



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2020 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 15th Sept 2020. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-09/ISSUE-09](http://www.ijiemr.org/downloads.php?vol=Volume-09/ISSUE-09)

DOI: 10.48047/IJIEMR/V09/I09/38

Title **A LSTM APPROACH FOR MODELING AND FORECASTING FINANCIAL TIME SERIES DATA**

Volume 09, Issue 09, Pages: 268-274

Paper Authors

K. Sai Teja Reddy, Dr. M. V. Rathnamma



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code



A LSTM APPROACH FOR MODELING AND FORECASTING FINANCIAL TIME SERIES DATA

¹K. Sai Teja Reddy, ²Dr. M. V. Rathnamma

¹Student, ²Associate Professor

¹National Institute of Foundry and Forge Technology, ²KSRM College of Engineering

Email: k.saitejareddy2001@gmail.com

ABSTRACT: Financial expectation of time series information is a most difficult work in present age, as it is influenced by a few socio economical elements, dubious conduct of stake holders, worldwide monetary situation, etc. Long Short Term Memory (LSTM) is a high level strategy belongs to Recurrent Neural Network (RNN) classification. It has shown set up execution across different datamining issue, for example, determining of financial time series information. In this article we have done near examination on conventional and LSTM-based determining model to investigate the ability of LSTM in anticipating such time series. ML and Deep Learning are extremely well known in conjecture demonstrating. A genuine Datasets of a few organizations or markets were picked to pass on the productivity of ML model for improving guaging accuracy.

Keywords: Financial time series, Recurrent Neural Network, Long Short Term Memory.

I. INTRODUCTION

Sequence of mathematical information is known as time series information. Forecasting of time series information is a prediction of any occasion by examining the chronicled information gathered from different sources. Forecasting of time series is perplexing errand on the grounds that questionable changes in monetary condition in one hand just as deficient data then again . Market instability as of late has presented genuine worries for monetary and financial time series forecasting. Thusly, surveying the accuracy of gauges is important while utilizing different types of forecasting techniques, and all the more specifically forecasting utilizing regression investigation as they have a few constraints in applications. The fundamental target of this article is to research which forecasting strategies offer best predictions as for lower gauge errors and higher accuracy of figures. In such manner, there are assortments of stochastic models in time series forecasting. The most notable technique is

univariate "Auto-Regressive Moving Average (ARMA)" for a solitary time frame series information in which Auto-Regressive (AR) and Moving Average (MA) models are joined. Univariate "Auto-Regressive Integrated Moving Average (ARIMA)" is a

Uncommon kind of ARIMA where separating is considered in the model. Multivariate ARIMA models and Vector Auto-Regression (VAR) models are the other most mainstream forecasting models, which thusly, sum up the uni-variate ARIMA models and uni-variate auto-regressive (AR), model by taking into consideration more than one advancing variable.

AI procedures and all the more critically deep learning calculations have acquainted new methodologies with prediction issues where the connections between factors are displayed in a deep and layered chain of importance. AI based methods like SVM, Decision Tree, Multi Layer Perceptron

and Random Forests (RF) and deep learning-based calculations, for example, RNN, and LSTM have acquired bunches of considerations lately with their applications in many disciplines including finance. Deep learning techniques are fit for distinguishing construction and example of information, for example, non-linearity and intricacy in time series forecasting. Looking at customary and progressed AI model LSTM out performs conventional models. To ascertain execution of a model by utilizing different factual measurements like Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) applying on various models to get error esteem which determine execution of the classifier. Negligible error rate shows most elevated accuracy of the model.

II. LITERATURE STUDY

A few analysts proposed various models and examined the anticipated upsides of financial market.

In [1] Aakanksha Sharaff, Meenakshi Choudhary this paper includes the similar investigation of different stochastic models, for example, ARIMA model, ANN and RNN to anticipate the end stock files of S&P Bombay Stock Exchange Sensex in 2012.

In [2] Sima Siami-Namini, Neda Tavakoli and Akbar Siami Namini did experimentation and similar investigation on ARIMA and LSTM in Forecasting Time Series data. It was seen that the quantity of preparing times, known as Epoch in deep learning, had no impact on the presentation of the prepared figure model and it showed a genuinely irregular conduct.

In [3] Richa Handa, A.K. Shrivastava, H.S. Hota (2018) utilizing Financial Time Series Forecasting utilizing

Back Propagation NN and Deep Learning Architecture. In this examination they have utilized three time series information i.e. BSE30 stock information, INR/USD Foreign Exchange (FX) information and Crude Oil Data for prediction. Numerous direct and non-straight models have been created for these time series prediction.

In [4] Zhang Guohui (2017). Research on Time Series Prediction and Its Application Based on Deep Belief Network [DBN]. Harbin Institute of Technology. They directed experimentation on Western holds college to fabricate preparing information. A flaw conclusion technique consolidating DBN and LSTM was proposed in this paper utilizing LSTM network has great versatility to successive information.

In [5] Bin Gui, Xianghe Wei and Qiong Shen (2014) Financial Time Series Forecasting utilizing Support Vector Machine. This paper proposes Information Granulation and Support Vector Machine to anticipate stock list price. The anticipated outcome is more attainable and believable in activity than exclusively by SVM.

In [6] Yang Yujun, Yang Yimei, Li Jianping research on financial time series forecasting dependent on SVM. The Research on financial time series forecasting dependent on the help vector machine. Albeit the speed of prediction measure is slow, it can further develop the prediction accuracy of the financial time series. The trial results show the prediction accuracy of this methodology dependent on the help vector machine.

In [7] Hongbo Sun, Jing Xu (2018) in these paper further developed methodology for financial market forecasting dependent on fixed time is clarified. This paper moves toward a few gatherings of answer for financial information forecasting. On one hand, On the econometric investigation based

techniques propose a further developed vector auto regression model named A-VAR. This paper utilizes a few strategies for forecasting the fixed time series information and consequence of this investigation clarify various techniques in various datasets.

In [8] K.Kanchymalay , Roselina Sallehuddin, Naomie Salim ,Ummi Rabaah Hashim (2008) "Time series based forecasting for unrefined palm oil cost using neural network calculations" This investigation expects to introduce time series based forecasting for Malaysian rough palm oil cost. The fundamental point of this exploration is to conjecture the rough palm oil costs dependent on two significant indicators, to be specific the cost of soy bean oil and cash exchange rates. By and large outcomes show that an expansion in the quantity of info components would impact the accuracy and execution model's neural network.

In [9]Bo-Sheng Lin,Wei-Tao Chu, &Chuin-Mu Wang(2018) use of stock investigation utilizing Deep learning this paper utilizes a neural network with memory ability: RNN. To work on its exhibition, LSTM design was used.This paper joins neural networks and stock history information to anticipate stock costs.

In [10] Wei Wang, Hong Zhao, Qiang Li, Zhixiong Liu (2009). A Novel Hybrid Intelligent Model for Financial Time Series Forecasting and Its Application.The study shows that the presentation of financial time series prediction can be altogether upgraded by utilizing EMD-SVR in examination with single SVR.

III. METHODOLOGY

A .Linear Regression

Linear regression is a Machine Learning technique used to build relationship between the forecast variable Y and single predictor variable X i.e

relationship between dependent and independent variable from the dataset.

Equation for linear regression is

$$Y=a_0+a_1X+e \quad (1)$$

Where a_0 represents intercept such as predicted values of Y when $X=0$, a_1 represents slope of the line such as average predicted change in Y when one unit increase in X.

b. Naïve Bayes Classifier

Naive Bayes classifier is a ML calculation dependent on probabilistic Bayes Theorem. It has different applications, for example, spam separating, arrangement records and sentiment prediction and so forth The name Naïve shows that changes esteem in one element doesn't effect on different provisions in the dataset. It is very force model since it makes prediction rapidly. The dataset was haphazardly rearranged, and afterward it was separated into two subsets Training dataset and testing dataset .Training dataset was utilized to prepare the naive Bayes classifier. Test dataset was utilized to get the outcomes.

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right).P(A)}{P(B)} \quad (2)$$

P (A/B) represents posterior probability P(B/A) is probability of likelihood of evidence P(A) indicates probability of evidence, P(B) indicates prior probability.

C. Support vector Machine

It is a supervised AI calculation which can be utilized for both grouping and regression challenges. SVM utilized for order, regression and time series forecasting undertakings. The primary target of SVM is to take care of non-straight regression assessment issues that makes SVM is effective in time series forecasting. In this calculation we plot every information as a point in n dimensional space. We perform arrangement by

tracking down the hyper-plane that separates the two classes. Estimating accuracy of a model by applying Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) are utilized for assessing results for time series forecasting.

$$Y = b + \sum \alpha (1) y(i) * x(i) . x \quad (3)$$

where $y(i)$ is the class value of t training example (i) , \cdot represents the dot product. The vector x represents a test example and the vectors $x(i)$ are the support vectors, b and $\alpha (1)$ are parameters that determine the hyperplane.

D. Multilayer Perceptron

A multilayer perceptron (MLP) is a class of feedforward counterfeit neural network (ANN). The term MLP is utilized vaguely, sometimes freely to allude to any feedforward ANN, sometimes stringently to allude to networks made out of numerous layers of perceptron (with limit actuation). Multilayer perceptron it get familiar with a capacity that is maps series of past perception to foresee future worth. This model is more adaptable relying upon covered up layer. Single information layer hubs utilized for preparing to take care of, covered up layer for changing its weight and yield layer for prediction.

$$Y = \sigma(\sum(w(i)x(i) + w(0)) \quad (4)$$

Where Y represents output, $x(i)$ is input of the neuron, $w(i)$ is the weights of the neuron

E. Auto Regressive Integrated Moving Average

Another normal Time series model that is exceptionally well known among the Data researchers is ARIMA. It represents Autoregressive Integrated Moving average. While remarkable smoothing models depended on a depiction of pattern and irregularity in the information, ARIMA models expect to portray the relationships in the information with one another. An improvement over

ARIMA is Seasonal ARIMA. Different ARIMA models get accuracy for a few models. ARIMA joins Autoregressive (AR) measure and Moving Average (MA) interaction and assembles a composite model of the time series. Autoregressive (AR) is a regression model which is determined connection between past perception and number of slacked perception. Moving Average (MA) is a methodology that used to determined connection between reliance perception and leftover error.

$$X(t) = C + \sum a_t * x_t + e \quad (5)$$

Where c is constant, a_t is autoregressive coefficient, e is residual error and x_t is stationary variable.

F. Decision Tree

Decision Tree has different applications in Machine Learning It works for both ceaseless also straight out information. Decision Trees that are developed truly deep to adapt profoundly unpredictable examples tend to overfit the preparation sets. Commotion in the information might make the tree fill in a totally surprising way. Arbitrary Forests conquer this issue via preparing various decision trees on different subspaces of the element space at the expense of somewhat expanded predisposition. This means that none of the trees in the timberland sees the whole preparing information. The information is recursively parted into allotments. At a specific hub, the split is finished by posing an inquiry on an attribute.

G. Random Forest

It is an AI procedures utilized for anticipating time series information .This model uses bootstrapping strategies to discover perceptions .In these irregular getting to of a few examples in preparing dataset and perform characterization task as per size of the dataset. This model functions admirably when provisions are appropriately ready.

H. Long Short Term Memory

It is a mainstream and incredible model for time series forecasting. Time series had greater intricacy when contrast with regression prescient displaying due to sequence relies upon input variable. LSTM is a kind of RNN that can demonstrate sequence of information so that each example can be thought to be subject to past ones with the capacity of recollecting the qualities from prior stages with the end goal of future use. It has extremely enormous models that can be prepared without any problem.

Artificial Neural Network (ANN): A neural network comprises of no less than three layers specifically: 1) an information layer, 2) covered up layers, and 3) a yield layer. The quantity of components of the informational collection determines the dimensionality or the quantity of hubs in the information layer. These hubs are associated through joins called "neurotransmitters" to the hubs made in the covered up layer(s). The neurotransmitters joins convey a few loads forevery hub in the info layer. The loads fundamentally play the LSTM network that is prepared utilizing Backpropagation Through time series and disappearing angle issue. It has memory impedes that are associated through layers. A square has segments that make it more astute than traditional neuron and memory or late sequences. A square contains entryways that oversee condition of the square and yield. LSTM network contains three doors which are neglect entryway, input door and yield entryway .neglect door chooses restrictively what data discard from block, input door chooses which esteems are taking or contribution to refresh memory state and yield entryway chooses yield dependent on information and memory of the square. These entryways comprise of loads that are picked up during the preparation system. Recurrent

Neural Networks are exceptionally amazing in dealing with the reliance among the information factors. LSTM network that can hold and gain from long sequence of perceptions. The calculation created is a multi-step univariate estimate calculation. To execute the calculation, Keras library along with Theano were introduced on PC.

These model is required gain from the series past perception to foresee next esteem in the sequence. LSTM network yield layer chooses prediction for time series dataset.

LSTM is a sort of RNN which is an incredible time series model which foresee irregular number of steps in future. LSTM is a blend of five extraordinary segments, called entryways which can demonstrate both long term just as short term information they are: Cell State, Hidden State, Information Gate, Forget Gate and Output Gate. Cell State (It addresses the interior memory of cell where LSTM can eliminate or add data. Secret State (This is yield state data ascertain regarding current information, past secret state and current cell state to anticipate the future information. Info Gate (Input Gate comprises of information and chooses the data streams from current contribution to cell state). Disregard Gate (Decides the progression of part of data current info and past cell state to current cell state). Yield Gate (It portrays the yield created by LSTM).

IV. EXPERIMENTAL RESULT

Dataset is gathered from kaggle site and furthermore get from UCI store .The datasets contains a few ascribes they are day by day opening expense, most exorbitant cost, least value, shutting

cost and exchange volume And change close cost and so on

	0	1	2	3	4
date	2013-02-01 00:00:00	2013-03-01 00:00:00	2013-05-01 00:00:00	2013-06-01 00:00:00	2013-01-15 00:00:00
date_block_num	0	0	0	0	0
shop_id	59	25	25	25	25
item_id	22154	2552	2552	2554	2555
item_price	999.00	899.00	899.00	1709.05	1099.00
item_cnt_day	1	1	-1	1	1
item_name	ЯВЛЕНИЕ 2012 (BD)	DEEP PURPLE The House Of Blue Light LP	DEEP PURPLE The House Of Blue Light LP	DEEP PURPLE Who Do You Think We Are LP	DEEP PURPLE 30 Very Best Of 2CD (Фирм.)
item_category_id	37	58	58	58	56
shop_name	Ярославль ТЦ	Москва ТРК "Атриум"	Москва ТРК "Атриум"	Москва ТРК "Атриум"	Москва ТРК "Атриум"

Fig: Data Set

For experimental analysis we are using Python software, Jupyter Notebook platform and Windows 10 operating system.

MAPE is determined by utilizing absolute error for every single perception then, at that point apply percentage like contrast among genuine and anticipated worth gap by number of information focuses then duplicate with hundred. This methodology is helpful when size of anticipated worth is significant in assessing.

$$MAPE = \sum \left(\frac{y_{actual} - y_{predicted}}{n} \right) * 100\% \quad (6)$$

MSE is an assessor for unnoticed information focuses. It very well may be estimated as average squared distinction among assessed and genuine worth. In these n is various perception, y esteems are one without a doubt and other for anticipated perceptions.

$$MSE = 1/n \sum (y_{actual} - y_{predicted})^2 \quad (7)$$

RMSE is an as often as possible utilized measure square root of the contrast between values anticipated by model and noticed qualities RMSE is a proportion of accuracy to look at forecasting error of various models for specific dataset and it is scale subordinate.

$$RMSE = \sqrt{\sum 1/N (y_{actual} - y_{predicted})^2} \quad (8)$$

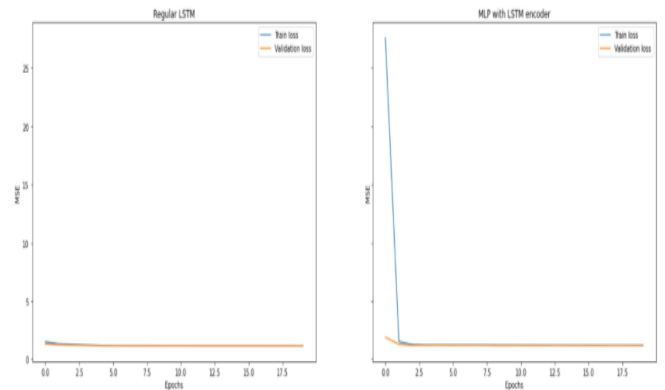


Fig: LSTM

Below Flowchart shows various steps in experimental analysis.

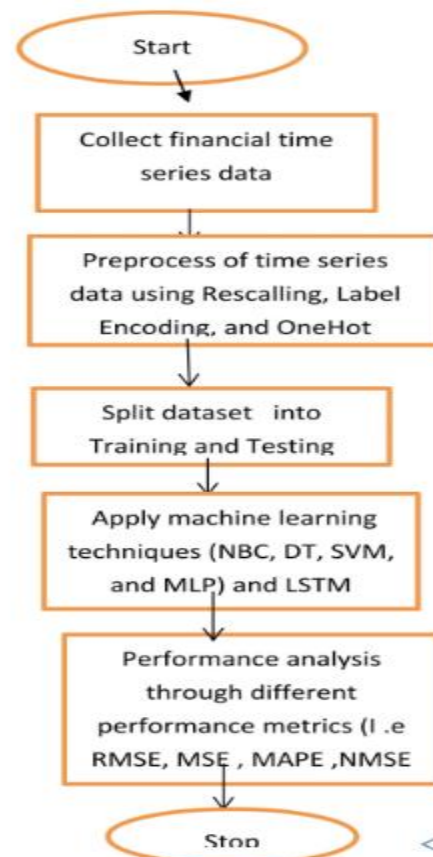


Fig.1: Flowchart for forecasting time series data. Below table shows Performance metrics for different classifiers.

Table -1: Comparison for various metrics

Models	MAPE	MSE	RMSE
Linear Regression	3.37	29.69	5.44
Naïve Bayes Classifier	3.56	27.68	5.32
Support Vector Machine	3.24	24.56	5.21
Multi-Layer perceptron	3.91	27.43	6.36
Decision Tree	3.82	25.78	6.02
Random forest	3.97	25.84	6.54

V. CONCLUSION

The financial time series is unpredictable, the customary forecasting models are less dependable. The dataset identified with the Financial Time Series or Stock Market show that RMSE utilizing LSTM got precise outcome when contrast and customary models. The Neural Network along RNN with memory and utilizations the LSTM engineering. LSTM taxi be utilized to address long-term time-subordinate problems Effectiveness of forecasting model has never been halted. Further we can utilize progressed Deep Learning models to examine in better manner.

REFERENCES

[1] Aakanksha Sharaff, Meenakshi Choudhary, "Comparative Analysis of Various Stock Prediction Techniques" 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI).

[3] H. S. Hota, A. K. Shrivastava, and R. Hota, "An Ensemble Model for Detecting Phishing Attack with Proposed Remove-Replace Feature Selection Technique," *Procedia Comput. Sci.*, no. 132, pp.900–907, 2018.

[4] Zhang Guohui. Research on Time Series Prediction and Its Application Based on Deep Belief Network [D]. Harbin Institute of Technology. 2017.6.08

[5] Bin Gui Xianghe Wei "Financial Time Series Forecasting Using Support Vector Machine" 2014 Tenth International Conference on Computational Intelligence and Security pp. 39-44 2014.

[6] Y. Zhao, J. Li, and L. Yu, "A Deep Learning Ensemble Approach for Crude Oil Price Forecasting," *Energy Econ*, vol. 66, pp. 9–16, 2017

[7] Hongbo Sun, Jing Xu "Improved Approaches for Financial Market Forecasting Based on Stationary Time Series Analysis".

[8] K. Kanchimalay, "Time Series based Forecasting for Crude Palm Oil Price Utilizing Neural Network Algorithms" *.Res*, vol. 3, no. 2, pp. 259–267, 2008.

[9] I. Haidar S. Kulkarni H. P. H. Pan "Forecasting model for crude oil prices based on artificial neural networks" 2008 Int. Conf. Intell. Sensors Sens. Networks Inf. Process. June 2018.

[10] Wei Wang, Hong Zhao, Qiang Li, Zhixiong Liu (2009) "A Novel Hybrid Intelligent Model for Financial Time Series Forecasting and Its Application" *Computer Engineering and Applications* vol. 43 pp. 21-08 2009