CHAPTER 2
LITERATURE AND THEORETICAL BACKGROUND

2.1. Literature Review

2.1.1. Previous Research

Saleem *et. al.* (2014) conducted a research to explore the impact on the cash flow behavior of Jordanian hotel companies on their profit. Based on the research, there exist a few findings on each of the variables that exist in the cash flow such as index of return and the profit, and also a non statistically significant inverse relationship on the index of operating cash flow and quality of profit.

Based on research conducted by Khan (2013), data mining activity does not particularly need specific software to be conducted. Generally, data mining activities use specific programming languages such as R, Python, and SQL. But it does not limit the main purpose of the activity itself when performing data mining by using other software. One of the most powerful software that is commonly used to process data in tabular form is Microsoft Excel. For example, research conducted by Khan (2013) studied student's performance to improve student's performance and decrease the ratio of dropouts in an educational institution.

Shi (2017) discussed purchasing management as the main topic in constructing a decision support system. In the research, a several theorem was created based on purchase quantity, financing decisions, purchase timing and capital or budget as the constraint. The model created in this research is created to explain the behavior on how the funds of a company is utilized.

Dutra (2020) took a perspective in understanding the system on a research about system identification and state estimation using maximum likelihood estimation. These methods are commonly used to evaluate the uncertainties estimation and detect various problem in the term of parameterization and other estimations. By understanding the method used, there will be so many matters and cases that can be examined to better understand the system observed.

Zhang *et. al.* (2019) constructed a simulation to show the length of stay distribution in a hospital using a Lognormal-based mixture model for robust fitting. In the research, there are three lognormal-based mixture model as its base that utilizes exponential distribution, gamma distribution, and lognormal distribution. The

parameters of each mixture-model are estimated using expectation-maximization algorithm, and the parameter of the models are put to test into a simulation to further validate the findings.

Phonsuwan and Kachitvichyanukul (2011) conducted a research on constructing the decision support system using Arena Simulation for Micro and Small Business in Thailand. In this research, the rapid-prototype conceptual model is used to create the decision-making support system based on the management policies regulated by the government. The main area being evaluated is the supply chain production and its impact to the lead time reductions.

Guo *et. al.* (2020) constructed a decision support system for information processing in an intelligent manufacturing using data mining as the main technology. In the research, the IoT-based DSS is constructed to help the decision-maker to make decision by providing many angles and perspectives by modeling, classifying, and clustering big sets of data, also finding the correlation between data. Based on the result of research, it is proven that data mining utilization can help a manufacturer achieve a system that is more effective.

Kadri *et. al.* (2013) created a simulation-based decision support system for the emergency department. In this research, a discrete-event simulation model was constructed to understand the strain situation in the emergency departments. The model explains the relationship between the strain situations that exist in the hospital and proposes a course of actions to handle those situations. The result of the research has proven that anticipation and strain situation management is important in the emergency departments.

Rupnik *et al.* (2006) defined the decision support system as a tool to integrate data and models to help identify and solve problems, also to help in making a decision. The research was about constructing a decision support system using data mining. In the research, a decision support system was made in the J2EE platform and Oracle-based on ODM to classify the type of customer for business owners.

Tan *et. al.* (2010) created a simulation-optimization framework for a global specialty chemicals enterprise as a decision support system. In the simulation, the behavior of entities, interactions, and other uncertainties can be observed and studied for optimization. In this research, the decision support in businesses is demonstrated in two case studies, which are on the design of optimal process reactor capacity and the policy of the optimal raw material reorder point.

2.1.2. Current Research

The essential element of performing the cash flow analysis is the cash flow itself. Cash flow generally is projected in a monthly time frame. The form of data obtained in the research conducted by Saleem *et al.* (2014) shows that the final balance or cash statement in months. A different situation that exists in the current research conducted is the data available in the final balance of the company in a daily time frame.

The current research conducted is inspired by Rupnik *et al.* (2006) where the output will be a decision support system constructed using Microsoft Excel Macro functions. The main differences from the research that was conducted before were the type of data and the model created in the research later.

2.2. Theoretical Background

2.2.1. Data Mining

Data mining is an activity to analyze sets of data to discover a certain pattern that contains implicit information (Fadli, 2011). This information can later be used to the advantage of the data owner. Data mining combines the study of statistics, computer programming, and machine learning as a tool to support the analysis.

The method is data mining is not limited to specific methods that use a certain algorithm. As long as the activity is to extract information from a big data set, the activity itself is considered as the activity of data mining (Urso *et. al.,* 2018).

Even though data mining uses computer programming in the practice, but one does not necessarily need to master the ability to code a program. There are other options for how data mining can be conducted. Data mining can be conducted by data processing software such as Microsoft Excel.

2.2.2. Data Distribution

In statistics, data distribution is a function that shows how often each of the values that exist in a data set is appearing. The most common way to understand the frequency of data appearing in a data set is to construct a histogram. Histogram is a bar chart that shows the frequency of each data that exist in a dataset. In a histogram, there exist two axes that represent different things. The horizontal axis shows the name of the value, and the vertical axis shows the frequency of the data in the horizontal axis is appearing. By observing the histogram constructed based on a certain dataset, the distribution shape of the data itself can be approximated (Siegel, 2012).

There are various types of data distribution and each of the data distributions has its common practical usage. Below is a few examples of data distribution that is commonly used in statistics.

a. Normal Distribution

A normal distribution of data is the very ideal type of data distribution with all of its randomness removed. The distribution of data if it follows the normal distribution will have most of its value of data swarmed near the average value of the data set. The statistic estimator that can explain the uniform distribution is the mean and standard deviation.

b. Uniform Distribution

A uniform distribution is the type of distribution where each value that exists in the data sets has the same probability to appear. When a data set is following a uniform distribution, the parameter needed to be obtained is the maximum and the minimum value.

2.2.3. Maximum Likelihood Estimation

Maximum likelihood estimation is a method to estimate parameters from a model. In this method, the statistical parameter is chosen based on its maximum value to make the assumption that a certain data fits a specific distribution (Erica, 2020).

In order to use the maximum likelihood estimation, some steps rules needed to be applied in the process. Those rules are:

a. Assuming a model of certain data sets

b. Deriving the likelihood function

There are several reasons why Maximum Likelihood Estimation is preferred in this research. The advantage of using Maximum Likelihood is mainly because the result provided is flexible and can be used easily in many applications and the mean squared error and variance are considered to be small. The first step in using the maximum likelihood estimation is to assume a certain probability distribution in a data set. Assuming the distribution of data can be simply visualized by creating a histogram that shows the characteristic of the data distribution.

Based on the independent and identically distributed random variables in the sample assumption, distribution parameters of θ are estimated by maximizing the following likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(x_i | \theta)$$

with xi as n observations of variable X and $f(.|\theta)$ as the density function of the assumed parametric distribution.

A package in R Studio called fitdistrplus is used to fit certain datasets to specific data distribution. In this package, the maximum likelihood is one of the methods that can be utilized. The results of utilizing this package are by observing three main parameters resulted. The parameters resulted are AIC (Akaike Information Criterion), Bayesian Information Criterion (BIC), and Loglikelihood.

By using the Maximum Likelihood Estimation, the parameter observed is the loglikelihood. The result that is going to be compared is the Loglikelihood. After fitting the data into several types of data distribution, the distribution type fitted that has the highest value of Loglikelihood is estimated to be the type of distribution that is closest to explaining the data distribution of the data.

2.2.4. Decision Support System

According to Segal (2020) on Investopedia, Decision Support System is an information system programmed to support the activities of making a decision and determining the course of actions. Decision Support System works by gathering and analyzing the data, and later, synthesizing it into information that can be used.

A decision support system can be developed for any type of industry, purpose, and business. The main function of using a Decision Support System is to interpret information in a better manner, so people from many backgrounds and professions can understand. The information generated from the Decision Support System later can be used to help one in making the best course of action according to the analyses resulted from the Decision Support System.

As it was discussed in the functionality of Microsoft Excel as a data processing tool, one of the abilities of the software is to create a decision support system. Below is an example of a decision-making support system by performing a Break-even Analysis.
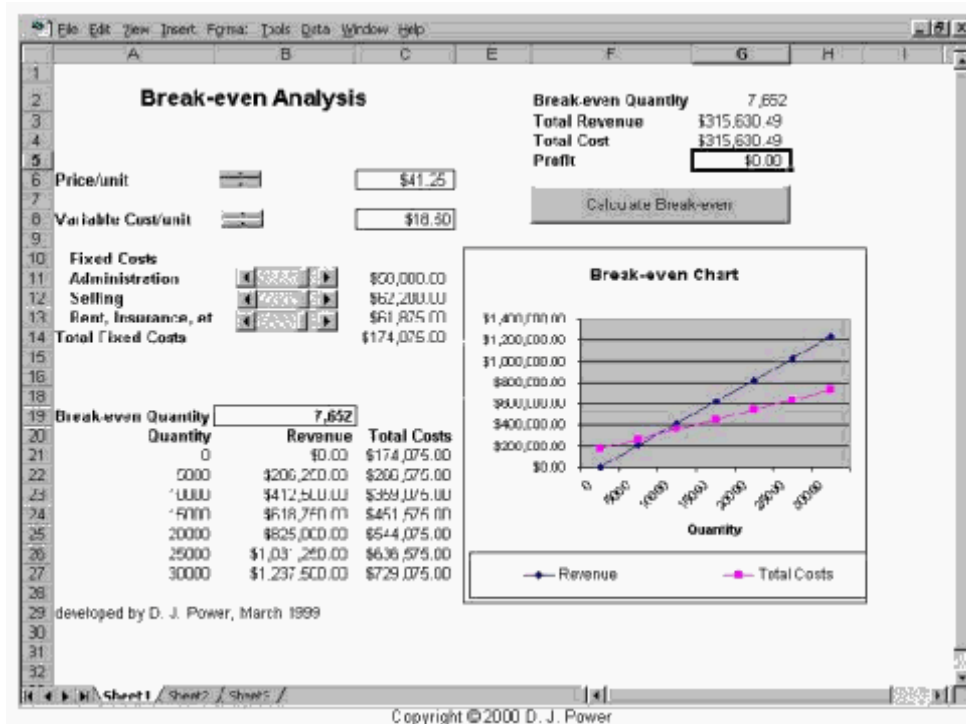
Figure 2.1. Model-Driven Decision Support System

(source: http://dssresources.com/tutorials/controls/sld001.htm)

The decision-making support system above is one of the examples of how Microsoft Excel can be utilized. The decision support system as above is used to perform the Break-even Analysis by only inputting data to the table that is already programmed to automatically calculates certain parameters and variables.

2.2.5. Simulation

A simulation is a form of replication or imitation of a system in real life that can be observed and modified its properties (Banks *et. al.*, 2009). By the definition, a simulation can also be explained as a form of a tool that can help one to observe a system or an environment without having to change anything in the real object. Generally, a simulation is developed based on "what-if" questions that will lead to the creation and the modification of the real system to find out the result.

A simulation can be created by utilizing a computer or any other instrument that is appropriate to depict the real condition of the system. The form of a simulation is often called a simulation model. A simulation model is usually constructed based on a mathematical model of a system that involves the changing of a variable's value in order to change the result of the model itself.

A simulation model is generally constructed in various versions according to the needs of the user and mostly relevant to be used as a decision making tool when

considered appropriate. The decision on how the utilization of a simulation as a decision-making tools depends on the case or the system observed. For example, a system that has a very complex behavior or considered as hard too imitate is unwise to be put into simulation since the accuracy of the system created will have smaller impact to the actual system.

By developing a simulation of a real system, there are a lot of advantages and also some disadvantages based on the system observed. Some of the advantages are, by using simulation as a decision-making tool, various insights based on the interaction of the variables that exist in the system can be obtained. Also a simulation can help its user to understand the behavior of a system without having to interact directly with the system itself. One of the disadvantages that can be found is that a simulation is often giving a result that can be considered hard to interpret. This will leads to the simulation model being irrelevant to be observed.