CHAPTER 1
INTRODUCTION

1.1 Background

Indonesia is an agricultural country, one of which supports the horticulture industry, which is an important aspect in the national economy. The number of individuals who live and work in the agriculture industry demonstrates this effect. West Java, for example, is a cool-climate region with several mountains appropriate for plantations, such as spinach, lettuce, mustard greens, and kale, and West Java is the major kale grower. Based on data from the Indonesian Ministry of Agriculture, specifically data for 2019, 56,229 tons, 2018, 58,228 tons, and 2017, 51,383 tons, there were relatively big changes in harvest seen in the year of observation, ranging from 2000 tons to 6000 tons [1].

Kale is a popular vegetable in Indonesia, particularly in West Java, which has a tropical climate. This plant is coming from the tropic regions, mainly Africa and Asia [2]. Kale is a vulnerable vegetables and unstable. Due to this unpredictability of kale supply, the fluctuating yield can lead to volatile market pricing. Weather sensitivity, limited resources such as producers, planting expertise, infrastructure, land availability, the impact of fertilizers, and climate change can all contribute to fluctuating yields. In order to keep kale commodity prices stable, the finest grade kale must be supplied in response to an area's sufficient demand. Soil moisture, room temperature, room humidity, and light intensity were specific elements noted for model analysis as growth variables for growing Kale. Therefore, the agricultural system of smart farming for growing Kale must be revitalized. Previously, a global system schema using Internet of Things (IoT) for Kale's farming was proposed. Then, that research technique upgraded with the current research technique, which employs a Machine Learning System to support Kale cultivation through smart farming

The aim of this IoT technique is to increase Kale production by providing farmers with the best Kale planting period model to raise the probability of optimal yields while also opening new chances to preserve Kale availability and quality in the market. A greenhouse is one of the facilities used to perform appropriate

planting. The greenhouse is designed to produce water Kale under ideal circumstances by paying close attention to the environment and reducing undesirable environmental elements.

The Internet of Things is a form of computing system in which little electrical devices equipped with sensors can detect their surroundings. Smart agriculture or smart farm is feasible to apply. The use of the Internet of Things to farming can save labor and money, improve temperature and humidity control, and collect more precise information about plant development variables [3]. MySQL, one of the most widely used Relational Database Management System solutions, is well known for its benefits. MySQL is also known for executing queries faster than other RDBMS systems [4]. In order to maximize the utilization of data in a database, Machine Learning can be used to investigate data usefulness for commercial purposes. Machine learning is the study and development of an algorithm that can learn and predict data using a computer system employing supervised learning or unsupervised learning techniques. This concept is acquiring the data model in various forms, depending on the machine's requirement and learning objective [28]. This technique will typically execute a model validation process on each learning process. Instead of following procedural instructions from the computer program, the learning algorithm predicts or determines the future by applying the model obtained from the sample dataset. This technique enables the machine to learn and increase its efficiency by learning from a sample dataset or the surroundings.

In this Thesis, a design architecture is formulated to create an ideal growth prediction model for kale, especially in the seedling phase. This research is based on three final projects published internally, the smart farm concept followed by an Internet of Things-based automation system [5]. MySQL database for raw data storage [6]. Machine learning techniques to develop models [7]. Finally, an integrated system is adopted, and combining concepts have been defined to create a seamless data set.

1.2 Problem Identification

The uncertainty of kale production for three years may lead to insufficient Kale supply resulting in higher demand in the future. Knowledge and technology illiteracy is seen as one of the other problems that may be felt by farmers. It is

difficult to decide which Kale seeds are preferred with quantity that matches quality. The impact of these problems require solutions to fulfill market demands, which are strengthened by ideas from farmers who claim that kale still has the potential to sell profitably. Kale is utilized in home and culinary menus. Currently, there are no references about growing kale that are good, comprehensive, systematic, and linked to databases. Then, this technique can be improved and become a profitable basis for farmers to determine the quality based on the quantity of kale plants to create products that do not significantly cut market prices.

1.3 Objective and Contribution

For the objective and the contributions, are:

1. Capable of implementing the global system scheme from IoT to the machine learning stage for generating the kale growth dataset and prediction model during the seeding phase.
2. Capable of achieving the highest Quality of Service and machine learning performance metrics value for the optimum dataset and prediction model output.
3. Be able to create an optimal kale growth model for the farmer environment as a future input for the finest kale crop.

1.4 Scope of the Work

For the scope of works, are:

1. This method is designed to create datasets based on the plants being studied.
2. The Raspberry Pi 3B+ model was utilized as the IoT device.
3. Python version 3.7 is the programming language used to design systems and prediction models.
4. The greenhouse temperature, humidity, light intensity, and soil moisture are monitored using the DHT22 sensor, YL-69 sensor, and GY-302-BH1750 sensor.
5. The database utilized is MySQL, which is an RDBMS.
6. The characteristic kale growth is not employed in predictive modeling.
7. Data collection and storage were performed on a Raspberry Pi, whereas kale prediction models were performed on a laptop.
8. The decision tree method is used to supervise learning in machine learning.

9. Pyton Library is used NumPy for data mining, pandas for data preparation, matplotlib and seaborn for data visualization, and scikit learn for machine learning in Python.
10. The kale growth dataset includes features collected during June and July in the Buah Batu area.
11. Data collection quality is determined by QoS (Quality of Service) parameters such as delay and throughput.
12. A confusion matrix and classification reports are used as machine learning performance metrics.
13. Network security to maintain data transmission between systems is not discuss.

1.5 Research Method

This thesis is divided into six work packages are as follows:

1. WP1: Literature Review

Collect the research and knowledge of formulating the dataset and prediction model using IoT, RDBMS, and machine learning system method from reputable sources such as articles, books, journals, or conference papers issued by credible publications, official organizations, or governments.

2. WP2: Model System

Organize the systematic design for the prediction model dataset from the data collecting phase in IoT, data storage in MySQL database, and data modeling by machine learning process in Python programming language are based on the literature review.

3. WP3: Application, Device, Tool, and Library Improvement

Ensure that the sensors are gathering data from the research object. Ascertain that the Raspberry Pi can store data and transmits it to the MySQL database for complete data storage. To build the prediction model, satisfy the Jupyter Notebook and its related library, which is installed on the laptop.

4. WP4: Dataset Preparation

Compiling the dataset in Microsoft Excel using manual parameter measurement characteristics, their associated values, and labels from the downloaded raw data in the MySQL database.

5. WP5: Analysis of the Kale Growth Model

In Jupyter Notebook, the prediction model is formulated by getting the specified dataset and establishing the model using the machine learning procedure.

6. WP6: Performance Metric Analysis

Analyze the Quality of Service in terms of data transmission throughput and delay. In addition, the classification report and confusion matrix are sorted by case, accuracy, and precision value to choose the optimum model.

1.6 Bachelor Thesis Organization

The rest of this thesis is organized as follows:

- Chapter 2 BASIC CONCEPT

  This chapter contains the explanation of the basic theory of the related system concept.

- Chapter 3 PROPOSED KALE GROWTH MODEL AND SYSTEM

  This chapter depicts the illustration of the algorithm, concept implementation, and flowchart diagram of the designed integrated system.

- Chapter 4 PERFORMANCE EVALUATION

  This chapter contains the results of quality of service and analysis from testing the smart farm automation system that has been created, testing the created web database, and the results of data modeling.

- Chapter 5 CONCLUSIONS

  This chapter contains the thesis' conclusion as well as the suggested idea for future research in the same field of interest.