CHAPTER I
INTRODUCTION

1.1 Background

Indonesia is a developing country with a broad population segment that is depending on their life in agriculture. The horticulture sector produces vegetable products, fruit plants, ornamental plants, and medicinal plants that can be traded. Based on data obtained from the Central Statistics Agency in 2018, the number of farmers is around 33.1 million, and there are 10.1 million horticultural farmers [1]. As a result of this population, agriculture is still one of the Indonesian economy pillars by contributing 14% of the GDP [2]. Besides helping the national income, the horticultural sector also contributing to its export activities and food needs. Again, the farmers receive the annual revenues sufficient for their lives. The other parties is involving as the supporting system in food supply distribution apply before consumption by the communities.

One of the leading agricultural commodities in Indonesia is bok choy. This plant is required as a supplementary food material in the household and industrial environment. Bok choy is considered as one of the plants classify in the mustard greens group. Along with high demand resulting in potential supply and price fluctuations, the mustard greens production rate is only increased by 2.63 % in 2019 over 2018 in the Central Statistics Agency of Indonesia data [3]. The public information also reinforces the Ministry of Agriculture data states that there has been a fluctuation rate in the population of West Java province food consumption by day and year from 2013 - 2018 [4]. The fluctuation is caused by several factors, including weather sensitivity, limited resources (producers, planting knowledge, infrastructure, and land availability), the influence of fertilizer, and climate change. In maintaining the stability of bok choy commodity prices, an adequate supply of the finest quality of bok choy is needed. The specific observed parameter to analyze is the soil moisture as the growth-related factor to plant the bok choy.

The objective is increasing the production of bok choy, which offers to farmers environment as the main player in producing the best choice of bok choy seeding phase in increasing the possibility of optimal crops and open opportunities to maintain the availability of bok choy quality on the market. One of the facilities used to do suitable planting is the greenhouse. The greenhouse is an architecture to plant

the bok choy following the desired optimal condition with a more attentive environment and minimize unwanted environmental factors. In addition to utilizing the environment independently, the researchers have conducted various researches on the use of greenhouses, including real-time monitoring and controlling greenhouse systems based on smart farms system [5]. An IoT device is connected through a network as an integrated system in retrieving real-time data is necessary. Raspberry Pi is used as an IoT device because it has an embedded wi-fi module and external instrument, such as sensors as a supporting system. The soil moisture sensor is defined as calibrating the water dose value of greater than or equal to 15 % as the lower limit and lower than equal to 25 % defined as normal condition [6]. Whereas the optimum temperature is in a range of 20 - 25 degrees Celcius [7]. The application of Raspberry Pi in the greenhouse as smart farm system implementation has been widely used by researcher [8][9].

Users can perform data retrieval results for a prediction model to determine the best composition in producing optimal bok choy seeding crops. This method has been proven by several researchers who made a collection of observation data results in tomatoes' and lettuce growth as it used for quality predictive models [10] [11]. The Raspberry Pi is retrieved data from the respective types of desired sensor monitoring and controlling and is automatically stored in the database. The platform is supported with the additional application of the database management system field, such as MySQL database. The designated raw data is divided into several groups. The group of data combines the dataset as the input for performing the prediction model. This integrated system of IoT and machine learning has also been explored by the scientist [12] [13].

A prediction model is generated using the machine learning technique in processing information from the available data sets split by the irrigation system scenario. ML performs with the help of complementary libraries, such as Numpy, Pandas, Matplotlib, Seaborn, and Scikit-learn of Python programming language that has been widely used in the related field to build the learning systems in producing a desired model [14]. A classification approach of supervised learning applies to develop the model strengthen by the designated output label that manually inputs. The model development is conduct in an open-source Jupyter notebook as it is interactive to the user to conclude the analysis of each machine learning phase. There are several procedures as a reference to perform the model testing. The train/test split procedure splits the dataset automatically and implements a decision tree algorithm to train the data. The evaluation of the model performance performs with three classification metrics, such as accuracy score, confusion matrix, and classifi-

cation table that correlated to each other as the analysis to determine the prediction model's quality. The model is saved to the local computer model for repeatedly implementing the model testing, which applies to the pickle library's given input.

In this thesis, a designated system aims to generate an ideal prediction model of bok choy growth crop, especially in the seeding phase. This research uses a machine learning technique to develop the prediction model, adopts an integrated system that has been established for creating the dataset. The smart farm system and MySQL database for data storing is discussed in the previous thesis published internally in the university field [15] [16]. This thesis design discussion is divided into five chapters, from the introduction to the conclusion and suggestion section.

1.2 Problem Identification

The production growth rate of the bok choy as one of the plants classified in mustard greens is still slightly increased. It may lead to a lack of supply if it is not comparable with the market's availability. One of the issues is caused by the lack of fully-compiled science and knowledge. It is difficult for farmers to determine superior seed that impacts the quality and quantity of the bok choy. However, market demand for bok choy is still outstanding and necessary for household and industrial material cooking. Nowadays, the bok choy optimal seeding model is still unavailable as a reference guide for generating predictions to work in the farmer environment.

1.3 Objective and Contribution

The objective and contribution of this thesis are as follows:

1. Capable of producing an ideal bok choy sedding period growth prediction model for the farmer environment.

2. Able to conduct a comparative analysis of performance metrics result of accuracy and precision parameter based on the confusion matrix and classification report achieves the ideal prediction model.

3. Obtain a suited well-work global system to generate the prediction based on the dataset in the seeding phase from the previous thesis.

1.4 Scope of This Thesis

The scope of this thesis are as follows:

1. The parameters of the dataset are greenhouse temperature, greenhouse humidity, greenhouse light intensity, moist level, the time received from the Raspberry Pi sensor, and the plant's height, irrigation level, and label that manually inputs that is taken in a greenhouse.
2. The range of the temperature is based on lowland.
3. The dataset contains data from two scenarios of the bok choy seeding phase, automatic and manual irrigation system.
4. The prediction model obtains by using the Python version 3.7.3 language.
5. Analysis and data modeling of the bok choy growth prediction model conducts in Jupyter Notebook stores on the local computer.
6. The dataset used as the input data is in the form of the Excel file.
7. The Python library are NumPy for data mining, Pandas for data preparation, Matplotlib and Seaborn for data visualization, and Scikit-learn for machine learning.
8. Machine learning technique used is the classification approach of supervised learning with a decision tree algorithm.
9. The bok choy growth dataset contains attributes in the seeding period gathered from 4th of January 2021 - 15th of January 2021 at 08.00 A.M. to 02.20 P.M. in Buah Batu region.
10. Camera's output is disabled in building the prediction model.
11. The parameter analyses are accuracy and precision score are based on the accuracy score and classification report.
12. The network security for maintaining dataset and prediction model guarantee is not discussed.
13. It covers only the learning stage of research based on interest and not ready for automatic integrated global system integration.

1.5 Research Method

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

1. WP1: Literature Study
   Gather the study and knowledge of generating dataset with Python programming language and formulate a prediction model with machine learning algorithm from credible references such as journals or the conference paper published by the Institute of Electrical and Electronics Engineers (IEEE) and other credible publications. Also, the articles, journals, and books are issued by the official organization or government.

2. WP2: Model Systematic Design
   Organize the global system design for a prediction model from the data mining phase to the prediction model testing phase. Determine the suitable dataset and machine learning algorithm for the desired model based on the literature review.

3. WP3: Application and Library Preparation
   Ensure the Jupyter Notebook and Python is installed on the local computer. Assure the library needed to perform the prediction model in the notebook is provided by install with the appropriate command and instruction available from a credible source on the internet.

4. WP4: Dataset Preparation
   Load the dataset in an Excel file from the MySQL database. Compile each dataset added with new attributes and variables column of the growth-related factor and label manually in Microsoft Excel.

5. WP5: The Prediction Model Realization
   Perform the establishment of a prediction model in the Jupyter Notebook. Retrieves two datasets that have been prepared. Implement dataset modeling from the data preparation stage to the model evaluation stage.

6. WP6: Analysis of the Prediction Model
   Analyze the classification report table of the prediction model generated from the concatenating dataset by the best accuracy and precision value used as the parameter to complete an ideal prediction.

1.6 The Structure of this Thesis

The rest of the structure of this thesis is organized as follows:

- Chapter II BASIC CONCEPTS
  Review the basic related concept and theory of trustworthy studies for this thesis.

- Chapter III PROPOSED BOK CHOY GROWTH PREDICTION MODEL AND SYSTEM
  Illustrate the global system and the workflow for a prediction model generated based on the phase that occurs when the dataset processing stage in the Jupyter Notebook.

- Chapter IV RESULT AND ANALYSIS
  Inform the prediction of data modeling and a comparative analysis of the model reviewed based on classification performance metrics.

- Chapter V CONCLUSION AND SUGGESTION
  Contain the conclusion taken from the prediction model and accommodate the suggested idea use for future research in the same field of interest.