

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Artificial Intelligence is a vast topic and point of discussion in this past few decades. The rising era of technology, compact smart computer and communication device , and increasing speed of network technology makes big amount of data transferred all over the world (big data) give rise to Artificial Intelligence.

Artificial Intelligence itself has a few branch and sub-branches, few of them are: Natural Language Processing, Machine Learning, Knowledge Management (the main branch of Data Mining), Computer Vision. Each branch have different usage for different problems.

According to the review paper written by Amir Mosavi et al. [1], the growth in the past two decades on energy system that utilized machine learning have emerged. In 2018, the paper about energy system that utilized machine learning almost hit 600 documents. Other than in the field of energy system, the utilization of machine learning in database administration seems increasing. The premise of machine learning implementation in database administration looks rather promising. Many of the tasks that these new solutions help to automate are those found to be the most monotonous, such as installation, configuration, maintenance and troubleshooting, along with patch management.

One of them is OtterTune, the method of tuning database to get the database to perform the best performance measured in metrics such as throughput (how fast it can collect data) and latency (how fast it can respond to a request). OtterTune use Google TensorFlow and Python's scikit-learn as an implementation. The result shows that OtterTune produce DBMS configuration that could achieves

58-94% lower latency compared to default settings or configurations generated by other tuning advisors [2].

Prediction using Neural Network method has been implemented in various field. Socially, Back-Propagation Neural Network proof to be applicative, feasible, credible and good discriminative effect for suicide attempt [3]. Economically, Artificial Neural Network using Levenberg-Marquardt algorithm experimental results show that relative error of prediction is less than 3 % on stock market prediction[4].

Forecasting implementation in a computer system itself already exists with the utilization for anomaly detection. For example, this implementation perform better with Hierarchical Temporal Memory unsupervised learning method to detect point-of-failure from the past experience in the system[5].

The problem that exists in Database Administration is that Database Administrator (DBA) must have long experience in each database to predict using feeling the time database is full with the help of alarming system. New Database Administrator need to have long experience before knowing and predict when the database is full. Even experienced Database Administrator can't predict when the database is full if the database is new. Alarming system also provide a problem that it only use precentage as threshold when the database is considered "full". If the started maximum size is 100MB then 10% of free space is 10MB, but if the maximum size reached 1000MB, 10% of it is 100MB which is still a lot of free space.

Based on the previous study and current problems, this research use Deep Neural Network approach to predict the full storage database with precentage and bytes as threshold (size considered full). The database that this study use is Oracle Database. This study use datasets simulated from an imitation of real-world production database from retail company as corpus. The result of this system is

predicted result as days from the neural network to produce days left until the cycle are full.

## 1.2 Problem Formulation

Some question that needed to be answered in this study are:

1. How accurate does the Deep Neural Network perform?
2. Can it be used by Database Administrator?

## 1.3 Scope

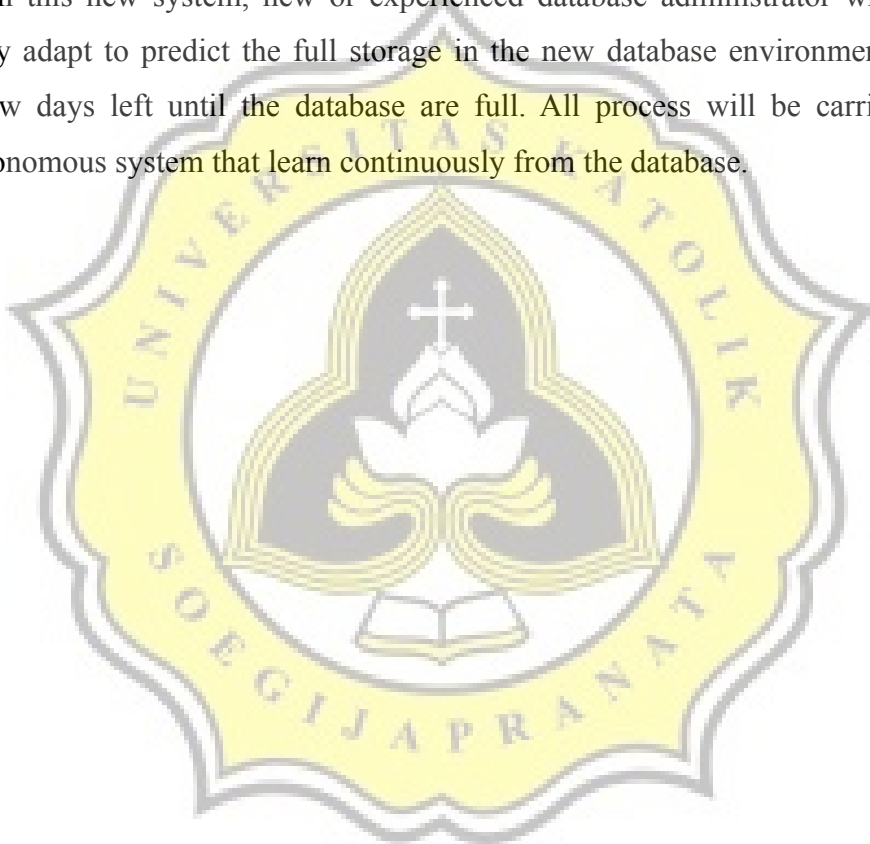
This study aims to investigate effectiveness of machine learning utilization in forecasting database full storage and the implementation result. The scope area of this study are:

- based on oracle database from a simulation of production database in retail company.
- This study will use 5 datasets: 1 tablespace with 2 tables and 6 cycles, 3 tablespace with 1 table and 3 cycles, and 1 tablespace with 1 table and 6 cycle (each cycle means from new datafile created until when the tablespace is full).
- The data sets produced 5 times per day which the tablespace size will be reduced on each time (hour) of the day.
- Using deep neural network algorithm and supervised method
- Result in a single output (time left until the tablespace is full)

The possible limitation of this study is real data sets from company's production database might be different from simulated data sets this study use. Another possible limitation of this study is the quantity of data sets might be lacking and it needs variety of tablespace to search for best use case scenario.

#### 1.4 Objective

The purpose / objective of this project is to make use of machine learning implementation in the database management process. The aimed result of this implementation is to lighten the database administrator job to monitor the system. With this new system, new or experienced database administrator who haven't fully adapt to predict the full storage in the new database environment can still know days left until the database are full. All process will be carried out by autonomous system that learn continuously from the database.



## CHAPTER 2

### LITERATURE STUDY

According to Russell and Norvig[8] there's 4 different approaches in the history to define what is Artificial Intelligence (AI). This approaches has 8 different definitions of AI (see fig 1). Top definitions are concerned with thought process and reasoning, and all on the bottom concerned with behavior. The definitions on the left measures success based on human performance, whereas on the right measure against ideal performance (rationality). A system is considered rational if it does the "right thing" given what it knows.

<p><b>Thinking Humanly</b></p> <p>"The exciting new effort to make computers think ... <i>machines with minds</i>, in the full and literal sense." (Haugeland, 1985)</p> <p>"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..." (Bellman, 1978)</p>	<p><b>Thinking Rationally</b></p> <p>"The study of mental faculties through the use of computational models." (Chamiak and McDermott, 1985)</p> <p>"The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)</p>
<p><b>Acting Humanly</b></p> <p>"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990)</p> <p>"The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)</p>	<p><b>Acting Rationally</b></p> <p>"Computational Intelligence is the study of the design of intelligent agents." (Poole <i>et al.</i>, 1998)</p> <p>"AI ... is concerned with intelligent behavior in artifacts." (Nilsson, 1998)</p>

Illustration 2.1: Some definitions of artificial intelligence, organized into four categories.

*Source: Artificial Intelligence: A Modern Approach [8]*

One of the sub field in AI is Machine Learning. In a research, Arthur Samuel [6] defining Machine Learning as "Field of study that gives computers the ability to learn without being explicitly programmed". The key part of this definition is computer behaves in the way which if done by human beings or