Journal of Theoretical and Applied Information Technology
15th May 2022. Vol.100. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IOT WEIGHT DATA DENOISE BASED ON EXPONENTIAL MOVING AVERAGE AND OUTLIERS

[1]HARRY SURYA, [2]ABBA SUGANDA GIRSANG

[1,2] Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

E-mail: [1]harry.surya@binus.ac.id, [2]agirsang@binus.edu

ABSTRACT

The emergence of Internet of Things (IoT) has delivered a new perspective about how technology can accomplish human needs, IoT primarily used to deliver rapid stream of data from sensors to variety options of storage system and used it to accomplish objective of the development. As a result of that, IoT may occurs an unavoidable challenges of data noise. In this paper we will present our result of data noise removal approach based on data stream of IoT scale machine that we built. Outliers and window functions will be used to pre-processing the raw data of IoT, by applying this approach, we aim to deliver a better data in the pre-processing phase as an effort to eradicate the data noise problem in IoT sensors. IoT device that we built was using NodeMCU microcontroller, HX711 amplifier and 200 kg loadcell. We also comparing Exponential Moving Average (EMA) to other Window Function to deliver an insight for reader of EMA performance across other window function.

Keywords: *Data Noise, IoT, Outliers, Window Function, Exponential Moving Average*

1. INTRODUCTION

IoT provides a data stream that can constantly deliver to the storage system through a variety of communication options such as: Wi-Fi, Bluetooth, etc. Data delivery is designed to get the identification, placement, tracking, analyzation, and administration [1]–[3]. Rapid development of IoT has enabled information of humidifies, temperature, weight, height, and many more to be retrieved from sensors. Data that is delivered from IoT needs to represent the real situation of the actual real-world condition so the next action of the IoT development can be delivered accurately using data from IoT. Carrying these challenges, IoT has developed since its first debut in 1999; IoT Manufacturer has invented more variance of sensors and microcontroller to be able to satisfy human needs of IoT.

Data noise that produced by IoT Sensors is caused by many factors: cabling, environment, and many more. Although the IoT Sensors manufacturer have tried to achieve the best measurement that deliver the real data based on the real world, Improvements are still needed to achieve a better data accuracy. According to several research on the data noise topic, IoT [4]–[11] shows the proof that data noise is a real problem of IoT implementation.

Data noise problem implies as the data being meaningless for further development. Based on our IoT scale machine project, we found a noise that produced by the weight sensor, carrying out the finding, we then deliver our approach using proposed method that based on outliers removal, EMA and averaging to solve the noisy dataset problem from the IoT weight scale machine.

Comparison between EMA and SMA based on analog scale machine will also be delivered in this paper to analyze the performance between smoothing data process, analog scale machine will be used to representing a real-world number that indicate the performance of our proposed method to denoising the IoT dataset.

2. LITERATURE REVIEW

Liu *et al.* [4] analyzing the problem in data noise of Industrial Internet of Things (IoT) by proposes method to remove noise, the method uses outlier to remove data that differ from the majority and then window function used to get the window score for the threshold to clarify noisy data using Q-Q Plot.

Faria *et al.* [5] focused on collection of sensors in Intel Berkeley Research Lab, analyzing several sensors of temperature, humidity, and luminosity. Using Discrete Wavelet Transform (DWT).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Lai *et al.* [6] researching on Air Quality based IoT that using Kalman Filter (KF), Simple Moving Average, and Autoregressive Integrated Moving Average. Based on that three method, KF 68.3% more better in a root mean square error (RMSE) across another algorithm.

Ma *et al.* [7] creating preprocessing of IoT humidity data using Autoregressive Moving Average (ARMA), to detect noise and remove it, the research shows that ARMA able to finish the task to detect noise in data of IoT based on humidity data characteristic.

Zhang [12] proposed a usage of empirical mode decomposition (EMD) and time-domain windowing to alter EMD only method, based on the result shows that the proposed method able to denoise data properly as shown on Figure 1.

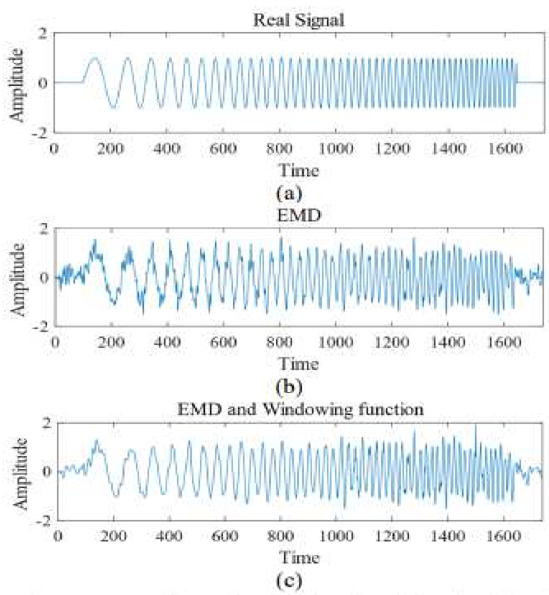

Figure 1: Denoising using EMD and Windowing Function [12].

Uyanik *et al.* [13] implementing Window Functions for Audio Fingerprint based on Audio Signal, the result shown that Hanning Filter able to outperform other window function and researchers also state that window function cannot generally be applicable to all application.

In several research [4]–[7] we conclude that IoT noise data problem is a challenge in IoT implementation. This research used a variance of methods to be able to distinguish real data that represents the actual information and represents the noise that is delivered from the sensors. Carrying this out, we will focus on the Window function to be able to remove noise from IoT Raw Data input.

3. RESEARCH METHODOLOGY

Data denoise in this paper are using outliers and window function to be able to get the useful data as our aim, We analyze the data characteristic of our dataset and understanding the needs of outliers as the first phase of denoise process, The human input process from our IoT scale machine start when user stepping the first foot to the scale machine platform, we then analyze this behavior as the first noise of the data that will be detected using outliers, then window function will be used to smoothing the remaining dataset and finally using averaging to get the single output as our finalized denoised value.

3.1 IoT Architecture of the Research

In this paper, we will be using data from our IoT device. This device was built using NodeMCU as a microcontroller, Hx711 as an amplifier, and Loadcell 200 kg as a force transducer. The IoT device's purpose that we built was to get weight data from users; users are an employee from a factory in Cibinong, West Java, Indonesia. By using our application, the data can identify which employee was doing the scaling process on the IoT device. Given that information, we can then calculate the estimate 2 second of the input process as a raw data input for that specific user. Prototype of the IoT device can be seen in Figure 2.

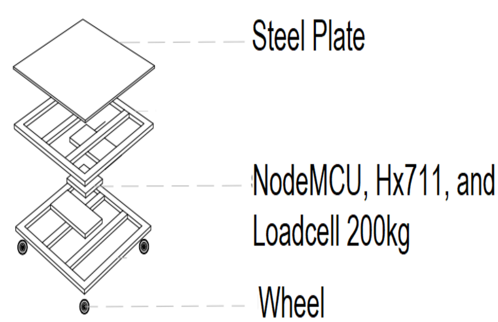

Figure 2: IoT Prototype

Figure 3 described that the sensor and microcontroller are placed in the center of the steel design. Employees will stand on steel plate, and all the data from sensors are delivered from microcontroller using Wi-Fi to our Cloud MySQL Database through API.

3.2 Outliers

Outliers are data instances that detached from most of the instances. This is defined as two perceptions, which is either the data are very useful

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and can be used as further analyzation or the data is misled information that we need to detect and remove. Aligned with this paper, we will define our outlier as misleading that needs to be removed [14]–[16].

Based on our analyzation of user behavior that doing input process of the IoT scale machine is by stepping one foot to the platform and followed by the next foot, given that process we then recognize the pattern of the data stream from IoT. Pattern shows that the first 1-2 second of the beginning and end of the data stream from IoT are noise, The optimal time of data stream that we wanted to capture are between that range that can be seen on Figure 3.

Figure 3: User Interaction to IoT Scale Machine

To capture the optimal data stream that replicate the human body weight, we then propose using outliers to find the outlier in the input data of specific range, and we will delete non-outlier data. We will use Euclidian distance function; the result is shown on Figure 4.

Based on our outlier testing shown on Figure 4, the blue dots contain an outlier and green are not an outlier. We will remove the blue dots before adding it to the Window Function.

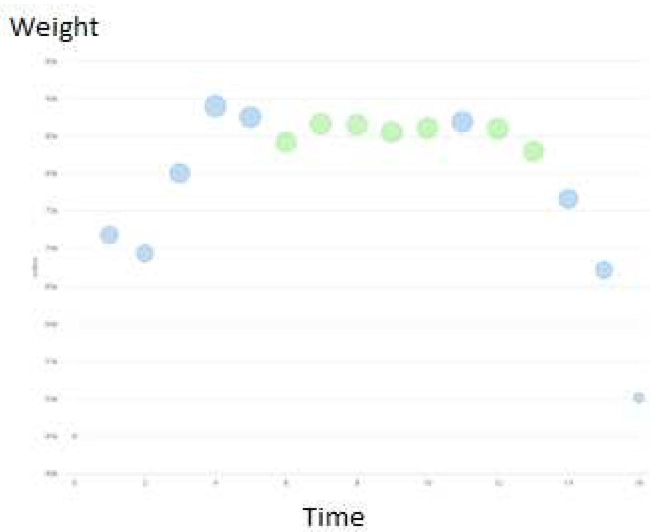

Figure 4: Outlier Detection

3.3 Window Function

In approach of denoise data, window function is considered as a major method to find the local estimate of the constraint, windowing is very robust for small noises but can resulted as an over-smoothing or over-sharpening for a high noise data [17]. In previous literature stated that data window function cannot be generalized because the characteristic of data stream affects the method used to solve the problem [12] In this section we will define the window function that will be used in this paper, such as:

1) Averaging

Averaging is method to sum all the number inserted and then divide it by the total of number inserted [18] this method represents on Equation 1.

$$A = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Equation 1: Averaging

2) Simple Moving Average (SMA)

SMA collecting all the data stream constraint in the given period or condition and all the constraints will be divided based on the amount of data stream [19], [20] SMA math function shown on Equation 2.

$$SMA = \frac{P_M + P_{M-1} + \cdots + P_{M-(n-1)}}{n}$$

Equation 2: Simple Moving Average (SMA)

Pm stands for data value for given time M, and n defining the total amount of data value that stated.

3) Exponential Moving Average (EMA)

EMA giving a weight to data that will be decreased if the data considered old and increase the weight if the data is new; therefore, So EMA is reactive upon data stream [21], [22]. EMA math function shown on Equation 3.

$$EMA_t = \frac{X_t + (1-\alpha)X_{t-1} + (1-\alpha)^2 X_{t-2} + \cdots + (1-\alpha)^t x0}{1 + (1-\alpha) + (1-\alpha)^2 + \cdots + (1-\alpha)^t}$$

Equation 3: Exponential Moving Average (EMA)

3.4 Proposed Method

Observing visualization of our data helps to understand more about our current problem, given example on Figure 5, the data is considered stable in between 90-80 kg range. Our aim is to get the smoothen value in that range as our approach for near to real data of human body weight from the noisy dataset.

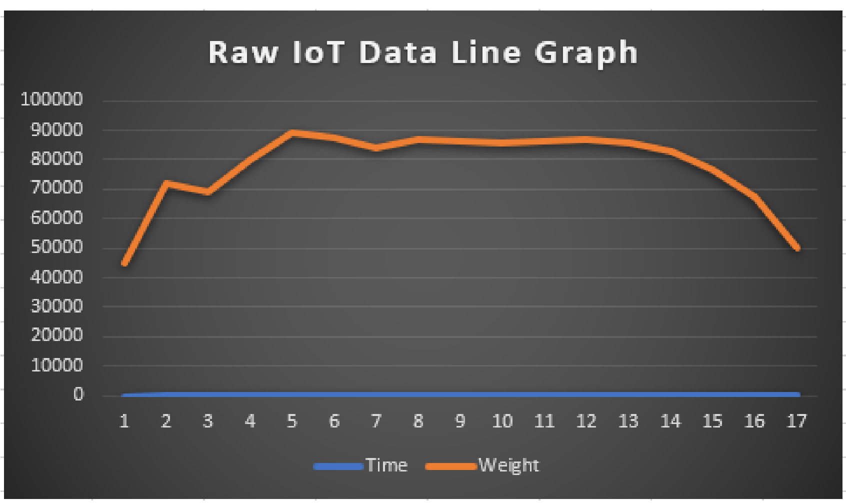

Figure 5: Raw IoT Data Line Graph

Figure 5 shows the data input from an employee. The graph shows that the human weight scaling will always start with a lesser amount of actual data such as 40kg; this data occurs when users try to step on our IoT Scale Machine with one foot only. We proposed that the first problem can be handled by using Outliers. The actual majority of data is distributed across 80kg, and, by using outliers' data, the lesser major will be removed.

After removing all the un-wanted data using outlier, we then use the Window function to smooth the data stream from IoT. We will then use averaging to get single data as a result. In this paper, we will be implementing our proposed method for our IoT Noise problem as shown on Figure 6. Output of this method would be the estimated value of human weight based on our Denoise approach of raw IoT noisy data.

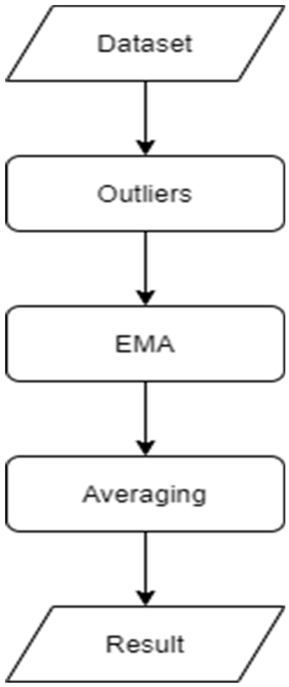

Figure 6: Proposed Method

4. RESULT AND DISCUSSION

After gathering data from employees for a month, we will obtain approximately 300.000 data. Carrying out this dataset, we then extract 1 sample dataset of employee interaction on our IoT device to be tested using our proposed method. Line graph of the raw data from IoT sensor can be seen on Figure 7.

Line graph on Figure 7, shows the sample dataset from sensor are full of noises and not represented by the human body weight of the actual employee. Sample dataset contains 31 rows of sensor input from a single employee interaction to IoT. The blue dots shown on Figure 7, represent the weight value sent by sensor in measurement of grams. We then implement outlier to the dataset as shown on Figure 8.

Based on Figure 8, the red dots are the outliers in our sample dataset. and the blue dots are the non-outliers. Outliers will be removed as we proposed that the outlier does not represent the actual weight data. The number of outliers to find for our experiment will be 75% of total dataset constraint, so, given 31 rows as our sample dataset, we then multiply it with 75% and resulted 23.

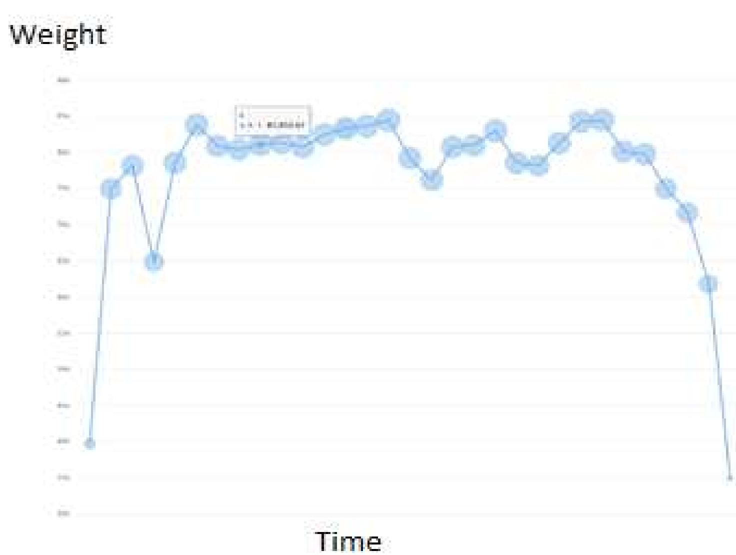

Figure 7: Sample Dataset Graph

Journal of Theoretical and Applied Information Technology
15th May 2022. Vol.100. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

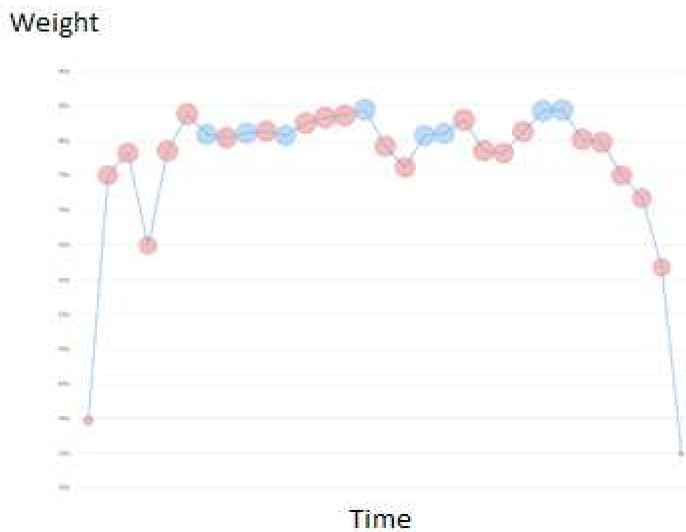

Figure 8: Outlier Detection of Sample Dataset Graph

Sample dataset will exclude all 23 outliers that colored as red dot in Figure 8, We then extract all the remaining data to deliver it to the next phase, The result of outliers removal shown on Table 1.

Table 1: Sample Dataset after Outliers Removal

Employee Name	Value (gram)
USER001	80,860
USER001	81,053
USER001	80,723
USER001	84,432
USER001	80,697
USER001	81,001
USER001	84,248
USER001	84,393

Using sample dataset as shown on Table 1, we then continue to the next phase to implement filtering to reduce the data variance. In this experiment, we will be using SMA and EMA to analyze the performance of each method to handle sample dataset.

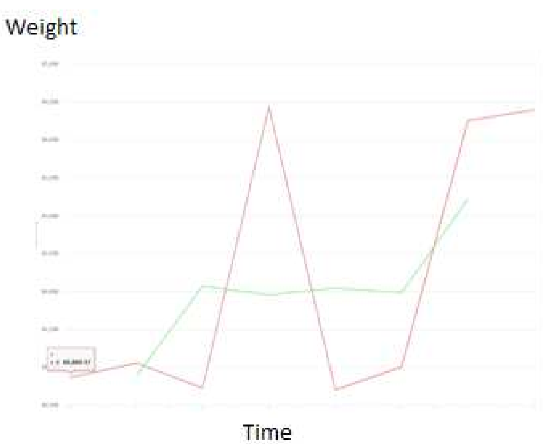

Figure 9: Window Function SMA Result

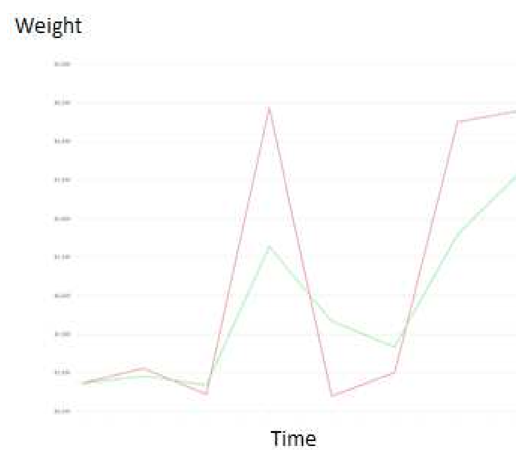

Figure 10: Window Function EMA Result

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Both SMA and EMA are shown as a green line on Figure 9 and Figure 10, while the red line is our sample dataset. The Window function provides smoother dataset compared to previous dataset. We then finalize our results by using averaging method to choose one value as a result and comparing it with actual employee body weight based on analog scale machine as seen on Table 2.

Table 2: Experiment Result

Window Function	Experiment Result (kg)	Body Weight (kg)	Variance
SMA	82	82.2	0.2
EMA	81.8	82.2	0.4

Variance listed on Table 2 shows that the EMA slightly able to replicate actual body weight based on our method compared to SMA. Given this result. we then elevate our research to do 10 experiments using 10 sample datasets to understand more about the performance between SMA and EMA. All the results are listed on Table 3.

Table 3: Advanced Experiment Result

Employee Name	SMA (kg)	EMA (kg)	Body Weight (kg)
USER005	72	72	72.1
USER038	86.4	86.5	86.9
USER070	86.4	86.3	85.5
USER082	86.5	86.4	86.1
USER091	85.4	85.4	85
USER098	82.4	82.4	83.1
USER102	66.4	66.4	66.7
USER105	86.5	86.5	87.1
USER108	82.4	82.4	82
USER112	90.8	90.9	91

We then summarize the variance between SMA to body weight and EMA to body weight based on Table 3, the result shows that SMA cumulative variance is 4.5 kg and EMA cumulative variance 4.1 kg. Average of variance using SMA is 0.82 kg and EMA is 0.75 kg, it concludes that EMA able to outperform SMA in transforming data based on our IoT device dataset.

5. CONCLUSION

In order to solve noisy human body weight dataset from IoT scale machine, we proposed a new method to denoise the dataset based on outliers removal, EMA, and averaging. Outliers removal at the first phase remarkably useful to reduce the noise of the dataset and our simulation to analyze the result between EMA and SMA to smoothing the dataset indicate that EMA able to outperform SMA compared to analog scale machine result.

As stated from previous research [16] that the correct denoising method for handling noisy data from IoT sensors tend to be segmented based on the IoT specification and environment and cannot be generally applicable to the other research, during our simulation we found out that the numbers of outliers that need to be removed are majorly influenced on the human behavior when doing the weighing process, further study to obtain adaptive number of outliers is needed to solve the given issue.

REFRENCES:

[1] K. K. Patel, S. M. Patel, and P. G. Scholar, “Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges,” *International Journal of Engineering Science and Computing*, 2016, doi: https://doi.org/10.4010/2016.1482.

[2] P. Radanliev *et al.*, “Definition of Internet of Things (IoT) Cyber Risk-Discussion on a Transformation Roadmap for Standardization of Regulations, Risk Maturity, Strategy Design and Impact Assessment,” *Preprints*, 2019, doi: https://doi.org/10.20944/preprints201903.0080.v1.

[3] D.-R. Berte, “Defining the IoT,” *Proceedings of the International Conference on Business Excellence*, vol. 12, no. 1, pp.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

118–128, May 2018, doi: https://doi.org/10.2478/picbe-2018-0013.

[4] Y. Liu, T. Dillon, W. Yu, W. Rahayu, and F. Mostafa, "Noise Removal in the Presence of Significant Anomalies for Industrial IoT Sensor Data in Manufacturing," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7084–7096, Aug. 2020, doi: 10.1109/JIOT.2020.2981476.

[5] M. L. L. de Faria, C. E. Cugnasca, and J. R. A. Amazonas, "Insights into IoT data and an innovative DWT-based technique to denoise sensor signals," *IEEE Sensors Journal*, vol. 18, no. 1, pp. 237–247, Jan. 2018, doi: 10.1109/JSEN.2017.2767383.

[6] X. Lai, T. Yang, Z. Wang, and P. Chen, "IoT implementation of Kalman Filter to improve accuracy of air quality monitoring and prediction," *Applied Sciences (Switzerland)*, vol. 9, no. 9, May 2019, doi: 10.3390/app9091831.

[7] Y. Ma, J. Jin, Q. Huang, and F. Dan, "Data Preprocessing of Agricultural IoT Based on Time Series Analysis," in *International Conference on Intelligent Computing*, 2018, pp. 219–230. doi: 10.1007/978-3-319-95930-6_21.

[8] K. , J. Židek, P. J. D., A. Hošovský, and P. Lazorík, "Data optimization for communication between wireless IoT devices and Cloud platforms in production process," *MMS 2018: 3rd EAI International Conference on Management of Manufacturing Systems*, p. 118, 2018, doi: https://doi.org/10.4108/eai.6-11-2018.2279672.

[9] A. M. Ezhilazhahi and P. T. v Bhuvaneswari, "IoT enabled plant soil moisture monitoring using wireless sensor networks," in *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS)*, 2017, pp. 345–349. doi: https://doi.org/10.1109/SSPS.2017.8071618.

[10] A. Mohsen, M. Al-Mahdawi, M. M. Fouda, M. Oogane, Y. Ando, and Z. M. Fadlullah, "AI Aided Noise Processing of Spintronic Based IoT Sensor for Magnetocardiography Application," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, Jun. 2020, pp. 1–6. doi: https://doi.org/10.1109/ICC40277.2020.9148617.

[11] R. da Rosa Righi, E. Correa, M. M. Gomes, and C. A. da Costa, "Enhancing performance of IoT applications with load prediction and cloud elasticity," *Future Generation Computer Systems*, vol. 109, pp. 689–701, 2020, doi: https://doi.org/10.1016/j.future.2018.06.026.

[12] K. Zhang, "A denoising algorithm for linear frequency modulation signal based on window function and empirical mode decomposition," in *Journal of Physics: Conference Series*, Jul. 2021, vol. 1971, no. 1. doi: https://doi.org/10.1088/1742-6596/1971/1/012082.

[13] H. Uyanik and M. Koseoglu, "Performance Evaluation of Different Window Functions for Audio Fingerprint Based Audio Search Algorithm," Oct. 2020. doi: https://doi.org/10.1109/ISMSIT50672.2020.9255225.

[14] A. Smiti, "A critical overview of outlier detection methods," *Computer Science Review*, vol. 38. Elsevier Ireland Ltd, p. 100306, Nov. 01, 2020. doi: 10.1016/j.cosrev.2020.100306.

[15] I. Ben-Gal, "OUTLIER DETECTION," 2005. doi: https://doi.org/10.1007/0-387-25465-X_7.

[16] F. Rehm, F. Klawonn, and R. Kruse, "A novel approach to noise clustering for outlier detection," *Soft Computing*, vol. 11, no. 5, pp. 489–494, Mar. 2007, doi: https://doi.org/10.1007/s00500-006-0112-4.

[17] M. Rakotosaona, V. la Barbera, P. Guerrero, N. J. Mitra, and M. Ovsjanikov, "<scp>PointCleanNet</scp>: Learning to Denoise and Remove Outliers from Dense Point Clouds," *Computer Graphics Forum*, vol. 39, no. 1, pp. 185–203, Feb. 2020, doi: https://doi.org/10.1111/cgf.13753.

[18] H. M. Saputra, "Perbandingan Average Filter dengan Hanning Filter pada Pengolahan Sinyal Load Cell," *Seminar Nasiona Itenas, Pusat Penelitian Tenaga Listrik dan Mekatronik, Lembaga Ilmu Pengetahuan Indonesia*, 2016, [Online]. Available: https://www.researchgate.net/publication/315455042

[19] S. Hansun, "A New Approach of Moving Average Method in Time Series Analysis," 2013. doi: https://doi.org/10.1109/CoNMedia.2013.6708545.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[20] D. K. Mak, "Simple Moving AverageSimple moving averages (SMA)," in *Trading Tactics in the Financial Market: Mathematical Methods to Improve Performance*, Cham: Springer International Publishing, 2021, pp. 29–55. doi: 10.1007/978-3-030-70622-7_3.

[21] Z. Cai, A. Ravichandran, S. Maji, C. Fowlkes, Z. Tu, and S. Soatto, "Exponential Moving Average Normalization for Self-supervised and Semi-supervised Learning," 2021. doi: https://doi.org/10.1109/CVPR46437.2021.00026.

[22] M. Nakano, A. Takahashi, and S. Takahashi, "Generalized exponential moving average (EMA) model with particle filtering and anomaly detection," *Expert Systems with Applications*, vol. 73, pp. 187–200, 2017, doi: https://doi.org/10.1016/j.eswa.2016.12.034.