



**PROJECT REPORT
FORECASTING DATABASE STORAGE
USING MACHINE LEARNING**

**CHRISTOPHORUS REYHAN TANTIO A.
16.K1.0028**

**Faculty of Computer Science
Soegijapranata Catholic University
2020**

APPROVAL AND RATIFICATION PAGE

FORECASTING DATABASE STORAGE USING MACHINE LEARNING

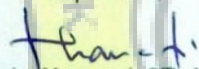
by

CHRISTOPHORUS REYHAN TANTIO A. – 16.K1.0028

This project report has been approved and ratified
by the Faculty of Computer Science on January, 9, 2020

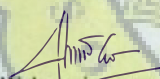
With approval,

Supervisor,

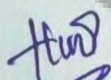

Rosita Herawati, ST., MIT
NPP : 05812004263

Examiners,

1.)


Shinta Estri Wahyuningrum, S.Si, M.Cs
NPP : 05812007272

2.)


Hironimus Leong, S.Kom, M.Kom
NPP : 05812007273


Dean of Faculty of Computer Science,

Robertus Sedawan Aji Nugroho, ST., MCompIT., Ph.D
NPP: 058.1.2004.264

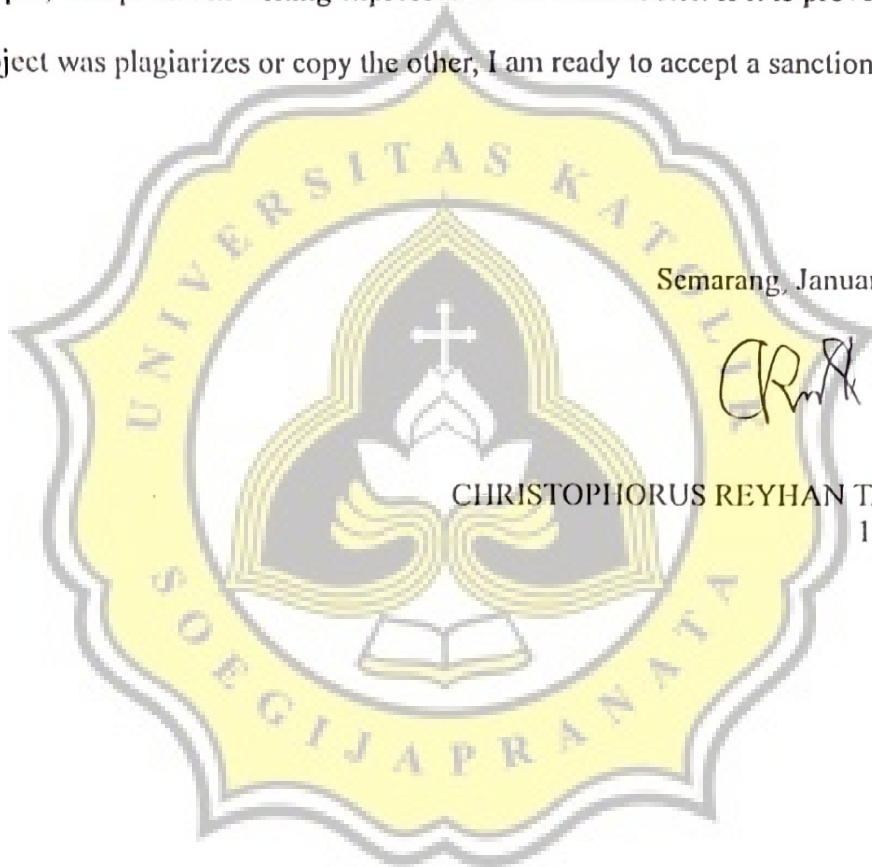
STATEMENT OF ORIGINALITY

I, the undersigned:

Name : CHRISTOPHORUS REYHAN TANTIO A.

ID : 16.K1.0028

Certify that this project was made by myself and not copy or plagiarize from other people, except that in writing expressed to the other article. If it is proven that this project was plagiarizes or copy the other, I am ready to accept a sanction.



Semarang, January, 9, 2020

CR

CHRISTOPHORUS REYHAN TANTIO A.
16.K1.0028

ABSTRACT

Space/storage management in database rely on the experience of Database Administrator (DBA) to predict time left until full and supported by script to alert when the database is full using percentage. There's no other way Database Administrator to predict days left until database is full without experience and feeling. For newly created database, Database Administrator can't predict when the database is being full in second, third, until twelve month (depends on how experienced the DBA is). The percentage threshold that is used in alert system also provide a problem because if the started size is 100MB then 10% of free space is 10MB, but if the size reached 1000MB, 10% of it is 100MB which is still a lot of free space.

This study trying to solve that problem by applying Machine Learning approach. This research use subset of machine learning, Deep Neural Network that use multiple layer of neurons in Neural Network to extract useful data from a pattern in the data. Deep Neural Network algorithms get implemented in the database system to forecasts time in days until the database considered full. The algorithm use simulation of oracle production database based on a retail company as corpus. Size and other various of data that are fetched from oracle database about a tablespace are processed so it can be used to get the best input parameters for training and predicting. The processed data that has output will be divided to training, validating, and testing, with the training and validating use most of the dataset. This testing used for monitoring the performance of Deep Neural Network and change the neural network settings and input to get the best performance based on the accuracy(what considered accurate are when the prediction is the same days as the real data or the prediction only has one day difference from the real data) . This system uses percentage and bytes as threshold that function as a flag if the current database size is considered full. When it's full, this system use the Deep Neural Network to train using past data to predict the future data. Supervised method is used with days until current data is full as output.

The result from this research is that Deep Neural Network can be used for prediction on days left until oracle database tablespace is full to some degree. Each cycle could have different pattern in adding extent and that makes the accuracy goes down. The most optimal structure is using 3 input neurons, 3 hidden layer with shape of diamond ($2 \times \text{total input} - \text{total input} + (\text{total input}/2) - \text{total input}$), and 1 output neurons.

Keyword: Database, Database Administrator, Machine Learning, Deep Neural Network, Forecasts, Oracle Database, Tablespace, Neurons, Storage

TABLE OF CONTENTS

Cover.....	i
ABSTRACT.....	iv
TABLE OF CONTENTS.....	v
ILLUSTRATION INDEX.....	vi
INDEX OF TABLES.....	vii
CHAPTER 1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Problem Formulation.....	3
1.3 Scope.....	3
1.4 Objective.....	4
CHAPTER 2 LITERATURE STUDY.....	5
CHAPTER 3 RESEARCH METHODOLOGY.....	15
3.1 Literature Study.....	15
3.2 Simulation Database.....	15
3.3 Analyzing and Designing the System.....	16
3.4 Implementing the system.....	16
3.5 Testing and Tweaking the Neural Network.....	17
CHAPTER 4 ANALYSIS AND DESIGN.....	19
4.1 Analysis.....	19
4.2 Design.....	22
CHAPTER 5 IMPLEMENTATION AND TESTING.....	30
5.1 Implementation.....	30
5.2 Testing.....	41
CHAPTER 6 CONCLUSION.....	50
REFERENCES.....	
APPENDIX.....	A

ILLUSTRATION INDEX

Illustration 2.1: Some definitions of artificial intelligence, organized into four categories.....	5
Illustration 2.2: The parts of a nerve cell or neuron.....	7
Illustration 2.3: A simple mathematical model for a neuron. The output activation of unit i is a_i (unit before this unit), $w_{i,j}$ is the weight of the link from unit i to this unit (unit j). The unit's output activation is.....	7
Illustration 2.4: 4 layer Neural Network.....	8
Illustration 2.5: ReLU, Leaky ReLU, PReLU and RReLU. For PReLU, α_i is learned and for Leaky ReLU α_i is fixed. For RReLU, α_{ji} is a random variable keeps sampling in a given range, and remains fixed in testing.....	11
Illustration 2.6: MSE Formula.....	13
Illustration 2.7: Leaky ReLU derivative.....	13
Illustration 4.1: Current oracle database storage administration process.....	20
Illustration 4.2: System database schema.....	22
Illustration 4.3: Testing database schema.....	23
Illustration 4.4: Change the settings.....	24
Illustration 4.5: Mine dataset process.....	25
Illustration 4.6: Training Process.....	26
Illustration 4.7: Predicting Process.....	28
Illustration 5.1: Sampling flowchart.....	42
Illustration 5.2: Train until cycle 1.....	42
Illustration 5.3: Train until cycle 2.....	42
Illustration 5.4: Train until cycle 3.....	43
Illustration 5.5: Lasso Model result.....	43
Illustration 5.6: Train cycle average accuracy result in diagram.....	48
Illustration 6.1: Example plot graph of loss. Taken from tblspc1, train until cycle 1, sample 1.....	AL
Illustration 6.2: Example of plot diagram of prediction and real value. Taken from tblspc1, train until cycle 1, sample 1.....	AL

INDEX OF TABLES

Table 2.1: Error rate of CIFAR-100 Network in Network with different activation function.....	12
Table 2.2: Error rate of CIFAR-10 Network in Network with different activation function.....	12
Table 2.3: Multi-classes Log-Loss of NDSB Network with different activation function.....	12
Table 5.1: Train cycle average accuracy result.....	47

