

Supervised Learning from Data Mining on Process Data Loggers on Micro-Controllers

Adi Dwifana Saputra^{1)*}, Djarot Hindarto²⁾, Haryono³⁾

^{1,2,3)}Universitas Pradita, Serpong, Banten

¹⁾adi.dwifana@student.pradita.ac.id, ²⁾sosbuk@gmail.com, ³⁾haryono@pradita.ac.id

Submitted : Nov 23, 2022 | **Accepted** : Nov 24, 2022 | **Published** : Jan 1, 2023

Abstract: In processing data science, data is needed as input. Sometimes the data needed is not available in public data, this is where the purpose of this research was made. The acquisition process is very important to process information into data. After that, the data is processed to make a decision. Microcontrollers in controlling conditions, such as temperature, and humidity are very common devices, and many studies have been carried out. Sometimes discussing it just shows how to serialize and save it on online platforms, like firebase, tinger.io and many other online platforms. So that the process of storing data on external or online platforms is an advantage for platform providers, where platform providers don't need to do business and get data for free. This is unnoticed by researchers who have produced microcontroller devices. The many platforms for storing data range from hardware and software. Some tools are paid or open source. This research uses software that is open source. Because using open source-based tools will be easy to develop and for further research purposes. The development of further research by entering the code into the microcontroller system or what is called an embedded system. Data is a very valuable asset. Because data is one of the most important components in processing data science. And better take care of the data logger. This study uses a microcontroller and ultrasonic distance sensor and potentiometer. The method from the results of the logger (dataset) is used for classification using the support vector machine and decision tree algorithms. Accuracy with the support vector machine reaches 97% and the decision tree reaches 100% accuracy.

Keywords: Arduino; Data Logger; Data Science; Distance Sensor; Micro-Controller Device; Potentiometer;

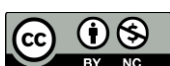
INTRODUCTION

A system controller or minimum system can generate data. For this reason, data from various microcontroller devices can be retrieved. This is very important considering that data is an asset. Many researchers do not take advantage of or even the data is exported or provided free of charge to the platform provider, this benefits the platform provider. The reasons from some parties say that there is no benefit with data from the device, but researchers who study data science will argue that data is an asset for data science. There are many service providers or frameworks for the Internet of Things, ranging from free to paid ones. Companies such as Google with its Firebase, Thingier.io, and many more data storage services are used by Internet of Things researchers as storage of data generated by the Internet of Things (IoT) devices. Features offered by storage providers or services such as displaying data with very attractive graphics. Storage service built with a cloud-based framework and real-time data access 7x24 hours.

Simply put, the resulting IoT devices use Arduino, Espressif (Ann, 2008), Raspberry or the cheapest and smallest devices that can send data to cloud-based storage services. Micro-controller devices that have wireless (WI-FI) features can already transmit data from sensors that have been connected to the micro-controller. As an example of using a data logger from the simple circuit in fig 1. In the experiment, the circuit will produce a data logger using 3rd party software. Lots of open source scripts to fetch data from arduino or expressive Esp8266. The result of data retrieval from the controller is a data mining process to serve as a dataset for data science purposes. This research indirectly discusses how to produce datasets

In addition to devices from micro-controllers (Yang et al., 2021) that have simple circuits, but if the data is generated from sensors such as ultrasonic sensors (Lawson et al., 2022), temperature and humidity sensors (Gancliev et al., 2019), water level detection sensors, CO, NO dangerous gas detection sensors and other sensors. Much of the research uses Arduino, esp8266 the resulting data is further processed. So that these data cannot be

*name of corresponding author



used as a decision maker. In this case it is a waste. Therefore, in this research a simple prototype is made, namely assembling the esp8266 micro-controller, adding an ultrasonic sensor and a potentiometer sensor. Then the series generates data, then the data is stored using 3rd party software to generate datasets. The process of producing a dataset is called the data mining process. The dataset is used by data science for processes such as machine learning where the goal is to be able to make supervised or unsupervised models depending on the needs.

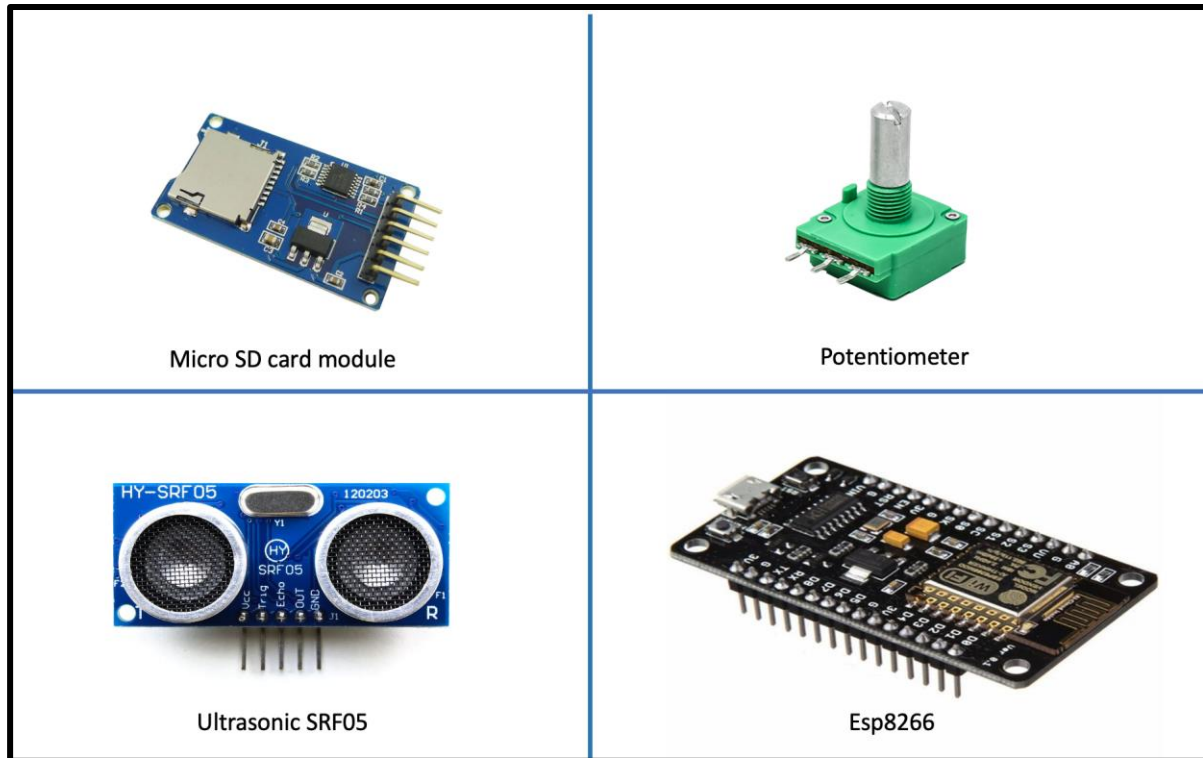


Fig. 1 components used in research
Source: Google Image

The explanation given raises several research questions related to this research.
Research Question 1, How is the micro-controller circuit developed?
Research Question 2, How to develop a simple data mining process?
Research Question 3, How do data mining processes become datasets and supervised machine learning processes produce supervised learning models?

LITERATURE REVIEW

Previous research discussing data loggers has been carried out a lot. In table 1, most research uses microcontrollers and data loggers that are not processed further and if they do the process, the data will provide insight to those who carry out the process.

Table 1. Research discussing data loggers

Reference	Topic	Strength	Weakness
(PUDIN & MARDIYANTO, 2020)	Design and Implementation of Data Logger for Measurement of Solar Panel Output Power and Solar Irradiation	This research has used a microcontroller to produce data mining using a series of data loggers.	The resulting dataset uses an SD card data logger, data science is not processed using data science such as supervised or unsupervised machine learning.
(Andi Setiono, 2010)	Manufacturing and Testing of ATMEGA32 Microcontroller-based Data Loggers for Monitoring of Land Shifting	The prototype used in measuring ground displacement uses a microcontroller. The	However, the results of the data logger are not used or further processes such as

*name of corresponding author



		measurement results are stored in an excel file.	machine learning are carried out, so they will gain insight into the data.
(Mahendra Sanjaya et al., 2018)	Visual Based Data Logger System Design On Solar Cells	This research stores information through a data logger in the form of an SD card and then saves it to a file. Then the data is entered into the database and visualized in graphs.	However, this research does not process with machine learning to gain insights into the data.
(Yandi, 2020)	Prototipe Data Logging Monitoring System Untuk Konversi Energi Panel Surya Polycrystalline 100 Wp Berbasis Arduino Uno	This research performs a data logger from a 100 Wp solar panel using Arduino Uno.	However, this research does not process with machine learning to gain insights into the data.
(Kholis et al., 2022)	Test the horizontal axis wind turbine blade length variation using a microcontroller	This research uses turbines and microcontrollers. The resulting data will be tested.	The data can be tested and it is better if the data is analyzed or processed with machine learning algorithms to gain insight into the data, so that the results are good enough for data analysis.
(Mahzan et al., 2017)	Design and Development of an Arduino Based Data Logger for Photovoltaic Monitoring System	Using a data logger to retrieve data from Photovolatic monitoring with the help of Datalogger. Data will be taken from the sensor then translated to the next micorcontroler, the data is stored.	The weakness of this research is not analyzing the data using machine learning algorithms, so that information can be obtained and can be processed with Machine Learning
(Mabrouki et al., 2021)	IoT-Based Data Logger for Weather Monitoring Using Arduino-Based Wireless Sensor Networks with Remote Graphical Application and Alerts	This research uses a variety of sensors such as the MQ-136, DHT11, MQ-7, and MQ-131 sensors. The results of the sensor readings will be sent to the host via the wireless network.	However, the data taken from these sensors is only displayed in the application. No further analysis was carried out using Machine Learning.

In table 1, it has explained the research that has discussed data loggers. And many data logger processes are actually simple mining processes, but no further processing is used. The gap in this research is that previous research on data loggers was not used for further processing. The state-of-the-art in this research processes data loggers using machine learning as data processing. Therefore this research does not look for weaknesses from previous research that discusses data loggers. This research complements the shortcomings of previous research using data loggers for data science processes.

METHOD

This research proposes a simple data mining method using a micro-controler from espressif, ultrasonic sensors, and potentiometer sensors. The data obtained from the ultrasonic sensor and potentiometer sensor is read by 3rd party software, resulting in a dataset in the form of excel or csv. The reason for using 3rd party software is because the software is open source, so that the software can be developed according to needs.

*name of corresponding author



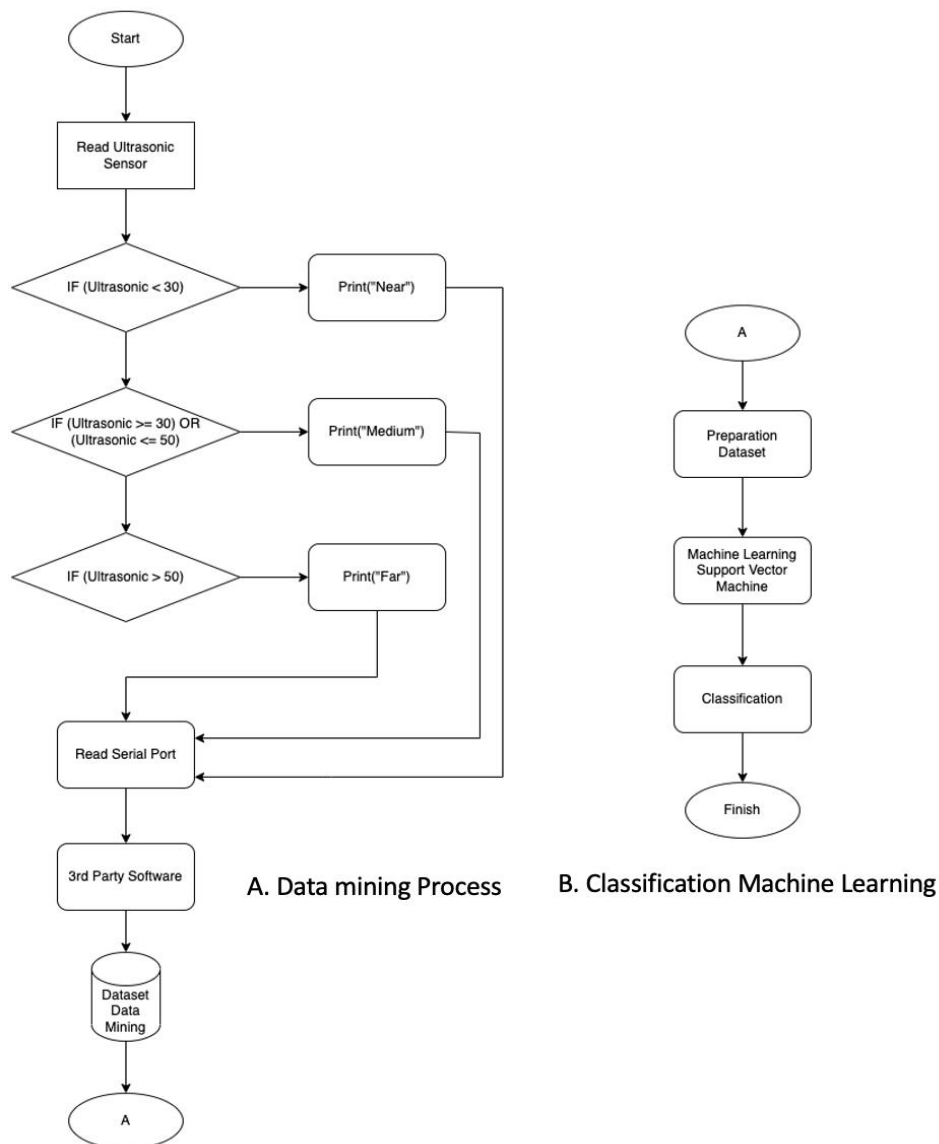
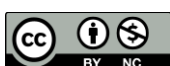


Fig. 2 A. Data Mining Process, and B. Classification Machine Learning
Source: Researcher Property

Explanation for fig 2, there is a Data Mining process, in simple terms the mining process can be used in various ways to get it. So much data and information is wasted. As shown above is the process for getting data from the Arduino micro-controller. What is mined is the ultrasonic sensor and potentiometer into datasets that are ready to be used for Machine Learning needs. The distance is less than 30 centimeters, the target or label is "Near". If the distance is greater than 30 or the distance is less than 50, then the label shows "Medium". The distance is greater than 50 then the label shows "Far". All processes are carried out and read using 3rd party software. Process results from 3rd party software produce datasets that can be in the form of excel files or csv formats.

After the data mining process is complete, it is followed by a classification process using machine learning. The method chosen is supervised classification because the mining dataset contains labels in the dataset. The algorithm used in this research is the Support Vector Machine. Actually, you can also use supervised learning algorithms such as decision trees, logistic regression. The merging of Arduino micro-controllers and machine learning has the following reasons. The output results from the micro-controller design produce private datasets. Meanwhile, for data science, sometimes a lot of research uses public datasets. As a wise data scientist, of course, you can use a mixture of datasets from public and datasets from private, so that you have a variety of datasets. For data science it is important with the dataset that will be researched, while researchers who focus on hardware sometimes do not pay attention to whether the output is processed or not, because the focus is only on creating prototypes and not carrying out the next process.

*name of corresponding author



Expressif, is an open source hardware platform. This platform is widely used, especially in the control field. Because it is good for control. There are many variations of micro-control products, starting from Esp8266, Esp32, Esp8685, Esp8285. This company engaged in the field of micro-controller devices has provided various hardware, software and documentation platforms. All of these devices can be connected to WIFI, Cloud, so that these devices are services called Internet of Things devices. A micro-controller (Sze et al., 2022) as long as it doesn't provide a device that is connected to the internet or cloud, is not yet called the Internet of Things. As a micro-controller manufacturer, Expressif produces products with complete features. Table 2, displays the features of the Expressif product.

Table 2. Expressive product features
Source: (Ann, 2008)

TCP/IP Stack	Bluetooth LE Stack	Wi-Fi MAC Library	Bluetooth Controller
Wi-Fi & Bluetooth	Developer Tools	Power Management	RTOS Kernel
Software Bootloader	SoC Support	Object Storage	POSIX and C++ Support
File Systems	Network Security	IDE Plugins	Crypto Library
Peripherals Drivers	Build System	Wi-Fi MAC Library	Manufacturing Utilities

The abundance of features on the Expressif micro-controller product makes this device widely used by developers in the field of Internet of Things research. So that a lot of Internet of Things devices are produced. The large use of micro-controllers causes a lot of data to be generated, for example devices from CCTV, devices for measuring temperature, humidity, and other products. There are so many billions of data and if it is not utilized it will not make decisions properly.

Sensor Ultrasonic, is a sensor that performs transmit and receive processes in detecting objects. This object detection process is inspired by bats. reflects sound waves and interprets the existence of an object with a certain sound frequency. The mention of ultrasonic sensors because this sensor uses ultrasonic waves. Electronic devices can convert DC electrical energy into mechanical energy in the form of sound waves. Ultrasonic waves can propagate through solids, liquids and gases. Ultrasonic waves are mechanical waves with longitudinal frequencies and above 20 Khz. Ultrasonic waves are sound waves that have a very high frequency of 20,000 Hz. Ultrasonic sound cannot be heard by the human ear. Ultrasonic sensors transmit ultrasonic waves (above the threshold for human hearing) and provide an output pulse that corresponds to the time it takes for the reflected waves to be received back by the sensor. By measuring the time lag between the sent pulse and the received pulse, the distance measured can be calculated.

Machine Learning is part of Artificial Intelligence and computer science that focuses on data. The method used in the learning process imitates the way humans learn by increasing the performance and accuracy of their learning. Machine learning is divided into several parts in how to do the learning. Machine learning has various methods. Machine learning has been widely used by various industries, which work using big data and recognize that machine learning algorithms help many problems in the industrial world. The use of Machine Learning as an insight into data insights in the company industry, and being able to solve problems that exist in companies or industries. Supervised machine learning, unsupervised machine learning and reinforcement learning. Supervised learning machines such as Support vector machines (Hindarto, 2022), Decision Trees, Logistic Regression are discussed in this research.

Support Vector Machine, this algorithm is quite famous for solving classification (Hindarto & Handri Santoso, 2021) or regression. But in some problems this algorithm is suitable for use as a classification. This algorithm solves the problem by dividing a point in a space by a boundary line or hyperplane. This hyperplane is used to separate points in space. Because the hyperplane is used as a dividing line in the room, it can be used as a support vector. Therefore, the process confirms the division of areas or classes, so this algorithm is called a Support Vector Machine. In some cases this algorithm produces high accuracy for labeled data. Also known as the supervised learning method. If there are no labels in the dataset, then the learning method uses unsupervised learning.

Machine Learning Decision Tree (Sarailidis et al., 2023), is a well-known supervised learning algorithm in dealing with classification problems. This algorithm processes similar to a decision tree to generate a decision. Doing the process by dividing the decision node, leaf node, root node and sub tree. This algorithm can solve problems for classification and regression at the same time.

*name of corresponding author



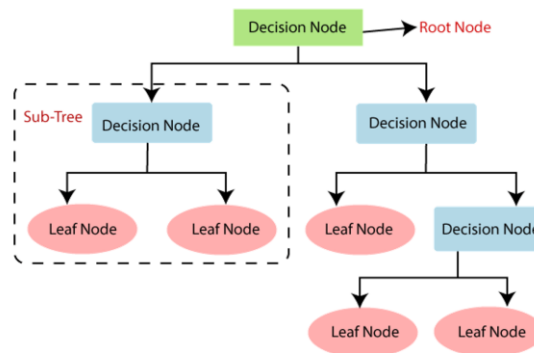


Fig. 3 Decision tree illustration

Source: <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>

Fig. 4 describes the decision tree in carrying out the classification process. Starting from the decision node or root node to form a branch or sub tree. All the features in the dataset are resolved to the leaf nodes. The result is that all features are completed by forming the next branch until the end point is a leaf node as the output or classification result.

Machine Learning Logistic Regression (Song et al., 2021) is a well-known Machine Learning algorithm, and is Supervised Learning. The benefit of Logistic Regression (van Eeden et al., 2021) is that it can be used to predict parameters or features depending on the class used. The prediction of the output depends on the dataset it processes. The result is a categorical value and can be used as a probability (prediction). Logistic Regression is very similar to Linear Regression, the difference is that Linear Regression is used to solve Regression problems, while Logistic regression is used to solve classification problems.

RESULT

The experimental results using the ESP8266 micro-controller produce a dataset that can be used for the needs of Machine Learning. Fig. 4 is a device for generating datasets.

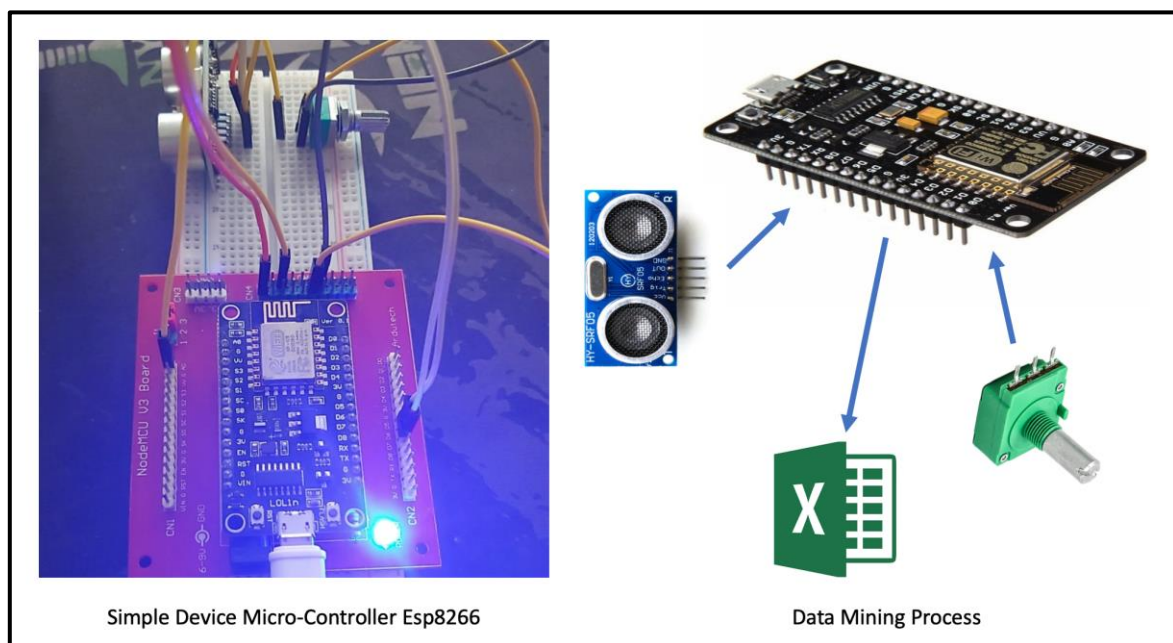
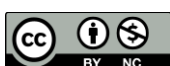


Fig. 4 Simple Process Mining with Micro-controller

Source: Researcher Property

Fig. 4 after the ESP 8266 circuit and the sensor are connected to the computer's USB Serial, the circuit has sent data to the computer. The Excel program is run and settings are made by entering options and selecting communication, the Data Streamer menu will enter the menu. Then click Connected to Device, it will be connected to the Serial Communication Micro-Controller. If it is connected, the data mining process is ready to run. As seen in Fig. 5.

*name of corresponding author



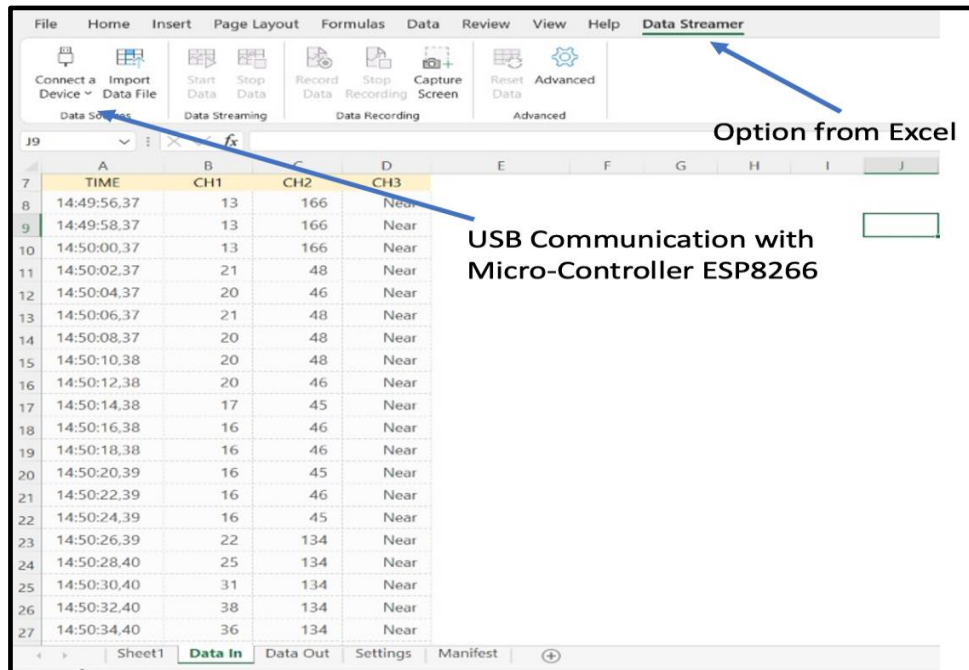


Fig. 5 Dataset for Machine Learning
Source: Researcher Property

Fig. 5 Dataset for Machine Learning, after running to retrieve the dataset, then the data is stored, as needed. The dataset is ready for processing with the Supervised Learning algorithm.

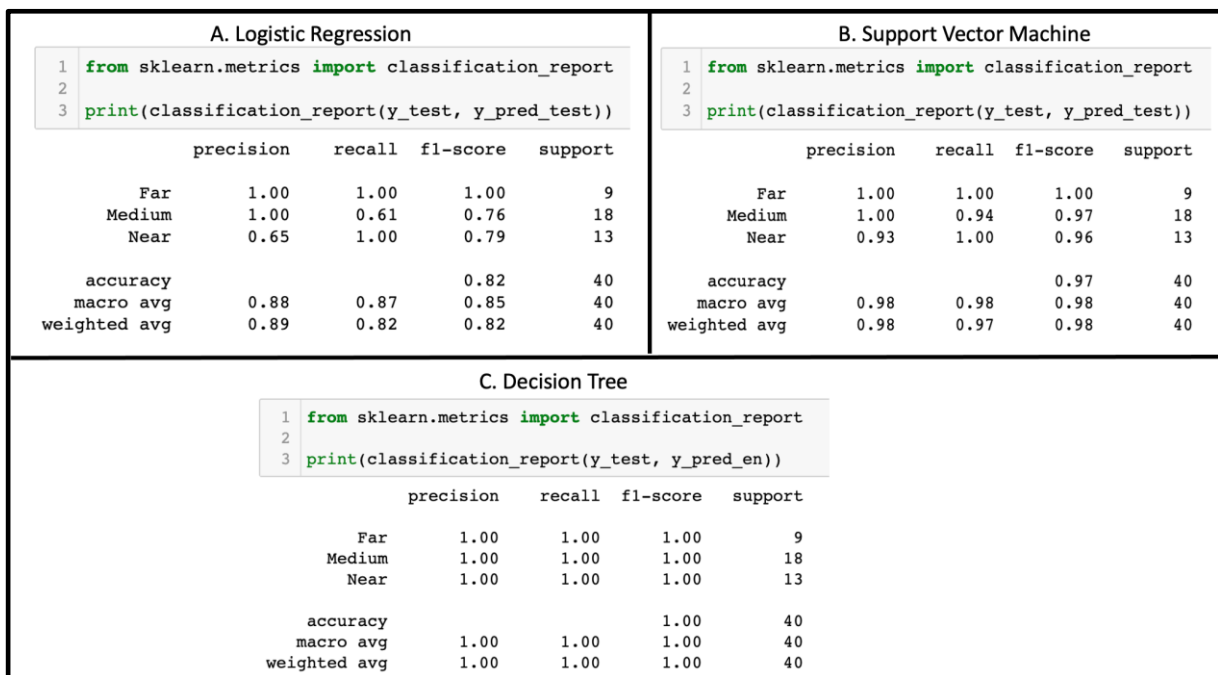


Fig. 6 Performance from Logistic Regression, Support vector Machine and Decision Tree
Source: Researcher Property

Fig. 6, regarding the performance of the three supervised learning algorithms shows different results. This dataset is a private dataset. Because of how to obtain or how to mine using your own device and storage. Fig. 6 part A. Logistic Regression, the dataset is done with data preparation, by checking if there are NaN, Empty and others. After that, the fitting is done for training. By dividing the 80% training dataset and 20% testing dataset, after training the accuracy results reach 82%. The performance of the Support Vector Machine algorithm reaches 97% accuracy. The Decision Tree algorithm achieves 100% accuracy. The highest accuracy in this research is the Decision Tree.

*name of corresponding author



DISCUSSIONS

Research Question 1, How is the micro-controller circuit developed? In the results section, we have discussed the micro-controller. The design of the ultrasonic sensor as a detection of object distance with a potentiometer as a resistance variable. Then the ultrasonic sensor for Trigger is inserted into pin D5 and Echo is inserted into pin D6. After completing the assembly the circuit is tested.

Research Question 2, How to develop a simple data mining process? The dataset obtained from the micro-controller is the easiest way in a simple mining process. There are a lot of 3rd party software in getting data and compiling it into datasets. Microsoft excel with data stream is software that contains excel macros that can connect the excel application with a USB port. This communication is used as a liaison between the micro-controller and the excel program.

Research Question 3, How do data mining processes become datasets and supervised machine learning processes produce supervised learning models? The data preparation process has been explained in the results section. After the dataset preparation is carried out, it is continued by training the dataset to become a classification model. The dataset already has a label or target. Supervised learning is used because the dataset already has a label or target.

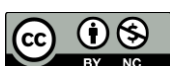
CONCLUSION

The conclusion drawn from this study is that a lot of previous research actually produced data, but data processing was not carried out using data science processes. So that a lot of the data that appears is not used to gain broader insights about the data. The presence of the Machine Learning method can solve problems from simple datasets. Such as the use of Machine Learning algorithms such as Support Vector Machine, Decision Tree, Logistic Regress, it turns out that it is capable of creating Supervised Learning models. From the experiments studied on the dataset of the results of the mining process from the microcontroller, the results were good. The accuracy of the logistic regression reaches 82%, the accuracy of the support vector machine reaches 97%. Decision tree accuracy reaches 100%.

REFERENCES

- Andi Setiono, P. P. dan B. W. (2010). PEMBUATAN DAN UJI COBA DATA LOGGER BERBASIS MIKROKONTROLER ATMEGA32 UNTUK MONITORING PERGESERAN TANAH. *Jurnal Fisika*, 10(242), 83–94.
- Ann, T. S. (2008). *Espressif Micro Controller*. <https://www.espressif.com/en>
- Gancliev, I., Taueva, A., Kutryanski, K., & Petrov, M. (2019). Decoupling Fuzzy-Neural Temperature and Humidity Control in HVAC Systems. *IFAC-PapersOnLine*, 52(25), 299–304. <https://doi.org/10.1016/j.ifacol.2019.12.539>
- Hindarto, D. (2022). Perbandingan Kinerja Akurasi Klasifikasi K-NN, NB dan DT pada APK Android. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(1), 486–503. <https://doi.org/10.35957/jatisi.v9i1.1542>
- Hindarto, D., & Handri Santoso. (2021). Android APK Identification using Non Neural Network and Neural Network Classifier. *Journal of Computer Science and Informatics Engineering (J-Cosine)*, 5(2), 149–157. <https://doi.org/10.29303/jcosine.v5i2.420>
- Kholis, N., Yuliza, E., & Ekawita, R. (2022). UJI VARIASI PANJANG BLADE TURBIN ANGIN SUMBU HORIZONTAL. 7(2), 1–6.
- Lawson, A. R., Giri, K., Thomson, A. L., Karunaratne, S. B., Smith, K. F., Jacobs, J. L., & Morse-McNabb, E. M. (2022). Multi-site calibration and validation of a wide-angle ultrasonic sensor and precise GPS to estimate pasture mass at the paddock scale. *Computers and Electronics in Agriculture*, 195(January), 106786. <https://doi.org/10.1016/j.compag.2022.106786>
- Mabrouki, J., Azrou, M., Dhiba, D., Farhaoui, Y., & Hajjaji, S. El. (2021). IoT-based data logger for weather monitoring using arduino-based wireless sensor networks with remote graphical application and alerts. *Big Data Mining and Analytics*, 4(1), 25–32. <https://doi.org/10.26599/BDMA.2020.9020018>
- Mahendra Sanjaya, I. P. G., Indra Partha, C. G., & Khrisne, D. C. (2018). Rancang Bangun Sistem Data Logger Berbasis Visual Pada Solar Cell. *Majalah Ilmiah Teknologi Elektro*, 16(3), 114. <https://doi.org/10.24843/mite.2017.v16i03p18>
- Mahzan, N. N., Omar, A. M., Rimon, L., Noor, S. Z. M., & Rosselan, M. Z. (2017). Design and development of an arduino based data logger for photovoltaic monitoring system. *International Journal of Simulation: Systems, Science and Technology*, 17(41), 15.1-15.5. <https://doi.org/10.5013/IJSSST.a.17.41.15>
- PUDIN, A., & MARDIYANTO, I. R. (2020). Desain dan Implementasi Data Logger untuk Pengukuran Daya Keluaran Panel Surya dan Iradiasi Matahari. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 8(2), 240. <https://doi.org/10.26760/elkomika.v8i2.240>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Sarailidis, G., Wagener, T., & Pianosi, F. (2023). Integrating scientific knowledge into machine learning using interactive decision trees. *Computers and Geosciences*, 170(October 2022), 105248. <https://doi.org/10.1016/j.cageo.2022.105248>
- Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. *International Journal of Medical Informatics*, 151, 104484. <https://doi.org/10.1016/j.ijmedinf.2021.104484>
- Sze, E., Hindarto, D., & Wirayasa, I. K. A. (2022). *Performance Comparison of Ultrasonic Sensor Accuracy in Measuring Distance*. 7(4), 2556–2562.
- van Eeden, W. A., Luo, C., van Hemert, A. M., Carlier, I. V. E., Penninx, B. W., Wardenaar, K. J., Hoos, H., & Giltay, E. J. (2021). Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve Bayes classifier, and traditional logistic regression. *Psychiatry Research*, 299(October 2020), 113823. <https://doi.org/10.1016/j.psychres.2021.113823>
- Yandi, W. (2020). Prototype Data Logging Monitoring System Untuk Konversi Energi Panel Surya Polycrystalline 100 Wp Berbasis Arduino Uno. *Jurnal Ecotipe (Electronic, Control, Telecommunication, Information, and Power Engineering)*, 7(1), 55–60. <https://doi.org/10.33019/ecotipe.v7i1.1486>
- Yang, R., Liu, Y., Yu, Y., He, X., & Li, H. (2021). Hybrid improved particle swarm optimization-cuckoo search optimized fuzzy PID controller for micro gas turbine. *Energy Reports*, 7, 5446–5454. <https://doi.org/10.1016/j.egy.2021.08.120>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.