ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 11(47), DOI: 10.17485/ijst/2018/v11i47/138080, December 2018 ISSN (Online) : 0974-5645*

A Hybrid Regression and Deep Learning LSTM Based Technique for Predicting Volatility IndeX (VIX) Direction of Change (Trend)

P. Hemanth Kumar[1] and S. Basavaraj Patil[2]

[1]Department of Computer Science and Engineering, Visvesvaraya Technological University, Jnana Sangama, Machhe, Belgaum – 590018, Karnataka, India; hemanth00kumar@gmail.com

[2]AMC Engineering College, Bannerghatta Main Road, Kalkere, Bangalore – 560083, Karnataka, India; dr.sharan9@gmail.com

Abstract

Objectives: The main of study is to predict the volatility trend with high accuracy. The improved accuracy can help profitable trades. Outliers in dataset can reduce prediction accuracy; an innovative dynamic approach is proposed to filter outliers from the dataset. **Methods**: In this study, an intelligent methodology is proposed to predict volatility trend with hybrid regression based outlier removal technique and advanced deep learning Long Short-Term Memory (LSTM) techniques. Regression techniques are used for prediction and classification problems, the same principle is applied here for identifying the outliers in the data. The random data set is trained and tested multiple trials with regression technique. The algorithm generates standard errors, residual error and 'p' value for all the predictions, these values are compared with standard threshold across all the trials. The similar errors with multiple occurrences are identified as outliers and removed from the dataset. The LSTM techniques trained and tested with different epochs and network configuration till the prediction results improve. **Result:** This study uses India Volatility Index (VIX) data for predicting next day volatility. The results show significant improvement in accuracy with the proposed approach. The results in this paper demonstrate the LSTM techniques out performs regression, decision trees, random forest, SVM, boosting techniques and neural network based techniques. **Application:** The paper also shows that usage of regression techniques for removal of outlier further improves the forecasting accuracy by 10%. The proposed approach can be applied for forex and option trading. The proposed approach can also used in other predictive modeling problems such as CRM, Healthcare, Credit risk and Auto insurance.

Keywords: Deep Learning, learning Long Short-Term Memory (LSTM), Prediction, Regression, Trend, Volatility

1. Introduction

The direction of the Volatility trend refers to movement of volatility or the risk fluctuation in a stock market in the future. Forecasting the trend is a practical issue that intensely impacts a financial trader's decision to buy or sell a stock. Precise forecasts of the patterns of the volatility trend can assist traders acquiring wealth and benefits in the stock and currency trade. Henceforth, exact estimating of the patterns of the volatility trend can be greatly invaluable for financial traders.[1] Many have perception that trading could be made profitable by a precise direction of movement of volatility in the stock list.[2] Their work recommended that financial forecasters and merchants should center on the precisely forecasting the volatility direction and limit the estimates' deviations from the real observed values. Additionally few researchers trust that exact expectations of the direction of volatility are critical for financial specialists.[3] In any case, the behavior of securities exchanges relies upon numerous subjective factors, for example, political, economic, and normal factors, among numerous others. The securities exchanges are dynamic and display wide variety, and the forecast of financial markets in this way turns into an exceptionally difficult due to the profoundly non-direct nature and complex dimensionality.[4,5]

*Author for correspondence

2. Background

Amid the 1990s, analysts made vital strides forward in forecasting sequences with neural networks. Few researchers identified some of the key scientific difficulties in forecasting long sequences.[6] As solution to these problems others presented the long short term memory or LSTM network to solve a portion of these difficulties.[7] LSTM is generally utilized for some, sequence modeling assignments, including natural language and dialect handling tasks at Google. The second wave of neural networks investigates endured until the mid-1990s. Once more, there was a decrease in deep learning research, because of unreasonable aggressive tasks that the current frameworks should solve. As of right now, deep networks were for the most part accepted to be extremely difficult to train. Rather those calculations, existing since the 1980s, work quite well, however this was not discovered around 2006. The issue is maybe essentially that these calculations were too computationally exorbitant to permit experimentation with the equipment accessible at the time.

An essential benefit of recurrent networks is their capacity to utilize relevant data when mapping between input and output data sequences. Tragically, for standard Recurrent Neural Network (RNN) structure, the scope of setting that can be accessed is constrained.[8] The issue is that the influence of a given contribution on the hidden layer, and in this manner on the network output, either decays or explodes exponentially as it cycles around the network's recurrent connections. Practically speaking this weakness alluded to in the writing as the vanishing gradient problem makes it hard for a RNN to learn tasks containing delays of more than around 10 time steps between pertinent input and target events.[9]

Over the previous decade, LSTM has demonstrated fruitful at a scope of synthetics tasks requiring long range memory, including learning setting free dialects,[10] reviewing high accuracy genuine numbers over broadened noisy sequences and different tasks requiring exact planning and checking.[11] In specific, it has tackled a few artificial issues that stay incomprehensible with some other RNN engineering. Various experiments were made in the 1990s to address the issue of vanishing gradients for RNNs. This included non-gradient based training techniques, for example, simulated annealing and discrete error propagation,[12] and unequivocally presented time delays[13] or time constants,[14] and progressive sequence compression. In any case, the most effective arrangement so far is the (LSTM) technique.

Moreover, LSTM has been connected to different genuine issues, for example, protein optional structure forecast,[15] musical generation, reinforcement learning[16] and discourse acknowledgment[17,18] and handwriting recognition.[19] As would be normal, its favorable circumstances are most articulated for issues requiring the utilization of long range logical data.

3. VIX Data

India VIX is a volatility index based on options costs of National Stock Exchange Fifty (NIFTY). India VIX is calculated utilizing the best bid and ask prices from the out-of-the-money near to and mid of the month Nifty options contracts which are exchanged on the F&O section of NSE. India VIX shows the speculator's impression of the market's volatility in the near term. The list portrays the normal market volatility throughout the following 30 calendar days. Higher the India VIX values, higher the expected volatility and the other way around.

3.1 India VIX Calculation System

India VIX utilizes the calculation system of Chicago Board Options Exchange (CBOE), with reasonable changes to adjust to the NIFTY alternatives arrange book utilizing cubic splines, and so forth.

The variables considered in the calculation of India VIX are specified beneath:

1. Time to expiry:
   The opportunity to expiry is processed in minutes rather than days with a specific end goal to touch base at a level of accuracy expected by proficient dealers.
2. Interest Rate:
   The pertinent residency NSE MIBOR rate (i.e 30 days or 90 days) is being considered as risk free interest rate for the individual expiry long periods of the NIFTY options contracts
3. The forward index level:
   India VIX is registered utilizing out-of-the-money option contracts. Out-of-the-money option contracts are distinguished utilizing forward index level. The forward record level helps in deciding the At-The-Money (ATM) strike which thus helps in choosing the options contracts which will be utilized

for registering India VIX. The forward index level is taken as the most recent accessible cost of NIFTY future contract for the individual expiry month. "VIX" is a trademark of ("CBOE") and Standard and Poor's has conceded a permit to National Stock Exchange (NSE), with authorization from CBOE, to utilize such check for the sake of the India VIX and for purposes identifying with the India VIX.

4. Bid-Ask Quotes:
The strike price of NIFTY option contract accessible just underneath the forward index level is taken as the ATM strike. NIFTY option call contracts with strike cost above the ATM strike and NIFTY option Put contracts with strike cost underneath the ATM strike are distinguished as out-of-the-money alternatives and best offer and solicit cites from such option contracts are utilized for calculation of India VIX. In regard of strikes for which proper statements are not accessible, values are touched base through interjection utilizing a measurable technique to be specific "Natural Cubic Spline" After distinguishing proof of the statements, the change (volatility squared) is figured independently for close and mid month expiry. The difference is figured by giving weight ages to every one of the NIFTY option contracts distinguished for the calculation, as per the CBOE technique. The weight age of a solitary option contract is specifically relative to the normal of best offer solicit cites from the option contract and conversely relative to the options contract's strike cost

3.2 Calculation of India VIX

The variances for the near and mid-month expiry registered independently are inserted to get a solitary variance an incentive with a consistent development of 30 days to termination. The square base of the registered fluctuation esteem is duplicated by 100 to touch base at the India VIX value. For additionally points of interest please refer to the white paper with the nitty gritty philosophy on calculation of India VIX [source NSE]

4. Deep learning

The key aspects that deep learning techniques so popular are the dataset size, model size, task complexity, learning algorithm and performance. Deep learning can also be defined as class of machine learning that has cascade of multiple layers of nonlinear processing units for data preprocessing steps such as feature extraction and transformation. All layers are connected and layer output of one unit is input to other layer. Deep learning technique mainly used for pattern recognition and they learn at multiple levels.

Most of the deep learning techniques are based on artificial neural network, also with few changes in the architecture. In deep learning each layer transforms the data into more abstract and learns from it. The deep word in deep learning refers to the number of layers deep through which the data undergoes transformation.

Initially deep learning was used only for image segmentation and image recognition problems, currently they are used in all the fields.

5. LSTM

A Recurrent neural network is said to store data in a combination of long-term and short-term memory. The short-term memory is framed by the activation of units, containing the current history of the network. The long-term memory is rather framed by the gradually changing weights of the unit transitions that are holding background based data about the framework. Long Short-Term Memory is an endeavor to broaden the time that a RNN can hold imperative data.

Rather than the hidden nodes in a customary RNN, a LSTM RNN makes utilization of memory units. The memory units are recurrently connected units that in themselves hold a system of units. Inside these memory units is the place the answer for the vanishing gradient issue exists. The memory units are made of a memory cell, an input gate, an output gate and a forget gate as shown in the Figure 1. The memory cell is the plain center of the memory unit, containing the data. To be capable to protect its state when no other information is available the memory cell has a self-recurrent connection. The forget gate watches this self-recurrent connection. Along these lines it very well utilized to adaptively dispose the cell state when it has turned out to be old. This is not just vital to keep the system data up and coming, yet in addition in light of the fact that not resetting the cell states can in some occasions, with consistent inputs, influence them to develop indefinitely. This would invalidate the point of LSTM.[20] The input gate figures out what data to store in the cell, that is, shields the cell from undesirable inputs. The output gate, then again, chooses what data should flow out of the from the memory cell and subsequently denies undesirable flow of data in the network.

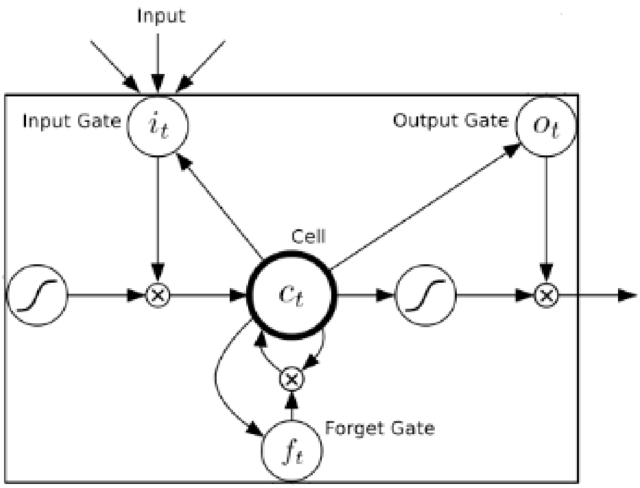

Figure 1. A LSTM memory block.

The cell's self-recurrent weight and the gating units develop by and large a steady errorflow through the cell. This errorflow is alluded to as Constant Error Carousel (CEC).[21] The CEC is the thing that makes LSTM networks ready to connect contributions to yields with in excess of 1000 time ventures in the middle them and in this way expanding the long range memory limit by a hundredfold contrasted with traditional RNNs. Approaching this long history of data is additionally the specific reason that LSTM networks can take care of issues that prior was incomprehensible with RNNs. This is today still the most effective answer for the issues with exploding and vanishing gradients.

6. Regression Model for Outlier Removal

Regression is statistical technique popular for solving prediction and classification problems. It is supervised machine learning technique and needs a target variable to learn from the data. The regression technique computes the statistical relationship between dependent (target) variable and independent variable. The general structure of regression model as:[22]

$$y = \beta 0 + \beta_1 x_1 + \beta_1 x_2 + \ldots. + \beta_k x_1 + \varepsilon \ (1)$$

Where:

y is the dependent (target) variable

$x_1, x_2, \ldots, x_k$ are the independent variables

β_i coefficients, its responsibility denote relationship with variable x_i

The estimation of the coefficient β_i chooses the dedication of the autonomous variable x_i and β_0 is before long the y-intercept. The coefficients $\beta_0, \beta_1 \ldots \beta_k$ are normally unknown and they are calculated by regression process.

The imaginative regression construct exception procedures works with respect to the rule that, regression being a best fit method will fit the correct information and create redress results. Though the outliers does not fit in the condition or regression bend and deliver errors. This guideline is accustomed to distinguishing and evacuating outliers. The stream chart of creative anomaly method is spoken to in figure.

The regression based outlier removing procedure block diagram is shown in Figure 2 and modules discussed beneath:

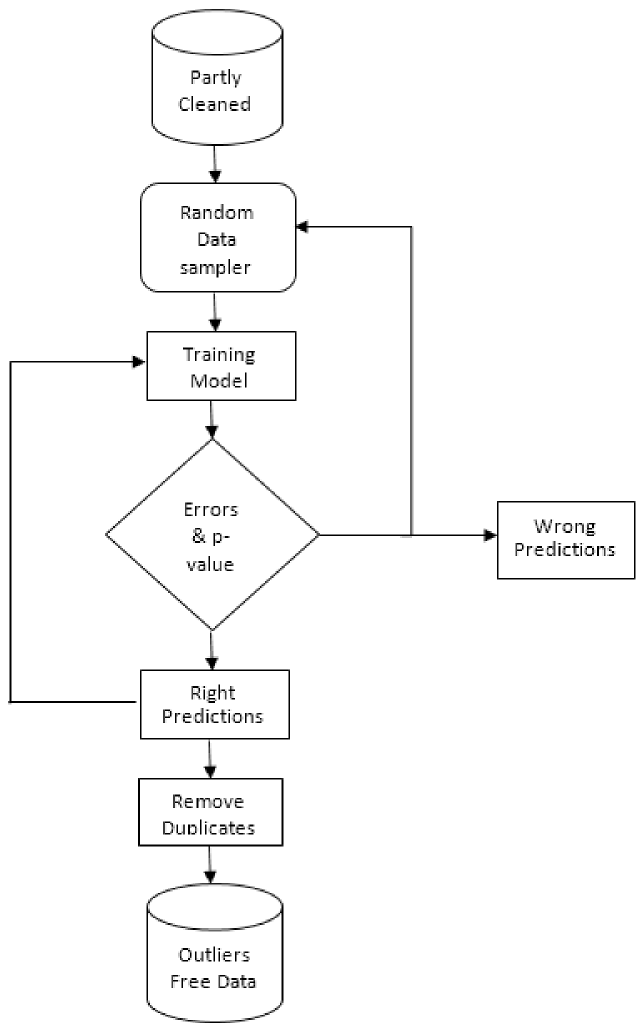

Figure 2. A Regression based outlier removal technique.

1. *Partly cleaned data:* Partly cleaned data refers to the dataset treated with missing values. The VIX data is checked for any missing or incomplete data generated while downloading or while uploading the data in website.
2. *Random data sampler:* This module is utilized to produce random samples of partly cleaned data. The output of this data is utilized for preparing training and test dataset. In regression based anomaly removal approach we may need to iterate the outliers separating numerous times till target thresholds are achieved. Each time we repeat ie prepare and predict with regression model, the data needs to be diverse samples. Consequently we are utilizing a random sample which rearranges the data and split them into 50:50 proportions each time data goes through this step. Each time the model trains with different samples of data, same with test dataset, each time data available for predictions are different datasets.
3. *Training Model:* The model is trained with regression technique utilizing the training dataset. The model built utilizing the training data set is utilized to predict the results with test dataset. The actual results and forecasted results are compared and right forecasts and wrong forecasts are put away independently. The procedure is repeated with various mixes of train and test dataset generated by random sampler. All the correct expectations and wrong forecasts are saved each time the model repeats.
4. *Errors and p-esteem:* In this step, measuring the forecasts and model accuracy are computed. The types of errors in forecasting problems are standard errors and residual error and 'p' value. In regression 'p' value refers to the accuracy of the model. The number of train and test iterations carried out is decided based on the errors. Each time the model predicts the result it also calculates standard errors, residual error and 'p' value; these error values should to be much less than standard thresholds. The iterations procedure is carried until the point when errors and 'p' value stays consistent.
5. *Right Predictions:* Every time the model is forecasts the outcome; these values will be contrasted and actual values. On the off chance that both the values are same they are considered as right predictions and they are saved. Same process is repeated for numerous cycles, each execution creates a specific measure of right predictions and these are saved with previous right predictions.
6. *Wrong Predictions*: If the actual and forecasted results are different, those results are named as wrong forecasts. Certain measure of wrong predictions are produced in every iterations, all the wrong forecasts are saved together.
7. *Remove duplicates*: In this step the proposed methodology expel duplicates data from the right forecasts. As the model is prepared with random data samples in each emphasis there is high likelihood of duplicates in dataset, to evacuate such qualities the algorithm search for duplicates and remove.
8. *Outlier free data*: During our tests it took 15 to 20 iterations to reduce errors and stabilize 'p' value. The data saved as right forecasts contains the qualities whose model errors are insignificant and the predictions match with actual. The rest of the data set after removal of duplicates can be referred as outliers free data as it fulfilled the prerequisites.

7. Methodology

The Methodology of regression and deep learning LSTM techniques based VIX trend forecasting is shown in the Figure 3. It consists of database containing VIX data, followed a preprocessing block to remove outliers and other missing and noisy dataset. From the cleaned data we compute Trend, the target variable from VIX data. Next, the methodology transforms data into a format allowing most recent data to have high impact on target variable; the last step is training the model with deep learning LSTM techniques and forecasting down trend and uptrend accuracy.

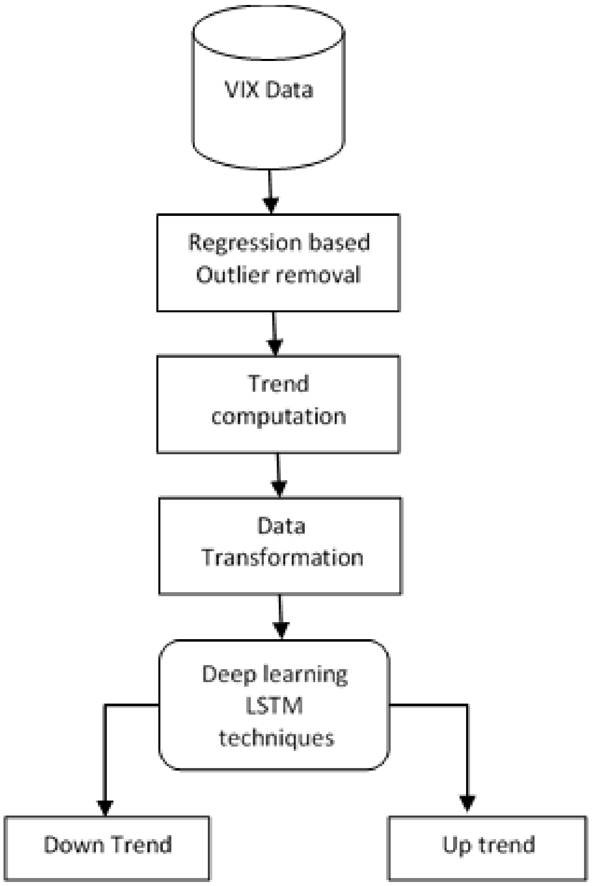

Figure 3. **Methodology of regression and deep learning LSTM technique.**

8. Data

Volatility Index 10 years data downloaded from National Stock Exchange (NSE) data ranging from 1st January 2008 till 2nd February. 2018. The data contains open, high, low close, prev close, change and % change columns as shown in Table 1, the closing prices are plot is depicted as Figure 4. This paper is focused on closing prices, trend calculation are made using the close value.

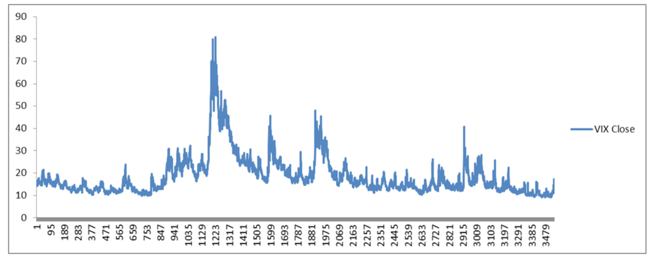

Figure 4. **Plot of closing prices of VIX.**

9. Trend Calculation

The trend indicates the direction of flow of curves. Here the trends are calculated based on difference of closing prices, yesterday closing price Close(i-1) and today's closing price Close(i). The trend is estimated as 0 if the difference value is negative and if positive then trend is 1.

In this section we calculated the trends from the closing prices using below formula

Trend = 0 if (Close(i-1) – Close(i))< 0 is downward trend (2)

Trend = 1 is (Close(i-1) – Close(i))> 0 is upward trend (3)

10. Data Transformation

The data is transformed in a matrix format with closing prices on rows, 25 values in every row. The first value in row 1 become second value on day 2, it carries on. This method is followed to predict the trend of VIX with 25 previous historical dataset. The sample data arrangement is shown in Table 2. With this type of unique arrangement, the trend depends on past 25 data, this method will be more error proof as trend depends on 25 values rather than just depending on just today's and yesterday's value. We may have increased the dependency of trend on 25 values but their need for capturing the features or patterns that exists in 25 values that make impact on the trend.

Table 1. Representation of sample VIX data

Date	Open	High	Low	Close	Prev. Close	Change	% Change
02-Jul-18	12.9375	13.9425	12.465	13.37	12.9375	0.43	3.34
03-Jul-18	13.37	13.7525	12.1875	13	13.37	-0.37	-2.77
04-Jul-18	13	13.34	12.575	12.6575	13	-0.34	-2.63
05-Jul-18	12.6575	12.9525	11.225	12.575	12.6575	-0.08	-0.65
06-Jul-18	12.575	12.68	11.6675	12.4425	12.575	-0.13	-1.05

Table 2. Table representation of data arrangement

Day 1	Day 2	…	…	…	…	Day 24	Day 25	Trend
12.94	13.94					12.47	13.37	1
13.37	13.75					12.19	13.00	0
13.00	13.34					12.58	12.66	1
12.66	12.95					11.23	12.58	1
12.58	12.68					11.67	12.44	0
12.44	12.58					11.22	12.39	0

11. Results

We trained the model with LSTM network 3 layers, input, hidden and output layer. Input layer contained 25 input nodes, 128 hidden nodes and 1 output node. The experimental results were recorded at epochs ranging from 250 to 1500 at interval of 250 epochs. The results are tabulated in the Table 3. From the Table 3. A result, The LSTM technique with regression based outlier approach has achieved an accuracy of 71% for downtrend, 77% for uptrend and overall accuracy of 74%. The maximum accuracy was achieved by LSTM technique with an epoch of 1500. In the Table 4. The results are tabulated without using the regression based outlier approach. It can be observed that the maximum accuracy are obtained by LSTM with 500 epochs, the down trend accuracy is 53%, uptrend accuracy of 69% and overall accuracy of 61%. The comparison of results with proposed approach regression with LSTM approach and forecast only with LSTM approach are compared in Table 5. An LSTM technique shows an accuracy increase by 14% on average taken in 6 attempts with varying epochs with proposed approach. The highest increase of accuracy being 19% with 1500 epochs. In one more set of experiments, this paper compares the performance of deep learning LSTM technique with other machine learning techniques. The experiments with regression based outlier removal data using machine learning technique such as decision trees, regression, random forest, SVM, boosting techniques and neural networks. The results tabulated in Table 6 shows that deep learning LSTM technique has outperformed all other technique.

Table 3. LSTM techniques results with regression and deep learning technique

LSTM Technique	Down Trend Accuracy	Up Trend Accuracy	Overall Accuracy
epochs = 250	71.87%	72.80%	72.34%
epochs = 500	68.31%	74.94%	71.70%
epochs = 750	69.95%	73.06%	71.57%
epochs = 1000	70.60%	72.05%	71.32%
epochs = 1250	60.56%	78.23%	69.42%
epochs = 1500	**70.80%**	**76.56%**	**73.73%**

Table 4. LSTM results on forecasting volatility trend

LSTM Technique	Down Trend Accuracy	Up Trend Accuracy	Overall Accuracy
epochs = 250	55.60%	64.22%	60.06%
epochs = 500	**53.37%**	**69.04%**	**61.10%**
epochs = 750	52.30%	59.59%	55.98%
epochs = 1000	55.89%	58.38%	57.12%
epochs = 1250	59.84%	55.48%	57.50%
epochs = 1500	57.36%	52.86%	55.12%

Table 5. LSTM techniques results comparison traditional with innovative approach

LSTM Technique	Traditional Approach	Innovative Approach	Increased Accuracy
LSTM epochs = 250	60.06%	72.34%	12.28%
LSTM epochs = 500	61.10%	71.70%	10.60%
LSTM epochs = 750	55.98%	71.57%	15.60%
LSTM epochs = 1000	57.12%	71.32%	14.20%
LSTM epochs = 1250	57.50%	69.42%	11.92%
LSTM epochs = 1500	55.12%	73.73%	18.61%
	Average increase		**13.87%**

Table 6. Comparison of results with machine learning techniques

ML Technique	Down Trend Accuracy	Up Trend Accuracy	Overall Accuracy
Decision Trees	71.20%	64.41%	67.64%
Random Forest	68.40%	72.00%	70.20%
Gradient Boost	69.13%	71.09%	70.11%
Support Vector Machines	70.80%	72.00%	71.40%
Regression	67.47%	79.42%	73.73%
Neural Network	69.60%	72.88%	71.32%
LSTM Technique	70.80% 76.56%		73.73%

12. Summary

In this study a regression based outlier removal approach was suggested to improve the forecasting accuracy in VIX data. The forecasts were implemented with deep learning LSTM techniques. From the results it was observed that there is improvement in the accuracy with application of regression based outlier technique. The experiments show that there is improvement of 14% overall average accuracy recorded in 6 LSTM experiments with different epochs. The deep learning LSTM technique performance was compared with several machine learning techniques such as regression, decision trees, random forest, boosting techniques and neural networks. The LSTM techniques outperformed all the listed machine learning techniques.

13. Conclusion and Future Scope

The main contribution of this paper is application of deep learning based machine learning techniques for forecasting volatility index trend. Another important contribution is usage of regression as preprocessing technique for removal of outliers, missing and noisy data. The study also contributes in a unique arrangement of data in the form of matrix with 25 most recent data. Trend is calculated from the varying prices and forecasted a day ahead. The study also proposes a unique approach to forecast uptrend and downtrend of VIX with their accuracy. Here it is also proven that improvement dataset by handling outlier and wrong data; there can be increase in the accuracy. This study also tests the performance of deep learning techniques with other machine learning languages and it once again proven that deep learning are most advanced and accurate techniques currently available for forecasting, regression and classification problems. With the proposed approach a balanced accuracy is obtained with respect to downtrends and uptrends instead of biasing with either of the trends. The future work that can be carried out in this research is application hybrid methodology at every step of forecasting. This study explores application regression for preprocessing; many such new hybrid innovations can improve the accuracy. For this analysis we have considered VIX data and handled outliers, there are many other unknown factors that have influence in accuracy, future work can explore the unknown events. In this paper, deep learning based LSTM (recurrent) approach is used to forecast accuracy, with development in technologies currently there exists more advanced and complex techniques such as CNN, back propagation, feed forward, gradient descent etc.

14. References

1. Gholamiangonabadi D, Mohseni Taheri SD, Mohammadi A, Menhaj MB. Investigating the performance of technical indicators in electrical industry in Tehran's Stock Exchange using hybrid methods of SRA, PCA and Neural Networks. 5th Conference Thermal Power Plants (CTPP). 2014, pp.75-82. https://doi.org/10.1109/CTPP.2014.7040698.
2. Leung MT, Daouk H, Chen A. Forecasting stock indices: a comparison of classification and level estimation models. International Journal Forecasting. 2000, 16(2), pp. 173-190. https://doi.org/10.1016/S0169-2070(99)00048-5.
3. Mostafa MM. Forecasting stock exchange movements using neural networks: Empirical evidence from Kuwait. Expert System Applications. 2010, 37(9), pp. 6302-6309. https://doi.org/10.1016/j.eswa.2010.02.091.
4. Guresen E, Kayakutlu G, Daim TU. Using artificial neural network models in stock market index prediction. Expert System Applications. 2011, 38(8), pp. 89-97. https://doi.org/10.1016/j.eswa.2011.02.068.
5. Lee T, Chiu C. Neural network forecasting of an opening cash price index. International Journal System Science. 2002, 33 (3), pp. 229-237. https://doi.org/10.1080/00207720110092216.
6. Untersuchungen zu dynamischen neuronalen Netzen. https://www.bioinf.jku.at/publications/older/3804.pdf. Date accessed: 15/06/1991.
7. Hochreiter S, Bengio Y, Frasconi P,Schmidhuber J. Gradient Flow in Recurrent Nets: the Diculty of Learning Long-term Dependencies. A Field Guide to Dynamical Recurrent Neural Networks. 2001, pp. 1-464.
8. Schmidhuber J. Learning Complex Extended Sequences using the principle of history compression. Neural Computation.1992, 4(2), pp. 234-242. https://doi.org/10.1162/neco.1992.4.2.234.
9. Schmidhuber J, Wierstra D, Gagliolo M, Gomez F. Training Recurrent Networks by Evolino. Neural Computation. 2007, 19(3), pp. 757-779. https://doi.org/10.1162/neco.2007.19.3.757. PMid: 17298232.
10. Gers F, Schraudolph N, Schmidhuber J. Learning Precise Timing with LSTM Recurrent Networks. Journal of Machine Learning Research.2002, 3, pp. 115-143.
11. Bengio Y, Frasconi P, Simard P. The problem of learning long term dependencies in recurrent networks. In IEEE International Conference on Neural Networks.1993, pp. 1183-1195. https://doi.org/10.1109/ICNN.1993.298725.
12. Lang J, Waibel AH, Hinton GE. A Time-delay Neural Network Architecture for Isolated Word Recognition.

Neural Networks. 1990, 3(1), 23-43. https://doi.org/10.1016/0893-6080(90)90044-L.

13. Lin T, Horne BG, Tino P, Giles CL. Learning Long-Term Dependencies in NARX Recurrent Neural Networks. IEEE Transactions on Neural Networks.1996, 7(6), pp. 1329-1338. https://doi.org/10.1109/72.548162. PMid: 18263528.
14. Mozer MC. Induction of Multiscale Temporal Structure. International Conferenceon Neural Information Processing Systems.1991, 4, pp. 275-282.
15. Bakker B. Reinforcement Learning with Long Short-Term Memory. International Conference on Neural Information Processing Systems. 2002, pp. 1-14. PMid: 12241036.
16. Graves J. Schmidhuber, Frame wise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. Neural Networks. 2005, 18(5-6), pp. 602-610. https://doi.org/10.1016/j.neunet.2005.06.042. PMid: 16112549.
17. Gerstner W. Long Short-Term Memory in Recurrent Neural Networks. Ecole Polytechnique Federale de Lausanne. 2001, pp. 1-102.
18. Hochreiter S. Long short-term memory. Neural computation. 1997, 9 (8), pp. 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735. PMid: 9377276.
19. Chen J, Chaudhari NS. Capturing Long-term Dependencies for Protein Secondary Structure prediction. Advances in Neural Networks ISNN 2004, International Symposium on Neural Networks.2004, pp. 494-500. https://doi.org/10.1007/978-3-540-28648-6_79.
20. Liwicki M, Graves A, Fernandez S, Bunke H, Schmidhuber J. A Novel Approach to On-Line Handwriting Recognition Based on Bidirectional Long Short-Term Memory Networks. In Proceedings of the 9th International Conference on Document Analysis and Recognition. 2007, pp. 1-5.
21. Lefebvre G, Berlemont S, MamaletF, Christophe G. Inertial Gesture Recognition with BLSTM-RNN. Artificial Neural Networks.2015, pp. 393-410. https://doi.org/10.1007/978-3-319-09903-3_19.
22. Neil A. Weiss N, Hassett, M. Introductory Statistics. Addison-Wesley publication. Biometrical Journal. 1982, 8, pp. 1-651.