Chapter 3
RESEARCH METHODOLOGY

3.1   Internet of Things

IoT is emerging as a big issue on the Internet. It is expected that billions of physical things or objects will be equipped with various types of sensors connected to the internet through networks as well as technology support such as embedded sensors and actualization, radio frequency identification (RFID), wireless sensor networks, real-time and web services, IoT cyber-physical systems or network from the network. With a large number of things/objects and sensors/actuators connected to the internet, massive and in some cases real-time data streams will be automatically generated by the connected things and sensors[1].

3.1.1 Internet of Things Research Methods

A systematic literature review is a research method to identify, evaluate and interpret all relevant research results related to certain research questions, certain topics, or phenomena of concern. Meanwhile, meta-analysis is one way to synthesize results statistically (quantitative technique). Another way to synthesize results is the narrative technique (qualitative technique)[2].
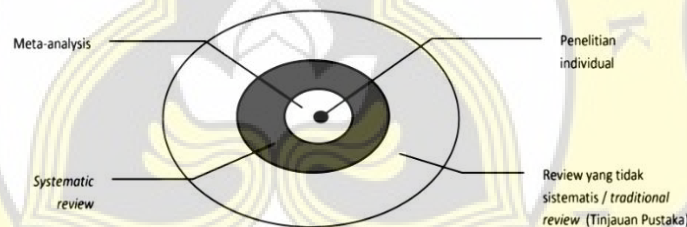

Figure 1: The position of the systematic literature review methodology[2]

3.1.2 How the Internet of Things Works

The way IoT works is that every object must have an Internet Protocol (IP) address. An Internet Protocol (IP) address is an identity on a network that allows it to be ordered from other objects on the same network. Furthermore, the Internet Protocol (IP) addresses in these objects will be connected to the internet network. Sensors and built-in IP and security functions are usually used to realize IoT products, such as Arduino Yun, Raspberry PI, BeagleBone Black, and so on.[2].

Figure 2: Examples of IoT in life and its elements[2]

3.2 Software Arduino

Arduino IDE is software provided on the arduino.cc website intended as a sketch development tool is use as a program on the Arduino board. IDE (Integrated Development Environment) means a form of integrated program development tool so that various needs are provide and expressed in the form of a menu-based interface[3].
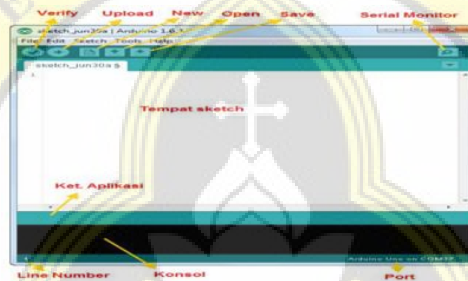

Figure 3: Arduino Sketch[3].

3.3 Arduino UNO

Arduino is an open-source microcontroller minimum system board. Inside the Arduino board is a 328 series AVR microcontroller is product of Atmel. Arduino has its advantages compared to other microcontroller boards being open source. Arduino also has its program in the form of C language. Besides that, the Arduino board has loader in the form of USB making it easier when programmed microcontroller in Arduino[4] .

Figure 4: Arduino Uno Driver[4]

3.4 **Temperature Sensor**

In measuring air temperature, a measuring instrument must be needed, namely an air temperature thermometer. In the world of electronics that focuses on digital electronics, the measurement of a parameter is always done using a sensor. These sensors are sensors that are well known in the electronics field, namely the LM35, DHT11, DHT22, and DS18B20 sensors[5].

3.4.1 DHT11

DHT11 is a sensor that can measure two parameters at once, namely temperature and humidity. This sensor has a digital signal output calibrated with temperature and humidity sensors. Apart from being of excellent quality, this sensor has a fast response, anti-interference capabilities, and the cost advantage of being able to measure two parameters at once[5].
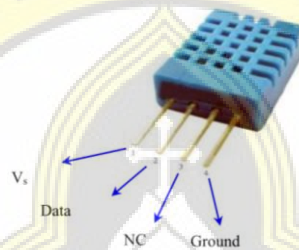

Figure 5: DHT11 Foot Configuration[5]

3.4 **XAMPP**

XAMPP is free software, which supports many operating systems, and is a compilation of several programs. Its function is as a stand-alone server (localhost), which consists of the Apache HTTP Server program, MySQL database, and a language translator written in the PHP programming language[6].

3.5 **PHP**

PHP is a programming language that is widely used for handling the creation and development of a web and can be used in HTML. PHP stands for "PHP: Hypertext Preprocessor". PHP is designed to work with database servers[6].

3.6 **MySQL**

The existence of an API (Application Programming Interface) facility owned by Mysql, allows various computer applications written in various programming languages to access the MySQL database[6] .

3.7 **Machine Learning**

Machine learning is part of artificial intelligence that is widely used to solve various problems. This article reviews problem-solving from the latest studies by

classifying machine learning into three categories: supervised learning, unsupervised learning, and reinforcement learning[7].

3.7.1 Supervised Learning

Supervised Learning is a classification method in which data sets are completely labeled to classify unknown classes. The supervised learning method is based on a collection of data samples that have labels. Supervised learning has several popular algorithms such as Linear regression, Random Forest, Support Vector Machines, Naive Bayesian, Decision Tree, and k-Nearest Neighbor. Then several algorithms for classification are also mentioned in such as Support Vector Machines (SVM), Normal Bayesian Classifier (NBC), K-Nearest Neighbor (KNN)[7].

3.7.2 Unsupervised Learning

In unsupervised learning, the system is provided with some sample input but no output is present. Unsupervised learning has several popular algorithms such as K-means, and Apriori[7].

3.7.3 Reinforcement Learning

Reinforcement learning problems are solved by learning new experiences through trial and error. The reinforcement learning algorithm is related to dynamic programming algorithms that are often used to solve optimization problems[7].

3.8 Naive Bayes Method

The advantage of using Naive Bayes is that this method only requires a small amount of training data (Training Data) to determine the parameter estimates needed in the classification process[8].

3.8.1 Naive Bayes Formula

The equation of Bayes' theorem is:

$$P\left(\frac{H}{X}\right) = \frac{P\left(\frac{X}{H}\right).P(H)}{P(X)} \text{ [8]}$$

Where :

X	: Data with unknown class
H	: Hypothesis data is a specific class
P(H|X)	: Probability of hypothesis H based on condition X (posteriori probability)
P(H)	: Hypothesis probability H (prior probability)
P(X|H)	: Probability of X based on the conditions on the hypothesis H
P(X)	: Probability X

To explain the Naive Bayes method, it should be noted that the classification process requires several instructions to determine what class is suitable for the analyzed sample[8].

3.8   Confusion Matrix

A table that describes the performance of a specific model or algorithm. Each row of the matrix represents the actual class of the data, and each column represents the predicted class of the data (or vice versa)[9].

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

Table 1:   Confusion Matrix[9]

3.8.1 True Positive
How much of the actual data class is positive, and the model also predicts positive[9].

3.8.2 True Negative
How much of the actual data class is negative, and the model predicts negative[9].

3.8.3 False Positive
How much of the actual data class is negative, but the model predicts positive[9].

3.8.4 False Negative
How much of the actual data class is positive, but the model predicts it is negative [9].

Through these 4 data, other data can be obtained that are very useful for measuring the performance of a model, including:

3.8.5 Accuracy
The overall total of how often the model correctly classifies. The accuracy formula can be written using the equation:

$$\frac{TP+TN}{Total} \qquad [9]$$

3.8.6 Precision

When the model predicts positively, how often is the prediction correct. The precision formula can be written using the equation:

$$\frac{TP}{FP+TP} \qquad [9]$$

3.8.7 Recall (Sensitivity / True Positive Rate)
When the actual class is positive, how often does the model predict positive. The recall formula can be written using the equation:

$$\frac{TP}{FN+TP} \qquad [9]$$

3.8.8 F1-Score
Is the average of the harmonics of Precision and Recall. The f1-score formula can be written using the equation:

$$2 * \frac{precision*recall}{precision+recall} \qquad [9]$$