

Prediction of Electrical Energy Consumption Using LSTM Algorithm with Teacher Forcing Technique

Sasmitoh Rahmad Riady^{1*)}, Tjong Wan Sen²

^{1,2}Information Technology, Faculty of Computing, President University
Email: ¹sasmitohrr@student.president.ac.id, ²wansen@president.ac.id

Abstract – Electrical energy is an important foundation in world economic growth, therefore it requires an accurate prediction in predicting energy consumption in the future. The methods that are often used in previous research are the Time Series and Machine Learning methods, but recently there has been a new method that can predict energy consumption using the Deep Learning Method which can process data quickly for training and testing. In this research, the researcher proposes a model and algorithm which contained in Deep Learning, that is Multivariate Time Series Model with LSTM Algorithm and using Teacher Forcing Technique for predicting electrical energy consumption in the future. Because Multivariate Time Series Model and LSTM Algorithm can receive input with various conditions or seasons of electrical energy consumption. Teacher Forcing Technique is able to lighten up the computation so that it can train and test data quickly. The method used in this study is to compare Teacher Forcing LSTM with Non-Teacher Forcing LSTM in Multivariate Time Series model using several activation functions that produce significant differences. TF value of RMSE 0.006, MAE 0.070 and Non-TF has RMSE and MAE values of 0.117 and 0.246. The value of the two models is obtained from Sigmoid Activation and the worst value of the two models is in the Softmax activation function, with TF values is RMSE 0.423, MAE 0.485 and Non-TF RMSE 0.520, MAE 0.519.

Keywords – Multivariate Time Series, Deep Learning, Teacher Forcing, LSTM.

I. INTRODUCTION

Energy is an important foundation for economic development in a country [1] and electricity is one of the main energy sources [2]. Therefore, energy policy for a country is very important, because it not only helps the development of the country but also affects the environment both in the field of industrial operations and in the realm of low-income to elite residential areas. Due to the large amount of capital investment and the length of time it took to expand the capacity of electrical energy projects. Therefore, a good estimate or prediction is one of the requirements in the development of a more effective energy policy, because it can reduce the possibility of errors in electrical system planning. Therefore, producing an accurate forecast of electricity consumption is very important [3].

In recent years, many studies have used techniques in predicting electrical energy consumption, either using machine learning [4] or deep learning [5]. In a study conducted by Karimbatar et al regarding data mining for energy consumption by comparing several algorithms in machine learning with Regression models [6], Neural Network [7], and SVM [8] and taking the best model obtained by the Regression model with a relative error of 0.9%, the output of this prediction shows that the average electricity consumption rate increases by about 3.2% per year and will reach 7,076,796 MW in 2020 from a population growth of 22.28% [9]. In another study for the realm of machine learning conducted by Choi regarding the energy consumption analysis for homes using the K-means clustering algorithm from the data generated for K-7 with a silhouette score of 0.799 [10] Another study was also conducted by Nallathambin et al to predict consumption. The electrical energy that will be applied to the USA using the Decision Tree and Random Forest algorithms and experimental solutions states that the RF model provides better accuracy, namely 95.78% than the DT model with an

accuracy of 91.6% and each model has an error rate 0.197 and 0.906 [11].

But at this time there is a new method, namely the Deep Learning method, which can process training quickly, at this time there are also many researchers conducting research on predictions using the method as reflected by Kim et al regarding the prediction of household electricity consumption using CNN-LSTM Hybrid Network [12], the proposed method can be quickly and accurately in predicting irregular energy consumption trends in the dataset of household power consumption. However, because the proposed method was processed earlier by the sliding window algorithm [13], this caused a prediction delay in the actual data [14], in other studies carried out by Young-Jun in electrical energy forecasting By comparing the models contained in the Deep Learning [15] including the LSTM, Gru, and SEQ2SEQ models with the results of the LSTM experiment get the best results with RMSE 0.96 [16], but this value is not good enough to use the actual data to use seasonal data features and Long term in forecasting more accurate electrical energy. Therefore, the researcher will propose a multivariate time series model [17] using the LSTM algorithm as a model and electrical energy prediction algorithm and Teacher Forcing Technique [18] to help in long-term predictions using public consumption datasets taken from the Smart Meters in London Some conditions or seasons. Because the Multivariate Time Series LSTM model algorithm can combine several input to training and testing and produce an output, therefore from various conditions or seasons for the consumption of electrical energy in the dataset will be used as input and will produce an output, namely predictions in the future Come accurately and precisely [19] But the algorithm has weaknesses in the long-term prediction because of the high computing side, then the teacher forcing will help in the long-term predictions because the algorithm can train repetitive networks quickly and efficiently due to output from Repeated LSTM will be used as a subsequent input so



that it will produce low computing using the basic truth from the previous time step as input [20]. With a dataset that the multivariate time series model and the algorithm can reduce RMSE [21] and can predict the consumption of electricity in the long term.

From the results of the training data, a comparison will be made with several activation functions such as ReLu, Softmax [22], Sigmoid [23] and the newest activation function found in RNN, namely Swish [24].

II. RESEARCH METHODOLOGY

At this research stage, there are steps taken by researchers including using literature studies, data analysis using data visualization techniques [25], then using preprocessing data [26] where the data that has been collected and processed is returned to be entered into the model architecture then the result data will be trained to get the results of a prediction of electrical energy consumption, as shown in Figure 1 below.

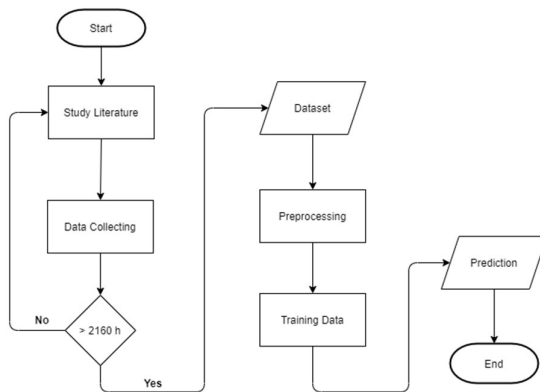


Figure 1. Flow Research Stages

A. Data Collecting

At this stage are the steps in data collection to be used as a dataset of several seasons that are correlated with data on electrical energy consumption for each housing block. The following is the flow of the data collection stages as in Figure 2 below.

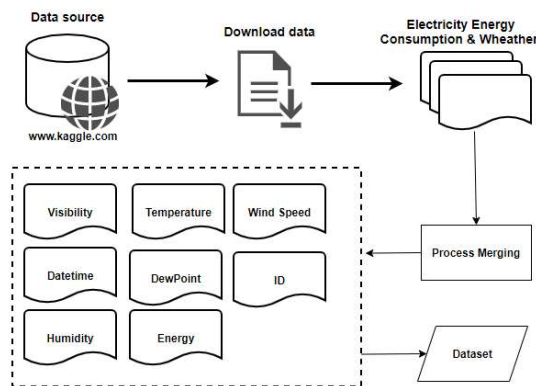


Figure 2. Data Collecting

As in Figure 2, where electricity consumption data is taken from public data, namely www.kaggle.com about smart meter data in London, data is downloaded by 2 GB and

analyzed, the data contains several folders and files. DataSet represents the reading of household electrical energy consumption in London in Kilowatt Hour (KWH) from November 2011 to February 2014. The file series is categorized into 112 blocks (block 0 to block 111) in the original dataset. In addition to electrical energy consumption data there are several files, including Acorn Details, Information Household, UK Holiday Bank, Weather Daily and Weather per hour. For this study we combine electrical energy consumption data with weather per hour for one housing block. The dataset will be trained is 2903 with variables datetime, visibility, temperature, dew point, pressure, windspeed, humidity, ID, and energy.

B. Training Model

The model that was built using the LSTM Multivariate Time Series Model with Teacher Forcing Technique, before going to the process, this study will propose to do some comparisons including comparing the LSTM algorithm with Non-Teacher Forcing and Teacher Forcing [27], to find which performance loss is better, then the results of the comparison will be continued using the Multivariate Time Series model and comparisons of several activation functions [28] contained in the Deep Learning method, the following is the flow of the training model that will be used in this study, as shown in Figure 3.

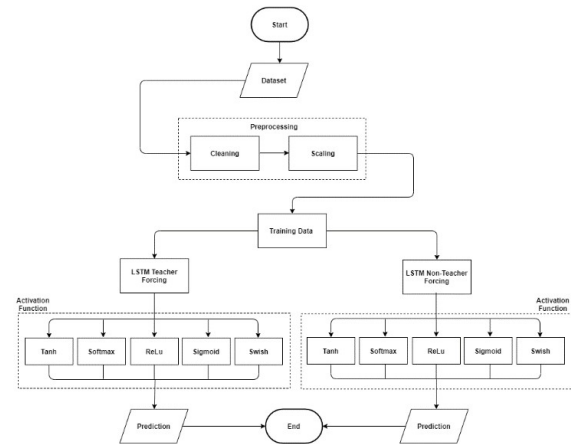


Figure 3. Training Model

C. Architecture Model

The following is an architectural model using a comparison between LSTM Non-Teacher Forcing and LSTM Teacher Forcing in order to find the best performance loss of the two models, then the model will be entered into the Multivariate Time Series model, then it will be compared with several activations to find RMSE results. better. As in Figure 4 and Figure 5.

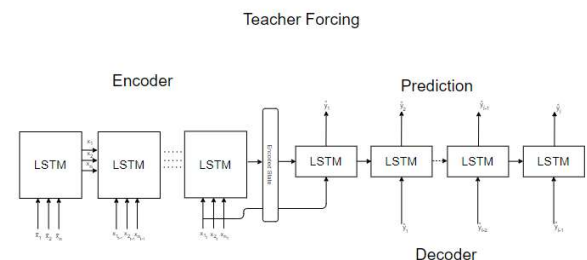


Figure 4. LSTM Teacher Forcing

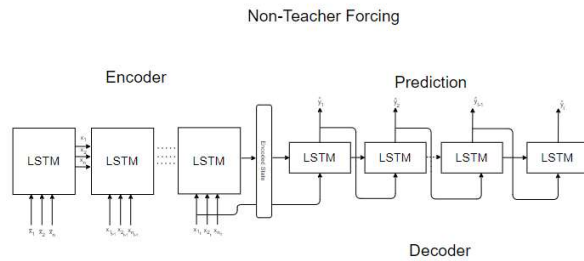


Figure 5. LSTM Non-Teacher Forcing

D. Activation Function

From the results of the eating model architecture, several activations will be compared, including sigmoid activation, ReLu, Sigmoid and Custom Activation. Development of the Sigmoid Activation, this activation functions to make the Neural Network non-linear [29]. Sigmoid will accept a single number and convert the x value into a value that has a range from 0 to 1, which has the following formula.

$$S(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

And next is the ReLU or Rectified Linear Unit activation which is a pretty good activation function because ReLU greatly accelerates the convergence process carried out with stochastic gradient descent when compared to sigmoid/tanh with the following formula.

$$f(x) = \max(0, x) \quad (2)$$

Because ReLU basically only creates a delimiter on the number zero, meaning that if $x \leq 0$ then $x = 0$ and if $x > 0$

then $x = x$. Their experiment shows that Swish [30] tends to perform better than ReLu on deeper models across a number of challenging data sets with the following formula.

$$f(x) = x \cdot \text{sigmoid}(x) \quad (3)$$

E. Measurement Step

experiment carried out which performance is better. Measurement using RMSE and MAE. Root Mean Square Error (RMSE) is the sum of the squared error or the difference between the true value and the predetermined predictive value. With the RMSE.

$$RMSE = \sqrt{\frac{\sum (Y' - Y)^2}{n}} \quad (4)$$

Mean Absolute Error (MAE) shows the mean error value which is the error of the true value with the predicted value. MAE itself is generally used for measuring error prediction in time series analysis. The formula for MAE itself is defined as follows:

$$MAE = \frac{|Y' - Y|}{n} \quad (5)$$

hat formula (5) shows that Y' is Prediction Value, Y is Actual Value, and n is Total of Data.

III. RESULTS AND DISCUSSION

A. Preprocessing

At this stage the data will be merged to form multivariate time series data so that data from several files such as weather and energy can be combined into a data file. The following is an example of a dataset that has been merged.

Table 1. Data Merging

visibility	temperature	datetime	dewpoint	pressure	Windspeed	humidity	id	energy
14.31	6.86	2013-10-29T23:00:00Z	4.61	1017.98	3.29	0.86	MAC000002	0.457
14.31	6.1	2013-10-30T00:00:00Z	4.09	1018.38	2.95	0.87	MAC000002	0.414
13.45	5.94	2013-10-30T01:00:00Z	4.17	1018.89	3.14	0.88	MAC000002	0.408
13.23	5.54	2013-10-30T02:00:00Z	3.92	1019.25	3.02	0.89	MAC000002	0.352
14.31	5.06	2013-10-30T03:00:00Z	3.28	1019.57	2.74	0.88	MAC000002	0.25
14.31	5.04	2013-10-30T04:00:00Z	3.24	1019.84	2.77	0.88	MAC000002	0.217
14.31	4.81	2013-10-30T05:00:00Z	3.29	1020.25	2.16	0.9	MAC000002	0.211
...
12.23	8.69	2014-02-27T17:00:00Z	1.55	1006.43	4.11	0.61	MAC000002	0.667
12.41	7.22	2014-02-27T18:00:00Z	2.16	1006.68	3.43	0.7	MAC000002	1.7
12.52	6.22	2014-02-27T19:00:00Z	2.07	1006.88	2.99	0.75	MAC000002	2.259
14.03	5.94	2014-02-27T20:00:00Z	2.07	1006.74	3.25	0.76	MAC000002	1
16.09	5.03	2014-02-27T21:00:00Z	1.67	1006.36	3.06	0.79	MAC000002	1.766
14	4.1	2014-02-27T22:00:00Z	1.64	1005.67	3.02	0.84	MAC000002	2.465



The following is a description of the attributes resulting from the merging process, namely Time, Visibility, Temperature, Dew Point, Pressure, Wind Speed, Humidity, ID and Energy. Produces 2904 raw data or 2904 hours of data. Then proceed with the indexing, cleaning, and scaling data stages. Table 2 is an example of the final preprocessing stage.

B. Teacher Forcing and Non-Teacher Forcing with Univariate Time Series Model

Figure 4 and Figure 5 have discussed the model architecture for teacher forcing and non-teacher forcing, the data that will be used first is the univariate time series data where the attributes to be learned are the data frame index and energy which have been changed from per hour to per minute, at this stage the data will be learned using an ADAM optimization of 15 epochs and a batch size of 100, here are the results of the model that has been trained.

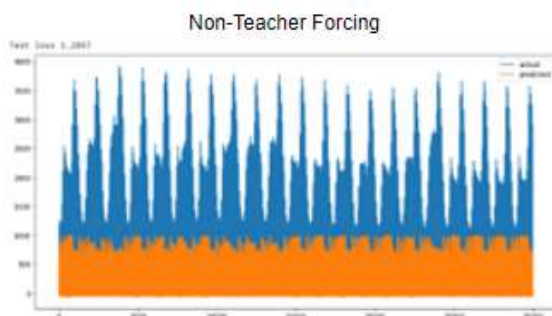


Figure 6. Teacher Forcing Univariate Model

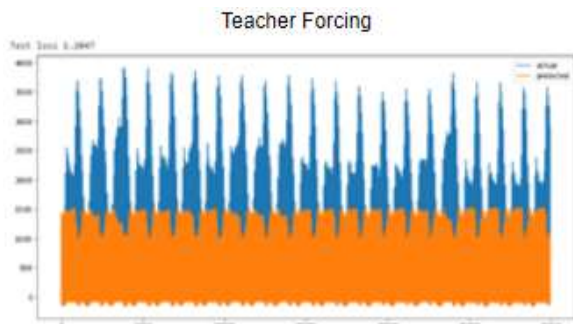


Figure 7. Non-Teacher Forcing Univariate Model

From the results of training and testing, the LSTM model using the Teacher Forcing technique is better for performance loss and prediction on the univariate model, next is training on electrical energy consumption data with the Multivariate Time Series model using Teacher Forcing and Non-Teacher Forcing Techniques

A. Teacher Forcing and Non-Teacher Forcing with Multivariate Time Series Model

At this stage is to conduct training on each attribute that will be used as input to be trained on the model that has been made, the following is a multivariate time series variable model that will be trained on the LSTM model using TF and Non-TF.

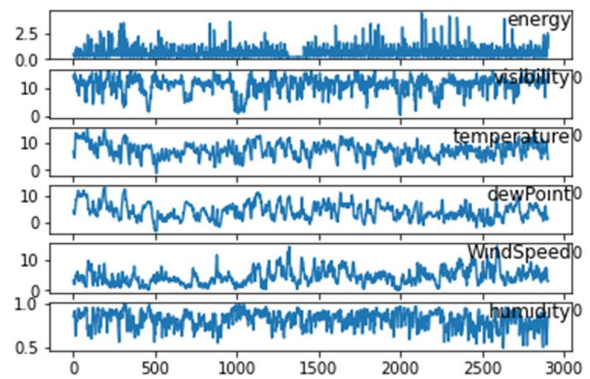


Figure 8. Multivariate Time Series Electricity Consumption

In the previous training, univariate time series data were trained without using an activation function, now multivariate data will be trained using Tanh, Softmax, ReLu, Sigmoid, and Swish activations. The image below is the result of the training.

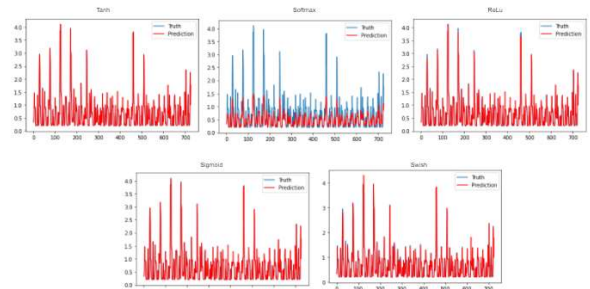


Figure 9. Prediction Result LSTM Teacher Forcing using Activation.

And next is to measure the prediction results for the Non-Teacher Forcing LSTM model as follows.

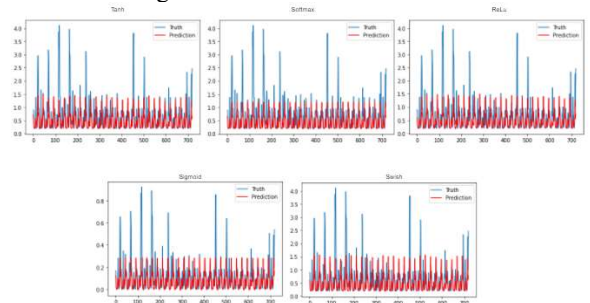


Figure 10. Prediction Result LSTM Non-Teacher Forcing Activation

Table 2. Activation Deferent

Activation	Teacher Forcing		Non-Teacher Forcing	
	RMSE	MAE	RMSE	MAE
Tanh	0.009	0.084	0.510	0.513
Softmax	0.423	0.485	0.520	0.519
ReLu	0.020	0.104	0.508	0.514
Sigmoid	0.006	0.070	0.117	0.246
Swish	0.025	0.117	0.505	0.515



IV. CONCLUSION

The training model in this study uses several activation functions including tanh, Softmax, ReLu, Sigmoid, and swish activation. The results obtained show that Teacher Forcing LSTM is better than Non-Teacher Forcing LSTM in terms of performance loss and prediction which results in quite a significant difference. The TF value is RMSE 0.006, MAE 0.070 and Non-TF itself has RMSE and MAE values of 0.117 and 0.246. The value of the two models is obtained from sigmoid activation and the worst value of the two models is in the Softmax activation function, with TF values namely RMSE 0.423, MAE 0.485 and Non-TF RMSE 0.520, MAE 0.519.

REFERENCES

[1] T. Ula, "Dampak Konsumsi Energi Terbarukan Terhadap Pertumbuhan Ekonomi: Studi di Asia Tenggara," *J. Econ. Sci.*, vol. 5, no. 2, pp. 26–34, 2019.

[2] dkk Ahmad Wahid, "Analisis Kapasitas Dan Kebutuhan Daya Listrik Untuk Menghemat Penggunaan Energi Listrik Di Fakultas Teknik Universitas Tanjungpura," *J. Tek. Elektro UNTAN*, vol. 2, no. 1, p. 10, 2014.

[3] O. Somantri, "Prediksi Kebutuhan Permintaan Energi Listrik Menggunakan Neural Network Berbasis Algoritma Genetika," no. September, pp. 1–135, 2015.

[4] S. Seyedzadeh, F. P. Rahimian, I. Glesk, and M. Roper, "Machine learning for estimation of building energy consumption and performance: a review," *Vis. Eng.*, vol. 6, no. 1, 2018, doi: 10.1186/s40327-018-0064-7.

[5] M. Alanbar, A. Alfarraj, and M. Alghieth, "Energy consumption prediction using deep learning technique case study of computer college," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 10, pp. 166–177, 2020, doi: 10.3991/ijim.v14i10.14383.

[6] B. Shyti and D. Valera, "The Regression Model for the Statistical Analysis of Albanian Economy," *Int. J. Math. Trends Technol.*, vol. 62, no. 2, pp. 90–96, 2018, doi: 10.14445/22315373/ijmtt-v62p513.

[7] B. Cetişli and A. Barkana, "Speeding up the scaled conjugate gradient algorithm and its application in neuro-fuzzy classifier training," *Soft Comput.*, vol. 14, no. 4, pp. 365–378, 2010, doi: 10.1007/s00500-009-0410-8.

[8] A. Abubakar *et al.*, "A support vector machine classification of computational capabilities of 3D map on mobile device for navigation aid," *Int. J. Interact. Mob. Technol.*, vol. 10, no. 3, pp. 4–10, 2016, doi: 10.3991/ijim.v10i3.5056.

[9] N. Karimtabar, S. Pasban, and S. Alipour, "Analysis and predicting electricity energy consumption using data mining techniques - A case study I.R. Iran - Mazandaran province," *2015 2nd Int. Conf. Pattern Recognit. Image Anal. IPRIA 2015*, no. Ipria, pp. 0–5, 2015, doi: 10.1109/PRIA.2015.7161634.

[10] Y. J. Lee and H. J. Choi, "Forecasting building electricity power consumption using deep learning

approach," *Proc. - 2020 IEEE Int. Conf. Big Data Smart Comput. BigComp 2020*, pp. 542–544, 2020, doi: 10.1109/BigComp48618.2020.000-8.

[11] S. Nallathambi and K. Ramasamy, "Prediction of electricity consumption based on DT and RF: An application on USA country power consumption," *Proc. - 2017 IEEE Int. Conf. Electr. Instrum. Commun. Eng. ICEICE 2017*, vol. 2017-Decem, no. 1, pp. 1–7, 2017, doi: 10.1109/ICEICE.2017.8191939.

[12] S. Chan, I. Oktavianti, and V. Puspita, "A Deep Learning CNN and AI-Tuned SVM for Electricity Consumption Forecasting: Multivariate Time Series Data," *2019 IEEE 10th Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2019*, pp. 488–494, 2019, doi: 10.1109/IEMCON.2019.8936260.

[13] H. Lim, Y. Kim, C. Y. Lee, and K. Cheun, "An Efficient Sliding Window Algorithm Using Adaptive-Length Guard Window for Turbo Decoders," *J. Commun. Networks*, vol. 14, no. 2, pp. 195–198, 2012, doi: 10.1109/JCN.2012.6253068.

[14] T. Y. Kim and S. B. Cho, *Predicting the Household Power Consumption Using CNN-LSTM Hybrid Networks*, vol. 11314 LNCS. Springer International Publishing, 2018.

[15] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, no. March, pp. 1–43, 2020, doi: 10.1016/j.physd.2019.132306.

[16] S. Goodman and N. Ding, "TeaForN: Teacher-Forcing with N-grams," pp. 8704–8717, 2020.

[17] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate LSTM-FCNs for time series classification," *Neural Networks*, vol. 116, pp. 237–245, 2019, doi: 10.1016/j.neunet.2019.04.014.

[18] K. Drossos, S. Gharib, P. Magron, and T. Virtanen, "Language modelling for sound event detection with teacher forcing and scheduled sampling," *arXiv*, no. October, 2019, doi: 10.33682/1dze-8739.

[19] A. A. Ismail, T. Wood, and H. C. Bravo, "Improving long-horizon forecasts with expectation-biased LSTM networks," *arXiv*, 2018.

[20] A. Goyal, A. Lamb, Y. Zhang, S. Zhang, A. Courville, and Y. Bengio, "Professor forcing: A new algorithm for training recurrent networks," *Adv. Neural Inf. Process. Syst.*, no. Nips 2016, pp. 4608–4616, 2016.

[21] A. S. Wardana and M. I. Ananta Timur, "Collaborative Filtering Recommender System pada Virtual 3D Kelas Cendekia," *IJEIS (Indonesian J. Electron. Instrum. Syst.)*, vol. 8, no. 1, p. 73, 2018, doi: 10.22146/ijeis.28729.

[22] C. E. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv*, pp. 1–20, 2018.

[23] Julpan, E. B. Nababan, and M. Zarlis, "Analisis



- Fungsi Aktivasi Sigmoid Biner Dan Sigmoid Bipolar Dalam Algoritma Backpropagation Pada Prediksi Kemampuan Siswa,” *J. Teknovasi*, vol. 02, no. 1, pp. 103–116, 2015.
- [24] M. Çelebi and M. Ceylan, “The New Activation Function for Complex Valued Neural Networks: Complex Swish Function,” no. July 2019, pp. 169–173, 2019, doi: 10.36287/setsoci.4.6.050.
- [25] N. Ahmad Syaripul and A. Mukharil Bachtiar, “Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi Dki Jakarta: Topik Ekonomi Dan Keuangan Daerah,” *J. Sist. Inf.*, vol. 12, pp. 15–29, 2016.
- [26] C. V. Gonzalez Zelaya, “Towards explaining the effects of data preprocessing on machine learning,” *Proc. - Int. Conf. Data Eng.*, vol. 2019-April, pp. 2086–2090, 2019, doi: 10.1109/ICDE.2019.00245.
- [27] M. Sangiorgio and F. Dercole, “Robustness of LSTM neural networks for multi-step forecasting of chaotic time series,” *Chaos, Solitons and Fractals*, vol. 139, p. 110045, 2020, doi: 10.1016/j.chaos.2020.110045.
- [28] A. Yadav, C. K. Jha, and A. Sharan, “Optimizing LSTM for time series prediction in Indian stock market,” *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 2091–2100, 2020, doi: 10.1016/j.procs.2020.03.257.
- [29] O. Contribution, “Nonlinear Neural Networks: Principles, Mechanisms, and Architectures,” *Pattern Recognit. by Self-Organizing Neural Networks*, vol. 1, 2020, doi: 10.7551/mitpress/5271.003.0004.
- [30] H. H. Chieng, N. Wahid, O. Pauline, and S. R. Kishore Perla, “Flatten-T Swish: A thresholded ReLU-Swish-like activation function for deep learning,” *arXiv*, vol. 4, no. 2, pp. 76–86, 2018.

