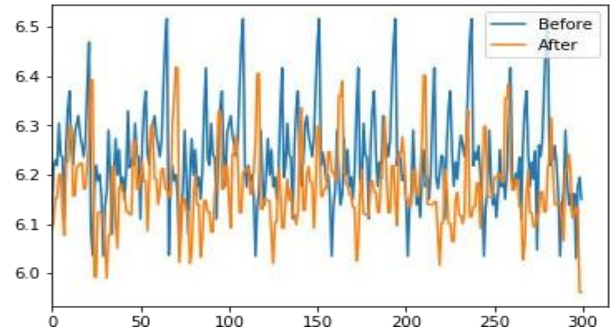


Figure 8 de noise over Axis Bank data

Table 1 Relating deviations of both family for de noise

Wavelet family	Bank	Old deviation	New deviation
haar	SBI	22.736	8.507
	PNB	39.874	3.859
	CANARA	49.502	5.461
	ICICI	25.626	4.836
	BOB	23.880	3.839
	Axis	44.859	14.005
db2	SBI	22.736	8.569
	PNB	39.874	3.849
	CANARA	49.502	5.290
	ICICI	25.626	4.791
	BOB	23.880	3.776
	Axis	44.859	13.796

From Table 1, the reading shows that the deviations obtained using the db2 family of wavelet transform are a little less than the haar wavelet family method of de noising the data. As, it is evident that the deviation is better for db2 family but at the same time the stock data over which we are working is dynamic in nature that needs the signals which are the rescaled version of time data series and hence, this quality is available with the haar wavelet family, so for further de noising we selected this method as our operandi. The de-noised data is then divided into train and test data to perform encoding and then evaluate it. Fig 9 and Fig 10 represent the reduction in loss value with each successive epoch for the particular bank. Table 2 shows the loss value for each bank.

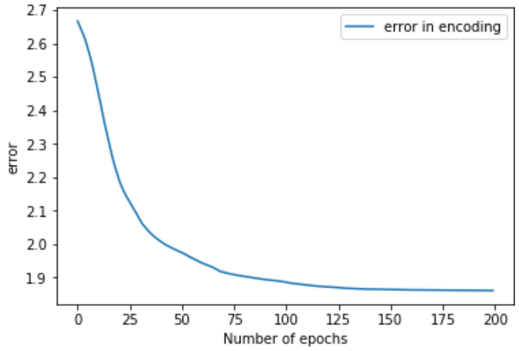


Figure 9 error rate axis bank

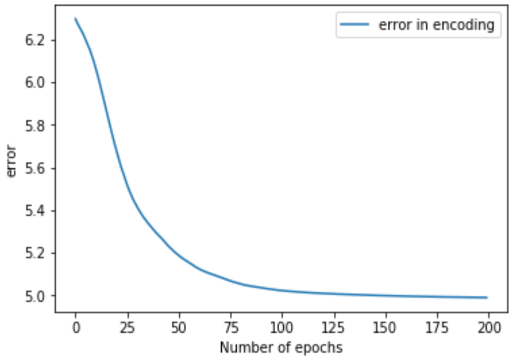


Figure 10 error rate PNB bank

Table 2

BANK	Loss value
SBI	1.003
CANARA	6.547
ICICI	2.976
BOB	7.218
PNB	4.046
Axis	2.517

It is quite evident that SBI has least loss value which can be attributed to its stable stock nature while BOB has highest loss value. Our target was to keep the loss value below 5 for better prediction.

From the above table it evident that the prediction is easy for stable bank stocks such as SBI, ICICI, Axis while more volatile stocks such as PNB has very high mean square error value. Also CANARA and BOB banks were moderate and there error can be further improved. Also from these graph we can predict that axis bank stocks are more profitable as the closing price of these stocks are not decreasing in its value rather it is increasing with time. Our model, can also work as a early recommender system for the stock market investors.

V. CONCLUSION

We were able to reduce the mean square error for different banks considerably with the help of de noising and auto encoding. But since stock market is too dynamic and depend on factors such as internal policy of the companies, market and political situation the prediction becomes too difficult. So we have tried to predict the movement of stock market data rather than predicting accurate values. The prediction can be further improved by changing the layers and parameters of auto encoders and LSTM layers. We can also try different wavelet family for de noising the data and change optimizer for better prediction.

ACKNOWLEDGMENT

It is optional. The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments. Avoid expressions such as “One of us (S.B.A.) would like to thank ... .” Instead, write “F. A. Author thanks ” *Sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page*.

REFERENCES

1. G Applying LSTM neural network for predicting stock closingprice, tingwei gao,yieting chai,yi lie, Journal of science and technology volume 68,jan8
2. Applied attention based LSTM neural networkin stock predictors, li chen cheng,yu hsing huang, mu enwu, Ieee explorer,10.01.11,jan14
3. Stock transaction prediction Modelling and Analysis based on LSTM, siyauan liu,guangz hong,yiftan ding, hubey laboratory of intelligent information processing and real time industrial system, Wuhan, China.
4. Prediction of Stock Market Price using Hybrid of Wavelet Transform and artificial Neural network, S.kumar, M.sumathi, S.N.sivananda,chennai.
5. Feature selection,rajajiram,Computer Science magazine,chapter .
6. Stock prediction using twitter sntiment analysis,Anshul Mittal,Arpit Goel,International conference on sience and technology,standford university,2012.
7. Predicting stock prices using data mining techniques,qasem a. al. radaideh,adel abu assaf,eman alnagi,the internatiomal arab conference on international technology,Dec 2013
8. Stock market prediction using artificial neural network,Bigul kutlun,Meltez Ozuran ,Bestan badus,white paper,istanbul,turkey
9. Stock market forecasting technique literature survey, international journal of computer science and mobile computing ,vol5,isue 6,june2016.
10. Using ai to make predictions on stock market ,alice zheng,jack jin,jornal of co,mputer science,2015,june.
11. Machine learning techniques and use of event information for stock market prediction: a survey and evaluation by Paul D. Yoo, Maria H. Kim,
12. Tony Jan 2007
13. A Multi agent approach to Q-learning for daily stock trading by Jae Won Lee, Jonghun Park, Member, IEEE, Jangmin O, Jongwoo Lee, and Euyseok Hong 2007
14. Intelligent Stock Trading System based on SVM Algorithm and Oscillation Box Prediction by Qinghua Wen, Zehong Yang, Yixu Song,
15. Peifa Jia
16. Forecasting Intraday Stock Price Trends with Text Mining Techniques by Mittermayer, University of Bern, Institute of Information Systems 2004.
17. Yin Hongcai, Zhao Chunyan. "Research on Stock Forecasting Based on Neural Network." Natural science journal of Harbin Normal University 23.3(2007):47-49.
18. Fenu, Gianni, and S. Surcis. "A Cloud Computing Based Real Time Financial System. " ACM Symposium on Applied Computing ACM, 2009:1219-1220.
19. Colah. Understanding LSTM Networks. http://colah.github.io/posts/ 2015- 08-Understanding- LSTMs/. 27.8(2015).
20. Gers F., A., J. Schmidhuber, and F. Cummins. "Learning to Forget: Continual Prediction with LSTM." Neural Computation 2.10(1999): 2451-71.
21. Sundermeyer, Martin, R. Schlüter, and H. Ney. "LSTM Neural Networks for Language Modeling." Interspeech 2012:601-608.

AUTHOR PROFILE

Dr. R. Kavitha received her Master’s in Software Engineering from College of Engineering , Anna University and Ph. D in Computer Science and Engineering from Vel Tech, Chennai. Her research areas are Data Mining, Image Processing and Software Engineering. Presently working as Associate Professor at Vel Tech, Chennai having 10 yrs of Teaching experience.