PREDICTING THE NEXT DIGIT IN A SEQUENCE USING LSTM.

Abdullahi Mohamud Salad

ABSTRACT

The fundamental project is grasping the patterns in the sequence of records and then the useof this sample to analyze the future. Deep Learning has been established to be higher in graspingthe patterns in each structured and unstructured data. To apprehend the patterns ina lengthy sequence of data, networks to analyze patterns are needed throughout time. Recurrent Networks is the one commonly used for mastering such data. Recurrent Networks are usually used for discovering such data. It is capable of sensing Long and Short Term dependencies or temporal differences. This proposed method is used to predict the next digit based on the given series of numbers. LSTM Deep learning Model is used to make predictions on the new data. LSTM has the advantage of feed forward neural networks (FFNN) and RNN can produce the patterns from long durations of time.

Keywords: LSTM, Number series prediction, Deep Learning

1. INTRODUCTION:

For many organizations, massive records – super volumes of uncooked structured, semi-structured, and unstructured records – is an untapped aid of the brain that can assist enterprise selections and decorate operations. As evidence continues to expand and change, so many techniques were implemented for predictive analytics, to faucet into that meaningful resource and advantage from facts at scale. Predictive analytics is pushed through predictive modeling. It's extra of a strategy than a process. Predictive analytics and laptop gaining knowledge go hand-in-hand, as predictive fashions normally encompass a desktop mastering algorithm. These fashions can be educated over time to reply to new statistics or values, turning in the consequences the enterprise needs. Predictive modeling generally overlaps with the subject of desktop learning. Predictive evaluation is the evaluation of historic facts as properly as current exterior records to locate patterns and behaviors. Predicting the next word, the next frame, the next location, the next time series, and so many applications are developed using Predictive evaluation.

Demand forecasting is the science of predicting the future. Analytics that predicts the future makes use of thousands strategies at statistics of geology, analytics, and simulation, computing device understanding, and synthetic brain to investigate contemporary facts can make forecasts for the next. It makes use of a variety of facts mining, Quantitative and predicting simulation strategies to deliver collectively the management, records technology, and modeling enterprise manner to make predictions about the future. The patterns observed in

enterprise manner to make predictions about the future. The patterns observed in

historic and asynchronous statistics could be applied to become aware of dangers and possibilities for the future. Predictive analytics finds associations pattern amongst many different elements to validate the complicated set of prerequisites to assign a score, or weightage. By efficiently making use of analytic forecasting , organizations are able successfully translate massive information for their own good. The facts extracting and textual content statistics alongside with statistical data, lets in the enterprise users to create predictive talent via uncovering patterns and relationships in each the structured and unstructured data. Thefacts which can be used simply for evaluation are structured data, examples like age, gender, marital status, income, sales. Unstructured statistics are textual statistics in name core notes, social media content, or different kind of open textual content which want to be extracted from the text, alongside with the sentiment, and then used in the mannequin constructing process.

Predictive analytics lets in businesses to grow to be proactive, ahead looking, expecting results and behaviors based totally upon the records and now not on a hunch or assumptions. Prescriptive analytics, goes in addition and recommend movements to gain from the prediction and additionally supply selection selections to gain from the predictions and its implications. Prescriptive Analytics robotically automate complicated choices and change offs to make predictions and then proactively replace hints primarily based on altering activities to take benefit of the prediction. The fig. 1.1 shows the predictive model.

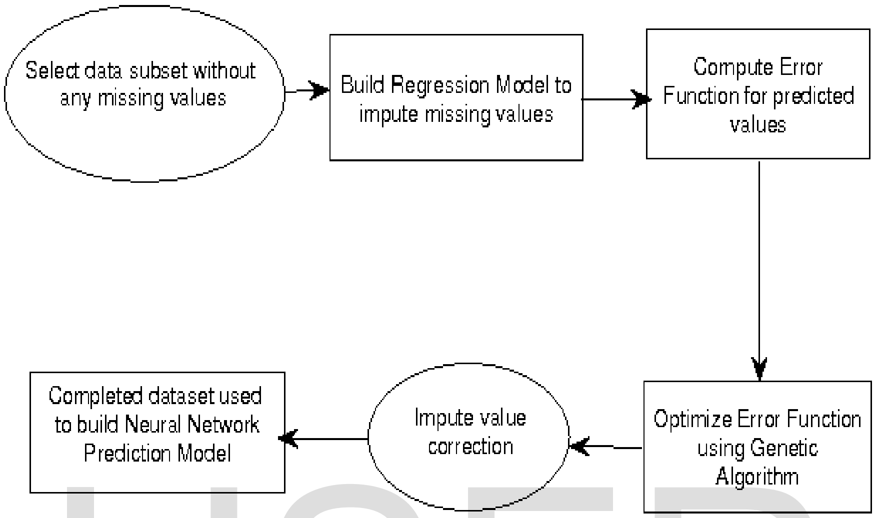

Fig.1.1 Prediction model

For making predictive analysis, all that have to do is collect the proper data, do construct the proper kind of statistical model, and be cautious of regarded assumptions. The proposed predictive analytic is going to yield higher effects in all the input that had provided for the system.

2. LITERATURE SURVEY

In [1], Violos et al. (2020) proposes the use of Artificial Neural Networks (ANN) with LSTM layers for the subsequent position prediction of moving objects utilizing a genetic algorithm and a transfer learning technique. The genetic program does a sophisticated analysis in the domain of science assess a c relative of optimum The algorithm and the ANN structure use learning algorithm. technique to utilize a wellspring of knowledge -qualified ANN models are used to accelerate up the training process. method.. In [2], Ghanbari et al. analyzed on the time-series dataset with several steps with various models of LSTM neural networks. The aim of this process is to show an accurate and important way using LSTM RNN used information containing genes. These methods are utilized in additional statistics that are comparable also The systems are designed to be used with a variety of number of databases. small correction. Authors did a comparison over the conclusions drawn out of the techniques between various systems. In this method, household electricity consumption dataset gathered for 4 years. Among the current framework algorithms, the preferred forecast yields the least number of inaccuracy are achieved through this. In [3] Liu et al (2019) implemented the model for forecasting the time series. Deep neural network with recurrent structures is implemented to get information from sequential data. LSTM is the kind of recurrent neural network that absorbs long term dependencies. It is appropriate for predicting time series with long- and short-term dependencies. This technique on time series leaning forecasting problems. The outputs which are brought from the conventional regression methods and also suggestively greater than the existing results over the stock data sets, those are very near to the random walk sequences.

. In [4], Wang et al. (2019) studied the complications of predictions in LSTM multi-step time series. By surveying two methods of multi-step seq2vec and input, the technique gives information for LSTM in the process of time prediction. The LSTM technique is analyzed to predict the chosen data

In [5], Suryo et al,(2019), analyzed the development of the forecast using Backpropagation and LSTM, and the report proved that this technique yield a value 0.8 for RMSE for LSTM and the backpropagation gives 0.10 for agriculture applications. Suryo et al. achieved Agricultural production is needed to boost agricultural production quickly. in small agriculture techniques. From this smart system, the environmental condition data such as water level, temperature, and other things are monitored.

In [6], wang et al. (2020) focused on trajectory prediction. This system is enhanced with LSTM deep learning, and it is employed clearly to understand mobility patterns over the user's historical trajectories and predict users' drive. In [6], extended to the prediction scheme based on region-oriented technique and proposed a framework based on multi-user multi-step which is incorporating the sequence-to-sequence (Seq2Seq) learning. In [7], Haoye et al. (2018) proposed the Period detection and vogue prediction algorithms associated to periodic information, duration detection strategies are nevertheless confined to the purposes of autocorrelation functions. Period detection phase tries one-of-a-kind numbers of Learning Automata algorithms (LAs) to study and predict the information trend. If the range of LAs deployed coincides with the length of the entered data, then all LAs are characteristic beneath stationary environment. The experimental consequences exhibit that for most of the cases, the variety of required observations stays at a low level. In [8], Bunker et al. (2017), focus on the software of Artificial Neural Network (ANN) to activity effects estimating. Authors pick out the getting to know methodologies utilized, statistics sources, the gorgeous ability of mannequin evaluation, and precise challenges of predicting activity results. In [9], Strohbeck et al. (2020) proposed a convolutional community that operates on rasterized actor-centric photographs which encode the static and dynamic actor-environment. Authors predict a couple of viable future trajectories for every visitor's actor, which consists of position, velocity, acceleration, orientation, yaw price, and function uncertainty estimates. To make higher use of the previous movement of the actor, authors recommend hiring temporal convolutional networks (TCNs) and counting on uncertainties estimated from the preceding object monitoring stage. In [10], Steven Elsworth and Stefan Guttel(2020), skilled in uncooked numerical time-sequence statistics showcase imperative barriers such as excessive sensitivity to the hyperparameters and even to the initialization of random weights. A mixture of a recurrent neural community with a dimension-reducing symbolic illustration is proposed and utilized for the reason of time collection forecasting.

In [13], Indriasari (2019) advise the Predictive Analyst for the Financial sector. The troubles that relate to the automatics and clever tracing of the transaction and banking patron turn out to be the foremost problems nowadays. Whereas most of the transactions that alreadymoved to digital and online transactions wanted extra wise facets on the Predictive Analyst. Predictive analytics (PA) makes use of quite a number of algorithms to find out one-of-a-kindpatterns in the huge information surroundings that would possibly create greater prices for corporations such as in economic institutions. In [15], ciaburro et al, (2021) take advantage of the statistics amassed in the past, attempting to perceive ordinary buildings in what came about to predict what should happen, if the equal constructions repeat themselves in the future as well. A time sequence represents a time sequence of numerical values determined inthe previous at a measurable variable. The values are sampled at equidistant time intervals, in accordance to a suitable granular frequency, such as the day, week, or month, and measuredin accordance with bodily devices of measurement. In computing device learning-based algorithms, the facts underlying the information are extracted from the information themselves, which are explored and analyzed in search of habitual patterns or to find out hidden causal associations or relationships.

3. METHODOLOGY

Predictive analytics is a branch of advanced analytics that makes predictions about future outcomes using historical data combined with statistical modeling, data mining techniques, and machine learning. Long short-term memory (LSTM) is a deep learning architecture that uses an artificial recurrent neural network (RNN). LSTM has feedback connections, unlike normal feedforward neural networks.. It can process whole data sequences as well as single data points.

3.1 Long Short-Term Memory (LSTM):

The long short-term memory (LSTM) contains distinct units called memory unit in the recurrent hidden layer. Every memory unit has 3 gates named as **Forget gate, Input gate, Output gate.** An input gate that selects whether or not to allow additional input, a forget gate that erases unimportant data, and an output gate that determines what data to output**.**

The input gate controls the flow of input activations to the activations into the rest of the network. The forget gate scales the internal state of the cell before adding it as input to the cell through the self-recurrent connection of the cell, therefore adaptively forgetting or resetting the cell's memory.

Forget Gate: This gate determines whatever information from the cell in that particular time stamp will be excluded. The sigmoid function determines this. For each number in the cell state Ct1, it looks at the previous state (ht-1) and the content input (Xt) andproduces a number between 0 (omit this) and 1 (keep this).

$$f_t = \sigma(w_f.[h_{t-1}, x_t] + b_f) \quad (1)$$

Update Gate/input gate: It decides how much of this unit is added to the current state.

$$i_t = \sigma(w_f.[h_{t-1}, x_t] + b_i) \quad (2)$$

$$C_t = tanh(w_c.[h_{t-1}, x_t] + b_c) \quad (3)$$

Output Gate: It decides which part of the current cell makes it to the output.

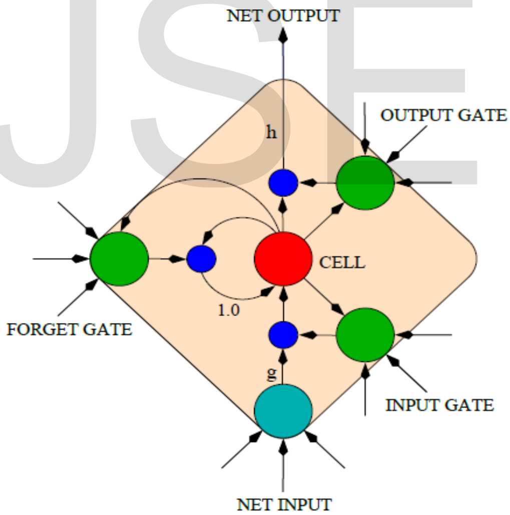

Fig. 3.3 LSTM memory cell

$$o_t = \sigma(w_o.[h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = O_t * tanh(c_t) \quad (5)$$

The sigmoid function determines which values are allowed to pass through 0,1, and the tanh function assigns weight to the values that are passed, determining their relevance level ranging from -1 to 1.

Those 3 doors are digital gates that function in the 0 to 1 area and are based on the sigmoid function. Figure .3.4 Explains it. A horizontal line that can be seen running through the cell represents the cell state.

In LSTM network, The cell state and the hidden state are both passed to the following cell. The main chain of data flow is the cell state, which permits the data to pass ahead largely unmodified. Certain linear transformations, meanwhile, might happen. Sigmoid gates can be used to add or remove data from the cell state. A gate is analogous to a barrier or a set of mathematical operations that each have their own set of distinct values. Because LSTMs use gates to manage the memorization process, they are created to avoid the long-term dependency problem.

The first stage in building a Lstm model is to represent the information that isn't needed and will be left out of the unit. The sigmoid function, which takes the output of the last LSTM unit (ht1) at time t 1 and the current input (Xt) at time t, determines the process of detecting and excluding data. The sigmoid function also selects whether sections of the old output should be removed. The forget gate (or ft) is a gate in which ft is a column with entries valuesbetween 0 and 1 that correlates to every digit in the lstm unit, Ct1.

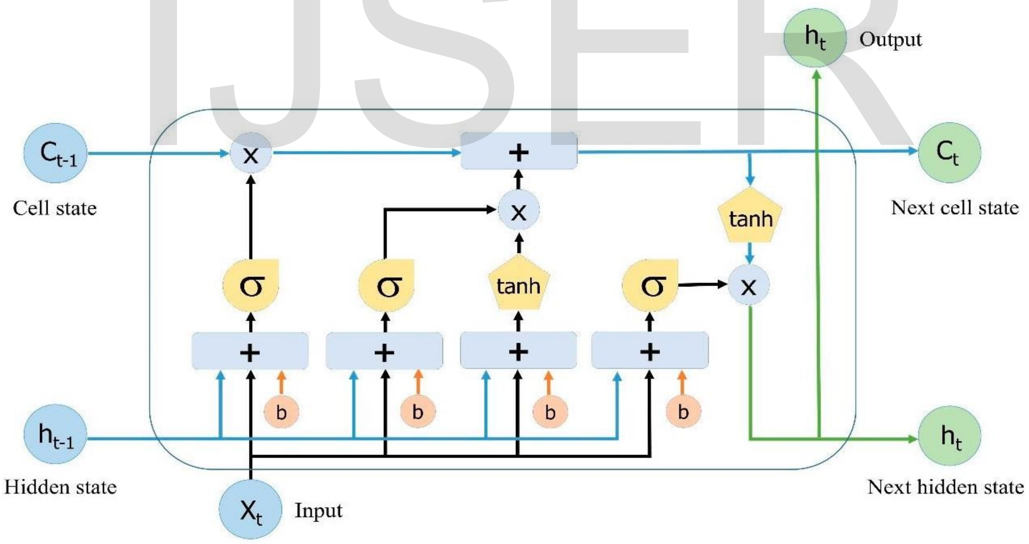

Fig. 3.4 Detailed Structure of LSTM

Table 3.1 LSTM Description

Inputs	**Outputs**	**Nonlinearities**	**Vector operations**
X_t – Current Input	C_t – New updated memory	σ – Sigmoid layer	**X** – Scaling of information
C_{t-1} –Memory from	h_t – current output	tanh – Tanh layer	+ – Adding

$$f_t = \_(W_f\ [h_{t-1}, X_t] + b_f\ ). \qquad (6)$$

Herein, σ The forget gate's weight matrices and bias, Wf and bf, are the sigmoid function respectively. The decision and storage of information from the new input (Xt) in the lstm unit, as well as the upgrading of the lstm unit, are the next steps. The sigmoid layer and the tanh layer are the two components of this stage. The sigmoid layer determines whether new data should be upgraded or discarded (0 or 1), while the transfer function assigns weighting to the numbers that move by, determining their relevance level (-1 to 1). To create the new cell state, the two parameters are combined. This new memory is then combined withCt1 to create Ct.

$$i_t = \_(W_i\ [h_{t-1}, X_t] + b_i), \qquad (7)$$

$$N_t = \tanh(W_n\ [h_{t-1}, X_t] + b_n), \quad (8)$$

$$C_t = C_{t-1} f_t + N_t\ i_t. \qquad (9)$$

Ct-1 and Ct are the unit stages at time t-1 and t, respectively, whereas W and b are the cell state's weight matrices and bias, respectively.

The output values (ht) in the final step are filtered versions of the output cell state (Ot). A sigmoid layer is used to determine whether aspects of the cell state make it to the output. The sigmoid gate's output (Ot) is then combined by the new numbers generated by thetanh layer from the cell state (Ct), with a result ranging from -1 to 1..

$$O_t = (W_o[h_{t-1}, X_t] + b_o), \qquad (10)$$

$$h_t = O_t \tanh(C_t). \qquad (11)$$

Here, W_o and b_o are the weight matrices and bias, respectively, of the output gate.

3.3 PREDICTION MODEL

The LSTM model will find out a function that connects a sequence of past observations as input to an output observation. As such, the sequence of observations must betransformed into multiple examples from which the LSTM can learn. For example, consider the following sequece is the input.

[10, 20, 30, 40, 50, 60, 70, 80, 90]

This sequence of input should be Samples of input/output patterns that are separated into many input/output patterns.. For example, For the one-step prediction that is being learned, three time steps are utilised as input and one time step is used as output.. This can be divided by train_test_split method of sklearn module.

Preparation of LSTM Module:

The input gate controls the flow of input activations to the activations into the rest of the network. The forget gate scales the internal state of the cell before adding it as input to the cell through the self-recurrent connection of the cell, therefore adaptively forgetting or resetting the cell's memory.

Training and Testing with LSTM model:

The deep learning algorithm is given the feature representation from the training data. LSTM network for training and in the end of each epoch, the network performance is found by validating the network with the validating dataset and the network is controlled to achieve the global convergence. Iterative gradient descent is used to reduce the overall error of an LSTM on a collection of procedures. algorithm such as backpropagation through time can be used to change each weight in proportion to the derivative of the error with respect to it. A problem with using For typical RNNs, the error gradients drop exponentially with the amountof the time lag between key events when using gradient descent.. But in LSTM units, however, when error values are backpropagated from output, they remain in the unit's memory. This "error carousel" mistake is fed back to each of the gates until they understand to chop off the value. As a result, regular backpropagation is a good way to train an LSTM. unit to remember values for long durations. After convergence the models is tested using the testing dataset and the performance of the classifier is noted down for sequence of data.

4. RESULTS AND DISCUSSION

In this section, the implementation process of LSTM is discussed. This section shows the training and testing methodology of LSTM with variable size of input. The implementation of this model is to predict the next digit of given series as input. To implement the LSTM in python, we have to import all the necessary libraries first sequential from keras models, dense and LSTM from Keras layers.

Train test split from sklearn.model_selection, numpy and matplotlib as usual. We can provide the sequence of data as the input in this system, but in this code here generates 100 vectors of five consecutive digits into the variable data, and the target will store the six consecutive digits. And when dealing with other cases like text or images, and all these vector sizes may change. All of the input sequence should be converted into NUMPY arrays and then check the shape, for applying the shape value for the input shapes as parameters when we'll be implementing LSTM. LSTM model creation is displayed in Fig. 4.2. Here, Sequential model is started and then LSTM layer is added. So while adding LSTM layers, we need to specify some parameters. The first parameter is the output size, which is one, for getting one output and the batch input describes the input shape of our data, and the format is the number of inputs, length of input sequences and length of each vector. Now if we don't know the number of inputs in the data, then it can be replaced with value of none and the number 5 is used to represent the length of input sequence because 5 digits has been given and one represents the length of each vector, so each vector is one cross one.

Stacked LSTM

A Stacked LSTM model is created by stacking different hidden LSTM layers one on bottom of the other. An LSTM layer needs a dual input, and by standard, LSTMs produce a double output as a result of the sequence's conclusion.. The Fig. 4.6 shows the stacked LSTM modeling.

Bidirectional LSTM And Data Collection

Allowing the LSTM model to study the input data either forward or reverse and combine both readings can be advantageous in some sequence prediction tasks.This is known as The LSTM is bidirectional. For unitary prediction, we can use a Bidirectional LSTM by covering the first hidden state with a Bidirectional wrapper layer.. The Fig. 4.7 shows the stacked LSTM modeling.

To estimate the performance of the classifier, and to allow comparisons, several rates have been considered.

Table.4.1. MAE, MSE and MAPE data for the prediction

Model	MAE	MSE	MAPE
LSTM(50 epochs)	0.185	0.064	0.0071
LSTM(100 epochs)	0.165	0.037	0.0025
LSTM(200 epochs)	0.140	0.030	0.0053
LSTM(400 epochs)	0.110	0.012	0.0035
STACKED LSTM (400 epochs)	0.020	0.015	0.0021
BIDIRECTIONAL LSTM (400 epochs)	0.030	0.016	0.0028

Data Collection 2

Time Series Forecasting

date	type	locale	locale_name	description	transferred
3/2/2012	Holiday	Local	Manta	Fundacion de Manta	FALSE
4/1/2012	Holiday	Regional	Cotopaxi	Provincializacion de Cotopaxi	FALSE
########	Holiday	Local	Cuenca	Fundacion de Cuenca	FALSE
########	Holiday	Local	Libertad	Cantonizacion de Libertad	FALSE
########	Holiday	Local	Riobamba	Cantonizacion de Riobamba	FALSE
########	Holiday	Local	Puyo	Cantonizacion del Puyo	FALSE
########	Holiday	Local	Guaranda	Cantonizacion de Guaranda	FALSE
########	Holiday	Regional	Imbabura	Provincializacion de Imbabura	FALSE
########	Holiday	Local	Latacunga	Cantonizacion de Latacunga	FALSE
########	Holiday	Local	Machala	Fundacion de Machala	FALSE
7/3/2012	Holiday	Local	Santo Domingo	Fundacion de Santo Domingo	FALSE
7/3/2012	Holiday	Local	El Carmen	Cantonizacion de El Carmen	FALSE

We have Collected thousands of Data From Kaggle but in paper we shared some of the Data about Time Series Forecasting

We trained the LSTM with different number of epochs and layers, based on the results in Table 4.1, the expected accuracy is achieved in 400 epochs.

5. CONCLUSION

Predicting next event, next time series, next frame and next values are important in data science. This proposed method is used to predict the next digit based on the given series of numbers. LSTM Deep learning Model is used to make predictions on the new data. In this system LSTM with various epochs has been tested to predict the data. Stacked LSTM and Bidirectional LSTM is also tested with the results. From the report, these three models are

producing the better accuracy when the size of the epochs increases. The prediction system strengthen the future data analysis in every department. In the proposed system stacked LSTM gives more accuracy than others.

REFERENCES

1. Violos, John & Tsanakas, Stylianos & Androutsopoulou, Maro & Palaiokrassas, Georgios & Varvarigou, Theodora. (2020). Next Position Prediction using LSTM Neural Networks. 232-240. 10.1145/3411408.3411426.

2. Ghanbari, R., & Borna, K. (2021). Multivariate Time-Series Prediction Using LSTM Neural Networks. 2021 26th International Computer Conference, Computer Society of Iran (CSICC). doi:10.1109/csicc52343.2021.9420543

3. Liu, Yuqi; Su, Zhongfeng; Li, Hang; Zhang, Yulai (2019). An LSTM based classification method for time series trend forecasting, 14th IEEE Conference on Industrial Electronics and Applications (ICIEA) - 402–406. doi:10.1109/ICIEA.2019.8833725

4. Wang, Yupeng; Zhu, Shibing; Li, Changqing (2019), Research on Multistep Time Series Prediction Based on LSTM, 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), 1155–1159. doi:10.1109/eitce47263.2019.9095044

5. Suryo Putro S, Budi Cahyo; Wayan Mustika, I; Wahyunggoro, Oyas; Wasisto, Hutomo Suryo (2019). Improved Time Series Prediction Using LSTM Neural Network for Smart Agriculture Application, 5th International Conference on Science and Technology (ICST), 1–4. doi:10.1109/ICST47872.2019.9166401

6. Wang, C., Ma, L., Li, R., Durrani, T. S., & Zhang, H. (2019). Exploring Trajectory Prediction through Machine Learning Methods. IEEE Access, 1–1. doi:10.1109/access.2019.2929430

7. Lu, Haoye; Srinivasan, Anand; Nayak, Amiya (2018), Period Detection and Future Trend Prediction Using Machine Learning Techniques, IEEE International Conference on Internetof Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 828–833. doi:10.1109/Cybermatics_2018.2018.00161

8. Bunker, Rory P.; Thabtah, Fadi (2017). A machine learning framework for sport result prediction. Applied Computing and Informatics,S2210832717301485–. doi:10.1016/j.aci.2017.09.005

9. Jan Strohbeck , Vasileios Belagiannis , Johannes Muller, "Multiple Trajectory Prediction with Deep Temporal and Spatial Convolutional Neural Networks", 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) October 25-29, 2020, Las Vegas, NV, USA (Virtual)

10. Steven Elsworth and Stefan Guttel, "Time Series Forecasting Using LSTM Networks: A Symbolic Approach", arXiv:2003.05672v1 [cs.LG] 12 Mar 2020

11. E. Indriasari, H. Soeparno, F. L. Gaol and T. Matsuo, "Application of Predictive Analytics at Financial Institutions: A Systematic Literature Review," 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI), 2019, pp. 877-883, doi: 10.1109/IIAI-AAI.2019.00178.

12. Ciaburro, G.; Iannace, G. Machine Learning-Based Algorithms to Knowledge Extraction from Time Series Data: A Review. Data 2021, 6, 55. https://doi.org/10.3390/ data6060055

13. P. Pawar, S. Hatcher, L. Jololian and T. Anthony, "Demand Forecasting using Machine Learning," 2019 SoutheastCon, 2019, pp. 1-8, doi: 10.1109/SoutheastCon42311.2019.9020393.

14. A. Matsunaga and J. A. B. Fortes, "On the Use of Machine Learning to Predict the Time and Resources Consumed by Applications," 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010, pp. 495-504, doi: 10.1109/CCGRID.2010.98.

15. Kumar, Vaibhav & L., M.. (2018). Predictive Analytics: A Review of Trends and Techniques. International Journal of Computer Applications. 182. 31-37. 10.5120/ijca2018917434.