nes-3

by siswanto@pktj.ac.id 1

Submission date: 15-Dec-2023 12:44AM (UTC+0800)
Submission ID: 2257748595
File name: JUITA-CYBER_ATTACKS-JOKO-CL.docx (1.67M)
Word count: 5954
Character count: 33081

Number of Cyber Attacks Predicted With Deep Learning Based LSTM Model

Joko Siswanto[1*], Irwan Sembiring[2], Adi Setiawan[3], Iwan Setyawan[4]

[1]Road Transportation Systems Engineering, Politeknik Keselamatan Transportasi Jalan, Tegal, Indonesia

[1,2,3]Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia

[4]Faculty of Electronics and Computer Engineering, Satya Wacana Christian University, Salatiga, Indonesia

[1]siswanto@pktj.ac.id, [2]irwan@uksw.edu, [3]adi.setiawan@uksw.edu, [4]iwan.setyawan@uksw.edu

Abstract - The increasing number of cyber attacks will result in various damages to the functioning of technological infrastructure. A prediction model for the number of cyber attacks based on the type of attack, handling actions and severity using time-series data has never been done. A deep learning-based LSTM prediction model is proposed to predict the number of cyberattacks in a time series on 3 evaluated data sets MSLE, MSE, MAE, RMSE, and MAPE, and displays the predicted relationships between prediction variables. Cyber attack dataset obtained from kaggle.com. The best prediction model is epoch 20, batch size 16, and neuron 32 with the lowest evaluation value on MSLE of 0.094, MSE of 9.067, MAE of 2.440, RMSE of 3.010, and MAPE of 10.507 (very good model because the value is less than 15) compared other variations. There is a negative correlation for INTRUSION-MALWARE, BLOCKED-IGNORED, IGNORED-LOGGED, and LOW-MEDIUM. The predicted results for the next 12 months will increase starting from the second month at the same time. The resulting predictions can be used as a basis for policy and strategy decisions by stakeholders in dealing with fluctuations in cyber attacks that occur.

Keywords: cyber attack, prediction, LSTM, deep learning

I. INTRODUCTION

Cybersecurity is a major problem for every service operating online[1] throughout the world[2] which has a detrimental impact on society[1]. Cybersecurity is challenged to accept the possibility and understand the occurrence of attacks in complex systems[3]. Threat intelligence properties are used to improve overall cyber security[4]. The demand for cyber security and protection against various types of cyber attacks is increasing according to the needs of the cyber world[5]. Hackers are targeting more organizations with a variety of distinct cyberattacks[2]. Cybersecurity experts are placing greater emphasis on approaches to assessment and mitigation[6]. Cybersecurity professionals have a duty to protect organizational data[7] in more ways[2]. Cybersecurity is related to the protection of data, information systems and digital assets of an organization [8]. A complete and related knowledge format is used to extract concepts and entities found in cyber security attacks[9]. Cyber attacks are an important system security challenge[10] and the biggest problem in the world[11]. Cyber attacks can occur intentionally and/or unintentionally[8] with targets increasing exponentially[12] as technology advances[8] with very bad impacts[13]. Attackers began to use non-standard schemes to implement attacks and employees of organizations as intermediaries to reduce the efficiency of breach detection[12]. Cyberattacks monitor overall application behavior using distributed tracing and detect anomalous cyberattack activity by calculating the frequency distribution of unique traces[2]. Criminals exploit weaknesses[14] or use the distinctive characteristics of emerging technologies[13]. Data protection and security is a big challenge in the modern technical world against cyber attacks[8]. Cyber attacks that often occur are ransomware, malware, social engineering, phishing, cryptojacking, zero day exploit, cross-site scripting (XSS), drive-by-downloads[14], man-in-the-middle, DDoS[6], port scan, bot, brute force, SQL injection, and heartbleed[8]. The increasing number of cyber attacks will result in various damages to the functionality of technological infrastructure[15][16].

Attack prediction can basically be done in two ways, namely a statistical approach and an algorithmic approach[7]. Cyber attack prediction strategies can be provided by artificial intelligence[14], machine learning[10], and deep learning[14][1]. Advanced cyber attack prediction based on Network Intrusion Detection Systems (NIDS) Intrusion Alert uses the intrusion Alert Correlation (AC) taxonomy with the result of providing a timely, concise and high-level view of the network security situation[17]. Prediction of cyber attacks with an intrusion detection system uses an artificial neural network (ANN) with an accuracy rate of 99%[1]. The Rotational Region Convolution Neural Network (R2CNN) model is used to predict the onset of cyber attacks on large connected IoT devices with results in increased accuracy and performance[18]. Prediction of computer attacks on critical information infrastructure (CII) based on comprehensive analysis of incident characteristics and system users can significantly improve the efficiency of incident detection[12]. Adaboost is used to predict DDoS cyber attacks with higher accuracy compared to naïve Bayes, logistic regression, and random forest[19]. Bi-Direction Recurrent Neural Network (BRNN) is used to predict cyber attacks based on real-time datasets and can have high accuracy (92%)[7]. ElasticNet Regression Model (ENetRM) is proposed to predict real-time cyber attacks on over-encrypted traffic in applications with consistency and accuracy capable of outperforming Intrusion Detection System (NIDS), Novel Nested-Arc Hidden semi-Markov Model (NAHSMM) and Density- Based Spatial Clustering of Applications with Noise (DBSCAN)[20]. Linear Support Vector Machine was found to be the most effective cyber attack method with an accuracy rate of 96.02%[6]. Decision Tree (DT) is used to predict cyber attacks correctly and provide patterns related to cyber attacks with 99% accuracy[8]. Holt-Winters, ARIMA, SARIMA, GARCH, and Bootstrapping are used to predict cyber attacks against systems based on time series, each of which has high accuracy[21]. HinAp can automatically predict cyber attack preferences for detection and defense with accuracy that can outperform SVM-B, KNN-B, Node2Vec, Esim, Metapath2Vec, Hin-att, and Hin-tran[22].

Efforts and progress in cyber security prediction are still unclear[13]. Successful cyber attacks are associated with inadequate handling, anticipation and prediction[12]. Most cyber attack prediction approaches focus on the malicious motivation[23] or the cyber attack event process[20]. It is important to observe cyber attack events to predict the future in designing security measures to protect socially sensitive data and critical infrastructure that can provide benefits to individuals, organizations and society[14]. New prediction models are needed by almost all platforms connected to the internet to protect user information from being hacked by intermediaries[18]. A prediction model for the number of cyber attacks using time-series data has never been done. Predictions of the number of cyber attacks can be grouped based on the type of attack, handling actions, and severity. Cyber attacks are recorded every time an attack occurs, so predictions are possible based on the date of the incident. A prediction model using deep learning-based LSTM is proposed to predict the number of cyber attacks in a time-series based on the type of attack, countermeasures, and severity level. 3 Dataset obtained from a time-series of the number of cyber attacks per incident for at least 3 years. Parameter variations were carried out to find the best model optimization and were evaluated using Mean-Squared Logarithmic Error (MSLE), Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) based on the lowest values in the 3 datasets. The prediction model produces 3 types of cyber attack predictions on 3 types of datasets in one model for the next 12 months. 3 types of datasets are used to predict types of cyber attacks, actions to handle cyber attacks, and the severity of cyber attacks. 3 types of predictions are needed to display developments and relationships respectively from the 1st to the 12th month. The proposed prediction model is already commonly used to predict, but in-depth prediction for 3 types of cyber attacks with in-depth data is something new that has been applied and can be used as a reference. The prediction results can be used as a reference for managers and stakeholders in making strategies, anticipating and developing models in dealing with the number of cyber attacks.

II. METHOD

Data collected from a CSV dataset on kaggle.com which contains cyber security attack data every day starting from January 1 2020 to October 11 2023[24]. The 3 types of data taken are attack type, action taken, and severity level which are made into 3 asset data. 3 data were taken and made into a dataset because the data was similar in format and type, and had high importance for prediction. The attack type dataset consists of the number of DDoS, Intrusion, and Malware attack types. The action taken dataset consists of the number of blocked, ignored, and logged actions. The severity level dataset consists of Low, Medium and High severity levels. The total data is 1,380 data based on daily cyber attack data. The data used in this dataset is the date and number of cyber attacks on each type of data every day. The prediction simulation environment uses the Python programming language running on Google Colaboratory with the macOS Sonoma 14.1.1 Operating System and 8 GB RAM. The deep learning framework used is Tensor Flow. The data is first processed using min-max feature scaling, then the dataset is divided into two segments (training and testing). The dataset is run using an LSTM model with different tuning parameters, so that the resulting model has the best suitability, stability and performance. The prediction dataset is compared with the training dataset and the prediction accuracy is evaluated. The selection of LSTM model parameters can be seen from the 5 model evaluation values (MSLE, MSE, MAE, RMSE, and MAPE). The lowest evaluation value from the experiment becomes the most optimal model. The system architecture produces prediction results for the number of cyber attacks with input from a dataset processed by the LSTM model with the best evaluation results as a model for predicting the number of cyber attacks(Fig. 1).

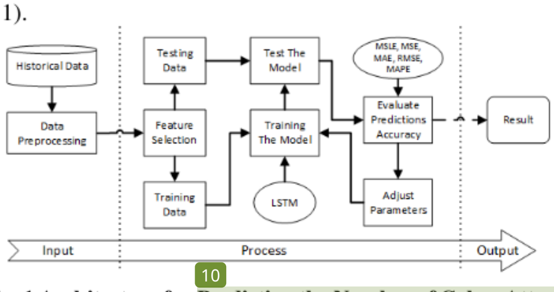

Fig. 1 Architecture for Predicting the Number of Cyber Attacks

The LSTM model is used to learn common case shapes, then predict future events and the time of occurrence[2]. LSTM neural networks have the main goal of modeling long-term dependencies and determining the optimal time lag for time series problems[25]. LSTM can be applied to supervised or unsupervised deep learning models for anomalous event prediction[26]. LSTM consists of an input layer, a recurrent hidden layer, and an output layer[18]. The difference between LSTM deep learning networks and other neural networks lies in the temporal relationships between LSTM units in the hidden layers[27]. The basic unit of the hidden layer is a memory block containing memory cells with independent connections that memorize the temporal state, and a pair of adaptive multiplicative gate units to control the flow of information in the block. Two additional gates named input gate and output gate respectively control the activation of input and output into the block[28](Fig. 2).

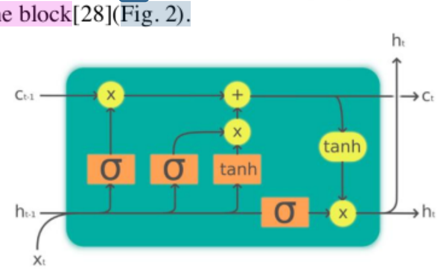

Fig. 2 Basic LSTM Cells Network Architecture

The performance of LSTM is evaluated by training and learning the behavior of logged cases from available data sets[2]. The proposed LSTM model for predicting the number of cyber attacks is evaluated for accuracy with model performance values. Model evaluation was carried out using MSLE, MSE, MAE, RMSE, and MAPE. 5 types of evaluation matrices are used to see the consistency of the proposed model's performance. MSLE is an evaluation metric used to measure the average error of model predictions on actual data on a logarithmic scale[29]. MSLE is useful when the variability between actual and predicted values is very large, and minimizes errors on a logarithmic scale[30] by taking the logarithm of the actual and predicted values, then squaring the difference between them(1). The advantages of MSLE include that this metric avoids the excessive impact of extreme values or outliers[31] and provides a better picture of the quality of model predictions. An MSLE value that is getting closer to 0 is a reflection of better model performance[32].

$$MSLE = \frac{1}{N}\sum_{i=0}^{N}(\log(y_i + 1) - \log(\hat{y}_i + 1))^2 \quad (1)$$

MSE is an evaluation metric that is commonly used to measure the average squared error between the value predicted by the model and the actual value in a dataset[33] which is suitable for predicting continuous values[34]. MSE is easy to calculate, gives large weight to large errors, and has good mathematical properties for model optimization[29]. For each observation, the difference between the actual value and the predicted value is calculated by squaring the difference and taking the average of all the squared difference values to get the MSE value[35](2). The lower the MSE value, the better the model performance in predicting real data[36].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{2}$$

MAE is an evaluation metric used to measure the absolute average error between the value predicted by a model and the actual value in a dataset[37] on the scale of the actual data without considering the direction of the error (positive or negative)[38]. MAE measures the extent to which the model predictions are from the actual values without regard to whether the model tends to overestimate or underestimate[29]. For each observation, the absolute difference (error) of the actual value and the predicted value is calculated and the average absolute difference is calculated(3). The lower the MAE value, the better the model is at predicting real data[39].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{3}$$

RMSE is an evaluation metric that is commonly used to measure the average error level between predicted values and actual values in a dataset[40]. RMSE gives an idea of how well the model can predict actual data and has properties similar to Mean Squared Error (MSE), but the RMSE value is taken as the square root of MSE[41]. RMSE is calculated by taking the square root of the average of the squared differences between the predicted value and the actual value[42] (4). A lower RMSE value indicates that the model is better at predicting real data[43].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4}$$

MAPE is a key performance indicator commonly used for prediction accuracy. MAPE divides each error based on each request[44]. High errors during periods of low demand can have a significant impact on MAPE[45] (5). The smaller the MAPE value, the higher the prediction accuracy. A MAPE value that is getting closer to 0 is a reflection of better model performance[46].

$$MAPE = \frac{1}{n}\sum_{i-1}^{n}\left|\frac{\hat{p}_i - p_i}{p_i}\right| \tag{5}$$

III. RESULT AND DISCUSSION

The data input format is in the form of numeric results from the sum of cyber attacks every day from January 1 2020 to October 11 2023. The data is processed and formatted in the Comma-Separated Values (CSV) file type with 1,380 results. The results of data processing are divided into 3 datasets, namely attack type, action taken, and severity level. The attack type dataset consists of DATE, DDOS, INTRUSION, and MALWARE columns(Fig. 3a). The action taken dataset consists of DATE, BLOCKED, IGNORED, and LOGGED columns(Fig. 3b). The severity level dataset consists of DATE, LOW, MEDIUM, and HIGH columns(Fig. 3c). The DATE column in each dataset contains the date, while the other column contains the number of cyber attack incidents. The data is divided into 2, namely training data (80%) and testing data (20%). Real data and predicted data are subjected to appropriate scaling, training and testing, then the results of the evaluation values from MSLE, MSE, MAE, RMSE, and MAPE are observed using different layers, and different units in 2 hidden layers and dense layers for prediction output.

	DATE	DDOS	INTRUSION	MALWARE
0	2020-01-01	10	13	7
1	2020-01-02	3	13	8
2	2020-01-03	12	10	10
3	2020-01-04	6	11	7
4	2020-01-05	5	8	11
...	...	...	...	...
1375	2023-10-07	10	8	7
1376	2023-10-08	11	5	12
1377	2023-10-09	12	12	8
1378	2023-10-10	9	7	6
1379	2023-10-11	3	8	5

(a)

	DATE	BLOCKED	IGNORED	LOGGED
0	2020-01-01	12	8	10
1	2020-01-02	7	4	13
2	2020-01-03	11	12	9
3	2020-01-04	8	4	12
4	2020-01-05	7	7	10
...	...	...	...	...
1375	2023-10-07	13	2	10
1376	2023-10-08	9	6	13
1377	2023-10-09	10	10	12
1378	2023-10-10	5	10	7
1379	2023-10-11	4	8	4

(b)

	DATE	LOW	MEDIUM	HIGH
0	2020-01-01	6	9	14
1	2020-01-02	7	9	8
2	2020-01-03	7	11	14
3	2020-01-04	7	11	5
4	2020-01-05	6	5	13
...	...	...	...	...
1375	2023-10-07	6	8	11
1376	2023-10-08	6	7	15
1377	2023-10-09	11	10	10
1378	2023-10-10	12	4	5
1379	2023-10-11	6	5	5

(c)

Fig. 3 Research Datasets

Training epochs should be selected in the best way to train the model according to the analysis of different epochs for LSTM models. The default LSTM model is with 2 hidden layers, the activation used is hyperbolic tangent (tanh), and a dropout of 0.20. There are 1 LSTM models used to train the training data with the optimizer used by Adam and Verbos. The variations of the LSTM model that were carried out in the experiment were number of neurons, epoch and batch size. The epoch 20, batch size 16 variation gives the lowest values for 8 neurons, 16 neurons, and 31 neurons. 8 neurons is the most optimal variation used with the lowest evaluation values in MSLE, MSE, MAE, RMSE, and MAPE. It turns out that having more neurons does not make the model better, in fact the opposite can happen.

Increasing the epoch and batch size does not always make the model better, the appropriate model variation is with epoch 20, batch size 16, and neurons 8. 32 neurons is a better variation in the number of neurons than the others, but the large number of epochs and batch sizes does not always make the model better than 4 evaluation values. The most optimal model variations are epoch 20, batch size 16, and neuron 32 with the lowest evaluation value of the 4 evaluation methods(TABLE I).

TABLE I
LSTM MODEL EXPERIMENT RESULTS

E	B	8 Neuron					16 Neuron					32 Neuron				
		E1	E2	E3	E4	E5	E1	E2	E3	E4	E5	E1	E2	E3	E4	E5
20	4	0.102	9.953	2.532	3.154	12.842	0.095	9.646	2.481	3.103	11.063	0.097	9.087	2.445	3.014	12.984
	8	0.101	9.979	2.530	3.160	12.710	0.096	9.657	2.485	3.107	11.966	0.096	9.076	2.441	3.013	11.855
	16	0.101	9.958	2.522	3.155	12.353	0.096	9.660	2.487	3.108	11.514	0.094	9.067	2.440	3.010	10.507
40	4	0.101	9.933	2.527	3.151	12.910	0.098	9.694	2.483	3.117	11.721	0.097	9.099	2.442	3.016	12.292
	8	0.103	9.969	2.536	3.157	12.781	0.098	9.665	2.488	3.113	11.516	0.097	9.092	2.444	3.015	11.752
	16	0.102	9.960	2.532	3.156	12.666	0.096	9.694	2.488	3.109	11.643	0.096	9.081	2.442	3.013	12.343
60	4	0.102	9.941	2.523	3.154	12.459	0.097	9.717	2.490	3.117	11.972	0.098	9.083	2.449	3.017	11.823
	8	0.102	9.967	2.535	3.158	12.852	0.097	9.689	2.488	3.112	11.354	0.098	9.081	2.447	3.016	11.066
	16	0.101	9.945	2.526	3.156	12.613	0.097	9.683	2.488	3.111	11.790	0.096	9.101	2.443	3.013	12.078
80	4	0.103	9.976	2.536	3.154	12.383	0.098	9.697	2.487	3.114	11.718	0.097	9.084	2.444	3.014	11.313
	8	0.101	9.998	2.529	3.162	12.919	0.098	9.720	2.491	3.117	11.614	0.098	9.107	2.447	3.017	11.081
	16	0.101	9.948	2.530	3.154	12.155	0.099	9.704	2.487	3.115	11.069	0.098	9.106	2.448	3.017	12.077
100	4	0.102	9.961	2.533	3.156	12.521	0.099	9.716	2.490	3.117	12.006	0.097	9.094	2.443	3.015	11.045
	8	0.100	9.919	2.521	3.149	12.080	0.097	9.682	2.488	3.111	12.336	0.098	9.114	2.447	3.019	12.137
	16	0.102	9.957	2.531	3.155	12.906	0.098	9.695	2.489	3.113	12.009	0.097	9.088	2.443	3.014	11.008

Note: E=Epoch, B=Batch Size, E1=MSLE, E2=MSE, E3=MAE, E4=RMSE, E5=MAPE.

The best LSTM model with an evaluation value from MSLE of 0.094, MSE of 9.067, MAE of 2.440, RMSE of 3.010, and MAPE of 10.507. The four evaluation models used have a value of less than 10, which means the LSTM model has very good performance quality and is acceptable. This is because if the value is more than 20, then the model needs improvement, even to the point where it is unacceptable. Comparison of loss and validation loss with 32 neurons, batch size 16, and epoch 20 on 3 datasets. Comparison of graphs with epoch variations in general, the training and validation lines are almost the same, so that epoch 20 with the lowest evaluation results is the most optimal model(Fig. 4).

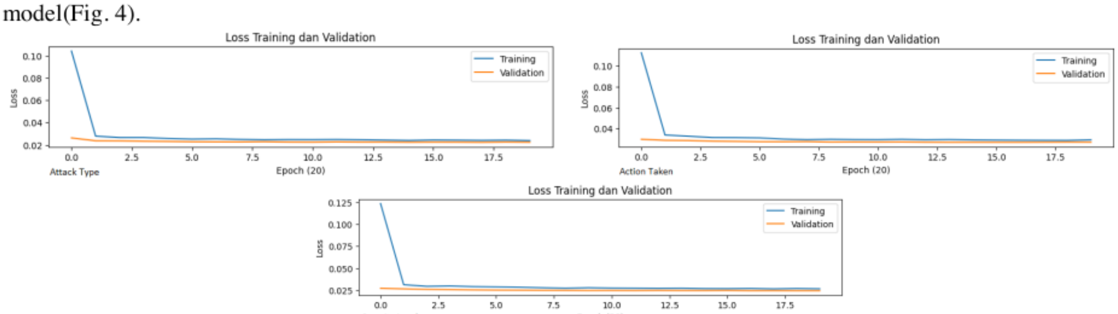

Fig. 4 Output Epoch by LSTM Model

The training and validation loss graph shows that in the three datasets the training and validation lines reach a parallel line. In the analysis of the Loss graph, you can see the difference between loss in training data and validation data. Validation Loss graphs provide insight into how well a model can predict never-before-seen data, and special attention is paid to potential overfitting or underfitting. The deviation between the Loss and Validation Loss graphs can provide important insight into the quality of the model's generalization to new data. Validation Loss which begins to increase will help in optimizing the LSTM model to improve prediction performance.

The prediction results using the proposed model are in accordance with the movement of testing and training data. Prediction results on training and testing data improve with more training carried out. Deep learning carries out deeper learning based on long and short term time by looking at the movement of the number of passengers on 3 types of data in each dataset which is getting better (**Error! Reference source not found.**). The fluctuations in the three movements show the same rhythm, although there are several times there are allusions between the data variants. The use of the LSTM model provides more reliable support in long and short term time modeling. Data recording is the key to being able to carry out learning using deep learning-based LSTM models.

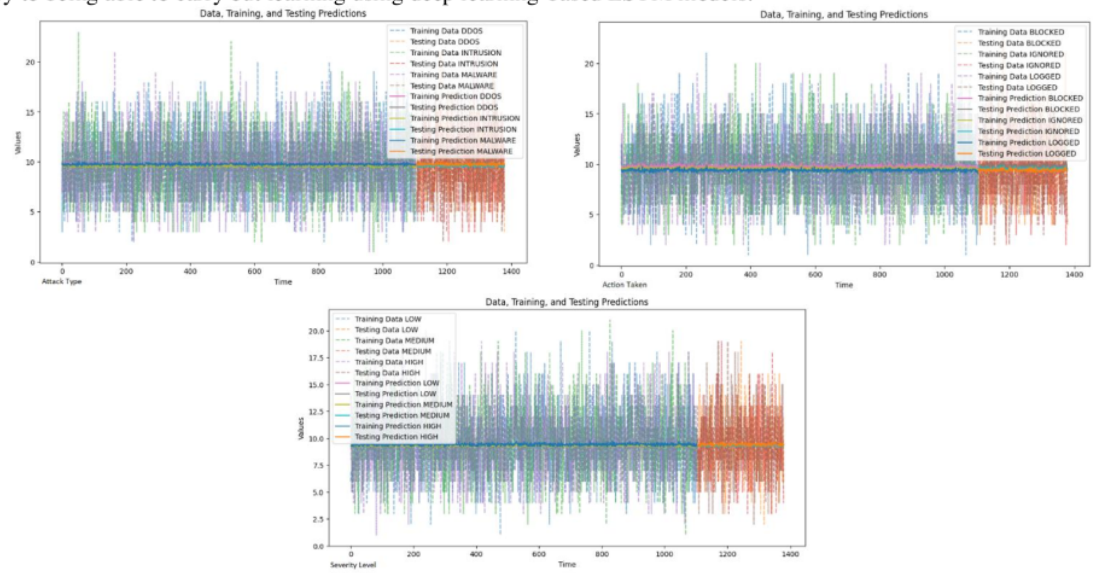

Fig. 5 Data, Training, and Testing Prediction

The three types of predictions produced are related to each other in each dataset. Linkages can be positive or negative. Linkage relationships can be presented with a correlation heatmap. In the correlation heatmap, cold colors such as blue indicate negative correlation and warm colors such as red indicate positive correlation. A positive link means that the types of predictions are directly proportional, while a negative link means that the types of predictions are inversely proportional. The attack type prediction has a strong positive correlation between DDOS-INTRUSION, a strong negative correlation between INTRUSION-MALWARE, and a weak correlation between DDOS-MALWARE. The action taken prediction has a strong negative correlation between BLOCKED-IGNORED and IGNORED-LOGGED, and a weak correlation between LOGGED-BLOCKED. Severity level predictions have a strong negative correlation between LOW-MEDIUM, a strong positive correlation between MEDIUM-HIGH, and a weak correlation between HIGH-LOW(Fig. 6).

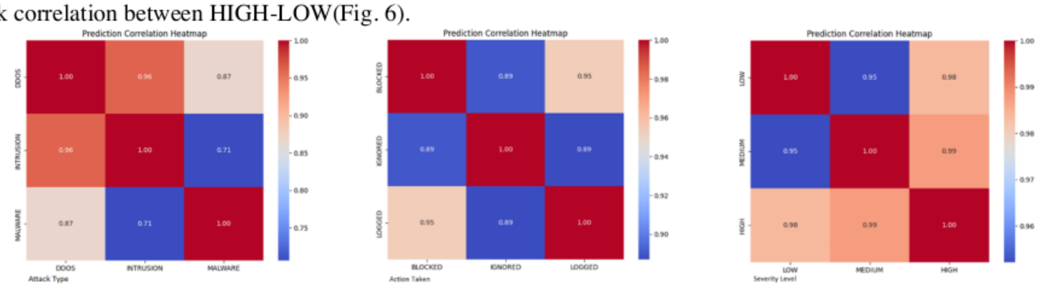

Fig. 6 Prediction Correlation Heatmap

The relationship between predicted variables in one dataset can be visualized with a prediction scatter matrix. The points on a scatter plot that move together or form a line become a consistent positive or negative pattern in the prediction model. If the points are concentrated in an area, it indicates that the variables are interdependent and can be used in predictions. In the three datasets there are variables that have a very strong relationship as indicated by the diagonal scatter plot. The attack type predictions that are closely interconnected and dependent are between

INTRUSION-DDOS and MALWARE-DDOS, while those that are less closely interconnected and dependent are between INTRUSION-MALWARE. Predictions of action taken that are closely interconnected and dependent are between LOGGED-BLOCKED and IGNORED-LOGGED, while those that are less closely interconnected and dependent are between BLOCKED-IGNORED. All severity level prediction variables are closely related and dependent(Fig. 7).

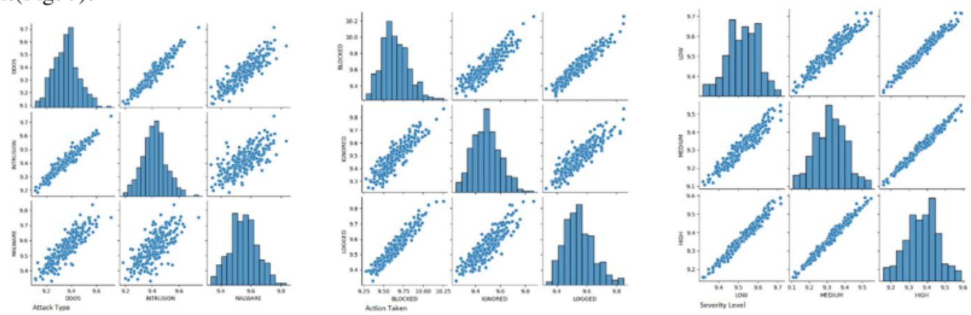

Fig. 7 Prediction Scatter Matrix

The number of cyber attacks in 3 predicted datasets, namely attack type, action taken, and severity level. The prediction results in the three datasets will see a spike starting in the second month onwards. All predicted numbers range from 9 to 10. The attack type dataset is predicted to be MALWARE, which was previously below DDOS and above INTRUSION, will outperform DDOS starting in the second month. The action type dataset is predicted to be BLOCKED which was previously below IGNORED and above LOGGED will outperform IGNORED starting from the second month. The severity level dataset is predicted to be HIGH, which was previously the same as medium and below LOW, which will outperform LOW starting in the second month(Fig. 8).

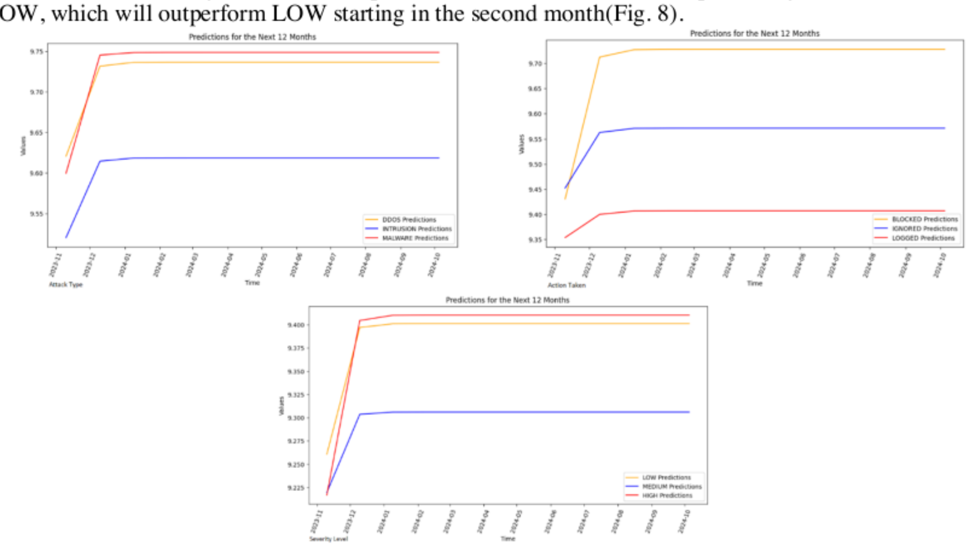

Fig. 8 Predictions For The Next 12 Months

Time is not considered in LSTM model experiments. Combinations and variations greatly influence the time required for processing and evaluation values. In-depth investigations were also carried out on large datasets to achieve the best accuracy and suitability of the desired predictions. Appropriate combinations and variations are important factors for training the model and design of the proposed framework. Empty data and data discrepancies are problems that must be resolved. The findings of the best LSTM model variations were tested on 3 datasets. The dataset is attack type, action taken, and severity level. The evaluation values in each dataset and the average have consistency and range close to the experimental results(TABLE II).

TABLE II
TESTING DATASET

No	Dataset	MSLE	MSE	MAE	RMSE	MAPE
1.	Attack Type	0.096	9.069	2.431	3.008	10.498
2.	Action Taken	0.093	9.068	2.438	3.009	10.238
3.	Severity Level	0.094	9.054	2.441	3.011	10.321
	Avarage	0.094	9.064	2.437	3.009	10,531

The deep learning-based LSTM model was the first to predict the number of cyber attacks on 3 datasets with 3 types of predictions each. The prediction model that is usually used is machine learning, but it is still rare to use a deep learning approach with an LSTM model. The accuracy of the proposed model is better than machine learning approaches such as ANN[1], R2CNN [18], BRNN[7], Cognitive Spectral Clustering [6], Decision Tree[8], ARIMA, SARIMA, GARCH, Bootstrapping[21], SVM, and KNN[22] whose accuracy value is a maximum of 90% with maximum 3 evaluation model. The prediction results and accuracy obtained are optimal, because the LSTM model parameter tuning is done first. Parameter tuning is very necessary to make the model work optimally. The difference between the studies that have been carried out is that in the previous prediction of the number of cyber attacks there were no parameter variants of neuron, epoch, and batch size whose performance was evaluated by 5 types of evaluation models. This study is the first to predict the number of cyber attacks and predict the correlation and relationship between prediction results on 3 different types of datasets. The number of cyber attacks based on attack type, action taken, and severity level can be used as an illustration of the number of attacks in the future. Presenting an overview of attacks can be used as a reference for creating policies and strategies to deal with them. Policies are needed to minimize the risks resulting from cyber attacks. Strategies are needed to ensure cyber security so that the number of cyber attacks can continue to be reduced. Policies and strategies can be adjusted based on predicted times. Timeliness is a solution to improving cyber security.

IV. CONCLUSION

Predicting the number of cyber attacks on 3 datasets, each of which has 3 types of predictions with time-series data, can be done using a deep learning-based LSTM model. The best prediction model is epoch 20, batch size 16, and neuron 32 with the lowest MSLE, MSE, MAE, RMSE, and MAPE evaluation values which are less than 10 compared to other variations. Prediction results on training and testing data improve with more training carried out. Negative correlation exists for INTRUSION-MALWARE, BLOCKED-IGNORED, IGNORED-LOGGED, and LOW-MEDIUM, apart from that it has a positive and weak correlation. The proposed prediction model makes predictions 12 months later for 3 types of predictions on each dataset simultaneously and can increase starting from the second month. The resulting predictions can be used as a basis for creating policies and strategies by stakeholders in handling fluctuations in cyber attacks that occur. Comparison of the LSTM prediction model with other models for predicting time series data is a step in finding the best modeling in future work.

ACKNOWLEDGEMENT

This work was supported by Politeknik Keselamatan Transportasi Jalan and Satya Wacana Christian University.

REFERENCES

[1] J. K. Jain and A. A. Waoo, "An Artificial Neural Network Technique for Prediction of Cyber-Attack using Intrusion Detection System," *Journal of Artificial Intelligence, Machine Learning and Neural Network*, no. 32, pp. 33–42, Feb. 2023, doi: 10.55529/jaimlnn.32.33.42.

[2] S. Jacob, Y. Qiao, Y. Ye, and B. Lee, "Anomalous distributed traffic: Detecting cyber security attacks amongst microservices using graph convolutional networks," *Comput Secur*, vol. 118, p. 102728, Jul. 2022, doi: 10.1016/j.cose.2022.102728.

[3] N. El Kamel, M. Eddabbah, Y. Lmoumen, and R. Touahni, “A Smart Agent Design for Cyber Security Based on Honeypot and Machine Learning,” *Security and Communication Networks*, vol. 2020, pp. 1–9, Aug. 2020, doi: 10.1155/2020/8865474.
[4] A. Yeboah-Ofori *et al*., “Cyber Threat Predictive Analytics for Improving Cyber Supply Chain Security,” *IEEE Access*, vol. 9, pp. 94318–94337, 2021, doi: 10.1109/ACCESS.2021.3087109.
[5] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, “IntruDTree: A Machine Learning Based Cyber Security Intrusion Detection Model,” *Symmetry (Basel)*, vol. 12, no. 5, p. 754, May 2020, doi: 10.3390/sym12050754.
[6] N. R. Rajalakshmi, S. V. E., C. K. Parameshwari, M. V., and P. M., “Cyber-security attack prediction using cognitive spectral clustering technique based on simulated annealing search,” *Applied and Computational Engineering*, vol. 6, no. 1, pp. 1360–1365, Jun. 2023, doi: 10.54254/2755-2721/6/20230791.
[7] A. O. David and O. O. Oluwasola, “Zero Day Attack Prediction with Parameter Setting Using Bi Direction Recurrent Neural Network in Cyber Security,” *International Journal of Computer …*, vol. 18, no. 3, 2020.
[8] M. A. Rahman, Y. Al-Saggaf, and T. Zia, “A Data Mining Framework to Predict Cyber Attack for Cyber Security,” in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, IEEE, Nov. 2020, pp. 207–212. doi: 10.1109/ICIEA48937.2020.9248225.
[9] C. Sun, H. Hu, Y. Yang, and H. Zhang, “Prediction method of 0day attack path based on cyber defense knowledge graph,” *Chinese Journal of Network and Information Security*, vol. 8, no. 1, 2022, doi: 10.11959/j.issn.2096-109x.2021101.
[10] P. Datta, N. Lodinger, A. S. Namin, and K. S. Jones, “Predicting Consequences of Cyber-Attacks,” in *2020 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2020, pp. 2073–2078. doi: 10.1109/BigData50022.2020.9377825.
[11] A. Bilen and A. B. Özer, “Cyber-attack method and perpetrator prediction using machine learning algorithms,” *PeerJ Comput Sci*, vol. 7, p. e475, Apr. 2021, doi: 10.7717/peerj-cs.475.
[12] I. M. Kosmacheva *et al*., “Predicting of cyber attacks on critical information infrastructure,” *J Phys Conf Ser*, vol. 2091, no. 1, p. 012062, Nov. 2021, doi: 10.1088/1742-6596/2091/1/012062.
[13] A. M. AL-Hawamleh, “Predictions of Cybersecurity Experts on Future Cyber-Attacks and Related Cybersecurity Measures,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023, doi: 10.14569/IJACSA.2023.0140292.
[14] A. Abu Bakar and M. F. Zolkipli, “Cyber Security Threats and Predictions: A Survey,” *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 5, no. 2, 2023, doi: 10.35629/5252-0502733741.
[15] N. Polatidis, E. Pimenidis, M. Pavlidis, S. Papastergiou, and H. Mouratidis, “From product recommendation to cyber-attack prediction: generating attack graphs and predicting future attacks,” *Evolving Systems*, vol. 11, no. 3, pp. 479–490, Sep. 2020, doi: 10.1007/s12530-018-9234-z.
[16] S. Altalhi and A. Gutub, “A survey on predictions of cyber-attacks utilizing real-time twitter tracing recognition,” *J Ambient Intell Humaniz Comput*, vol. 12, no. 11, pp. 10209–10221, Nov. 2021, doi: 10.1007/s12652-020-02789-z.
[17] H. Albasheer *et al*., “Cyber-Attack Prediction Based on Network Intrusion Detection Systems for Alert Correlation Techniques: A Survey,” *Sensors*, vol. 22, no. 4, p. 1494, Feb. 2022, doi: 10.3390/s22041494.
[18] P. S. Prabha and S. M. Kumar, “A Novel Cyber-attack Leads Prediction System using Cascaded R2CNN Model,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 2, 2022, doi: 10.14569/IJACSA.2022.0130260.
[19] Prof. A. Narote, V. Zutshi, A. Potdar, and R. Vichare, “D-Dos Attack Prediction Using Machine Learning Algorithms,” *Int J Res Appl Sci Eng Technol*, vol. 10, no. 3, pp. 2303–2312, Mar. 2022, doi: 10.22214/ijraset.2022.41131.
[20] S. Srinivasan and P. Deepalakshmi, “ENetRM: ElasticNet Regression Model based malicious cyber-attacks prediction in real-time server,” *Measurement: Sensors*, vol. 25, p. 100654, Feb. 2023, doi: 10.1016/j.measen.2022.100654.
[21] M. Zuzcák and P. Bujok, “Using honeynet data and a time series to predict the number of cyber attacks,” *Computer Science and Information Systems*, vol. 18, no. 4, pp. 1197–1217, 2021, doi: 10.2298/CSIS200715040Z.
[22] J. Zhao *et al*., “Automatically predicting cyber attack preference with attributed heterogeneous attention networks and transductive learning,” *Comput Secur*, vol. 102, p. 102152, Mar. 2021, doi: 10.1016/j.cose.2020.102152.
[23] A. H. Matey, P. Danquah, and G. Y. Koi-Akrofi, “Predicting Cyber-Attack using Cyber Situational Awareness: The Case of Independent Power Producers (IPPs),” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022, doi: 10.14569/IJACSA.2022.0130181.
[24] A. Agur and U. Venugopal, “Cyber Security Attacks.”
[25] M. Hasan, A. Al-Maliki, and N. Jasim, “Review of SQL injection attacks: Detection, to enhance the security of the website from client-side attacks,” *International Journal of Nonlinear Analysis and Applications*, vol. 13, no. 1, 2022, doi: 10.22075/ijnaa.2022.6152.
[26] O. Alkadi, N. Moustafa, B. Turnbull, and K.-K. R. Choo, “A Deep Blockchain Framework-Enabled Collaborative Intrusion Detection for Protecting IoT and Cloud Networks,” *IEEE Internet Things J*, vol. 8, no. 12, pp. 9463–9472, Jun. 2021, doi: 10.1109/JIOT.2020.2996590.

[27] N. Zhang, S.-L. Shen, A. Zhou, and Y.-F. Jin, "Application of LSTM approach for modelling stress–strain behaviour of soil," *Appl Soft Comput*, vol. 100, p. 106959, Mar. 2021, doi: 10.1016/j.asoc.2020.106959.
[28] R. M. Alguliyev, R. M. Aliguliyev, and F. J. Abdullayeva, "The Improved LSTM and CNN Models for DDoS Attacks Prediction in Social Media," *International Journal of Cyber Warfare and Terrorism*, vol. 9, no. 1, pp. 1–18, Jan. 2019, doi: 10.4018/IJCWT.2019010101.
[29] N. Singh, P. Sharma, N. Kumar, and M. Sreejeth, "Short-Term Load Forecasting Using Artificial Neural Network and Time Series Model to Predict the Load Demand for Delhi and Greater Noida Cities," in *Lecture Notes in Networks and Systems*, vol. 177 LNNS, 2021, pp. 443–455. doi: 10.1007/978-981-33-4501-0_41.
[30] Y. Liu, W. Zhang, Y. Yan, Z. Li, Y. Xia, and S. Song, "An Effective Rainfall–Ponding Multi-Step Prediction Model Based on LSTM for Urban Waterlogging Points," *Applied Sciences*, vol. 12, no. 23, p. 12334, Dec. 2022, doi: 10.3390/app122312334.
[31] M. S. Devi, S. Basheer, and R. M. Mathew, "Exploration of Multiple Linear Regression with Ensembling Schemes for Roof Fall Assessment using Machine Learning," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 134–139, Oct. 2019, doi: 10.35940/ijitee.L3474.1081219.
[32] M. S. Devi, R. M. Mathew, and R. Suguna, "Regressor Fitting Of Feature Importance For Customer Segment Prediction With Ensembling Schemes Using Machine Learning," *Int J Eng Adv Technol*, vol. 8, no. 6, pp. 952–956, Aug. 2019, doi: 10.35940/ijeat.F8255.088619.
[33] J. Liao, Y. Liang, and J. Pan, "Deep facial spatiotemporal network for engagement prediction in online learning," *Applied Intelligence*, vol. 51, no. 10, pp. 6609–6621, Oct. 2021, doi: 10.1007/s10489-020-02139-8.
[34] Q. Wang, Y. Wei, C. Zhu, and K. Tian, "Research on Traffic Accident Risk Prediction Based on Spatio-Temporal Graph Convolutional Network," *Jisuanji Gongcheng/Computer Engineering*, vol. 48, no. 11, 2022, doi: 10.19678/j.issn.1000-3428.0062961.
[35] J. Song, L. Zhang, G. Xue, Y. Ma, S. Gao, and Q. Jiang, "Predicting hourly heating load in a district heating system based on a hybrid CNN-LSTM model," *Energy Build*, vol. 243, p. 110998, Jul. 2021, doi: 10.1016/j.enbuild.2021.110998.
[36] İ. Kırbaş, A. Sözen, A. D. Tuncer, and F. Ş. Kazancıoğlu, "Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches," *Chaos Solitons Fractals*, vol. 138, 2020, doi: 10.1016/j.chaos.2020.110015.
[37] T. Li, M. Hua, and X. Wu, "A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5)," *IEEE Access*, vol. 8, pp. 26933–26940, 2020, doi: 10.1109/ACCESS.2020.2971348.
[38] D. Fan, H. Sun, J. Yao, K. Zhang, X. Yan, and Z. Sun, "Well production forecasting based on ARIMA-LSTM model considering manual operations," *Energy*, vol. 220, p. 119708, Apr. 2021, doi: 10.1016/j.energy.2020.119708.
[39] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin, "An LSTM-based aggregated model for air pollution forecasting," *Atmos Pollut Res*, vol. 11, no. 8, pp. 1451–1463, Aug. 2020, doi: 10.1016/j.apr.2020.05.015.
[40] Y. Liu, W. Duan, L. Huang, S. Duan, and X. Ma, "The input vector space optimization for LSTM deep learning model in real-time prediction of ship motions," *Ocean Engineering*, vol. 213, p. 107681, Oct. 2020, doi: 10.1016/j.oceaneng.2020.107681.
[41] J.-Y. Wu, M. Wu, Z. Chen, X.-L. Li, and R. Yan, "Degradation-Aware Remaining Useful Life Prediction With LSTM Autoencoder," *IEEE Trans Instrum Meas*, vol. 70, pp. 1–10, 2021, doi: 10.1109/TIM.2021.3055788.
[42] S. Al-Janabi, M. Mohammad, and A. Al-Sultan, "A new method for prediction of air pollution based on intelligent computation," *Soft comput*, vol. 24, no. 1, pp. 661–680, Jan. 2020, doi: 10.1007/s00500-019-04495-1.
[43] B. Du *et al*., "Deep Irregular Convolutional Residual LSTM for Urban Traffic Passenger Flows Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 972–985, Mar. 2020, doi: 10.1109/TITS.2019.2900481.
[44] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 6910–6920, Nov. 2021, doi: 10.1109/TITS.2020.2997352.
[45] S. Liyanage, R. Abduljabbar, H. Dia, and P.-W. Tsai, "AI-based neural network models for bus passenger demand forecasting using smart card data," *Journal of Urban Management*, vol. 11, no. 3, pp. 365–380, Sep. 2022, doi: 10.1016/j.jum.2022.05.002.
[46] L. Liu *et al*., "Transient rotor angle stability prediction method based on SVM and LSTM network," *Dianli Zidonghua Shebei/Electric Power Automation Equipment*, vol. 40, no. 2, 2020, doi: 10.16081/j.epae.202001009.

nes-3

ORIGINALITY REPORT

16%	10%	14%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	Feifan Jia, Haiying Li, Xi Jiang, Xinyue Xu. "Deep learning-based hybrid model for short-term subway passenger flow prediction using automatic fare collection data", IET Intelligent Transport Systems, 2019 Publication	1%
3	Md. Arif Istiake Sunny, Mirza Mohd Shahriar Maswood, Abdullah G. Alharbi. "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model", 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020 Publication	1%
4	I M Kosmacheva, N V Davidyuk, SV Belov, Yu Kuchin, I Yu Kvyatkovskaya, M F Rudenko, V I Lobeyko. "Predicting of cyber attacks on critical information infrastructure", Journal of Physics: Conference Series, 2021 Publication	1%

5	Stephen Jacob, Yuansong Qiao, Yuhang Ye, Brian Lee. "Anomalous Distributed Traffic: Detecting Cyber Security Attacks amongst Microservices using Graph Convolutional Networks", Computers & Security, 2022 Publication	1%
6	www.hindawi.com Internet Source	1%
7	Md Anisur Rahman, Yeslam Al-Saggaf, Tanveer Zia. "A Data Mining Framework to Predict Cyber Attack for Cyber Security", 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2020 Publication	1%
8	www.researchgate.net Internet Source	<1%
9	Submitted to Queen's College Student Paper	<1%
10	"Setting the Optimum Time for a Special Audit to Improve the Enterprise's Cyber Security", International Journal of Engineering and Advanced Technology, 2020 Publication	<1%
11	mspace.lib.umanitoba.ca Internet Source	<1%
12	peerj.com Internet Source	

<1 %

13 Lai, Yu-Kuen, Chung-Chie Lee, Bo-Hsun Huang, Theophilus Wellem, Nan-Cheng Wang, Tze-Yu Chou, and Hargyo Tri Nugroho. "Real-time detection of changes in network with OpenFlow based on NetFPGA implementation", Microprocessors and Microsystems, 2014.
Publication
<1 %

14 web.archive.org
Internet Source
<1 %

15 www1.rmit.edu.au
Internet Source
<1 %

16 Elissa Penfound, Eric Vaz. "Analysis of Wetland Landcover Change in Great Lakes Urban Areas Using Self-Organizing Maps", Remote Sensing, 2021
Publication
<1 %

17 www.econstor.eu
Internet Source
<1 %

18 Martin Husak, Jana Komarkova, Elias Bou-Harb, Pavel Celeda. "Survey of Attack Projection, Prediction, and Forecasting in Cyber Security", IEEE Communications Surveys & Tutorials, 2018
Publication
<1 %

19	Preuss, Stephanie, and Wolfgang Stein. "Comparison of Two Voltage-Sensitive Dyes and Their Suitability for Long-Term Imaging of Neuronal Activity", PLoS ONE, 2013. Publication	<1 %
20	Submitted to Van Lang University Student Paper	<1 %
21	ici2016.org Internet Source	<1 %
22	Lidong Zhang, Jiao Li, Heng Zhang, Tianyu Hu, Yuanjun Guo, Zhile Yang, Xiandong Xu. "Chapter 7 Study on the Prediction of Indoor Temperature and Thermal Load on Different Floors in a Community", Springer Science and Business Media LLC, 2021 Publication	<1 %
23	Liming Tao, Renbo Xia, Jibin Zhao, Tao Zhang, Yinghao Li, Yueling Chen, Shengpeng Fu. "A high-accuracy circular hole measurement method based on multi-camera system", Measurement, 2023 Publication	<1 %
24	ijeces.ferit.hr Internet Source	<1 %
25	journalofbigdata.springeropen.com Internet Source	<1 %

26	journalofcloudcomputing.springeropen.com Internet Source	<1 %
27	jurnal.ugm.ac.id Internet Source	<1 %
28	link.springer.com Internet Source	<1 %
29	Andri Pranolo, Yingchi Mao, Aji Prasetya Wibawa, Agung Bella Putra Utama, Felix Andika Dwiyanto. "Robust LSTM With Tuned-PSO and Bifold-Attention Mechanism for Analyzing Multivariate Time-Series", IEEE Access, 2022 Publication	<1 %
30	Yan-Hui Lin, Liang Chang. "An Unsupervised Noisy Sample Detection Method for Deep Learning-Based Health Status Prediction", IEEE Transactions on Instrumentation and Measurement, 2022 Publication	<1 %
31	jurnal.borneo.ac.id Internet Source	<1 %
32	rgu-repository.worktribe.com Internet Source	<1 %
33	"Proceedings of 6th International Conference on Recent Trends in Computing", Springer Science and Business Media LLC, 2021	<1 %

	Publication	
34	J. Pineda-Jaramillo, F. Bigi, T. Bosi, F. Viti, A. D'Ariano. "Short-term arrival delay time prediction in freight rail operations using data-driven models", IEEE Access, 2023 Publication	<1 %
35	Prasanjit Dey, Soumyabrata Dev, Bianca Schoen Phelan. " CombineDeepNet: A Deep Network for Multi-Step Prediction of Near-Surface PM Concentration ", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023 Publication	<1 %
36	e-sciencecentral.org Internet Source	<1 %
37	jit.ndhu.edu.tw Internet Source	<1 %
38	openaccess.altinbas.edu.tr Internet Source	<1 %
39	unsworks.unsw.edu.au Internet Source	<1 %
40	www.comsis.org Internet Source	<1 %
41	www.igi-global.com Internet Source	<1 %

42 Chang-Min Lee, Byung-Gun Jung, Jae-Hyuk Choi. "Experimental Study on Prediction for Combustion Optimal Control of Oil-Fired Boilers of Ships Using Color Space Image Feature Analysis and Support Vector Machine", Journal of Marine Science and Engineering, 2023 <1%
Publication

43 Juan Pineda-Jaramillo, Federico Bigi, Tommaso Bosi, Francesco Viti, Andrea D'ariano. "Short-Term Arrival Delay Time Prediction in Freight Rail Operations Using Data-Driven Models", IEEE Access, 2023 <1%
Publication

44 Rasim M. Alguliyev, Ramiz M. Aliguliyev, Fargana J. Abdullayeva. "Deep Learning Method for Prediction of DDoS Attacks on Social Media", Advances in Data Science and Adaptive Analysis, 2019 <1%
Publication

45 Gaojie Fan, Pengyao Yu, Qiang Wang, Yongkang Dong. "Short-term motion prediction of a semi-submersible by combining LSTM neural network and different signal decomposition methods", Ocean Engineering, 2023 <1%
Publication

46	Sathiyandrakumar Srinivasan, P. Deepalakshmi. "ENetRM: ElasticNet Regression Model based malicious cyber-attacks prediction in real-time server", Measurement: Sensors, 2022 Publication	<1%
47	Sepp Hochreiter, Jürgen Schmidhuber. "Long Short-Term Memory", Neural Computation, 1997 Publication	<1%
48	Sohani Liyanage, Rusul Abduljabbar, Hussein Dia, Pei-Wei Tsai. "AI-based neural network models for bus passenger demand forecasting using smart card data", Journal of Urban Management, 2022 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		