DATA QUALITY MINING IMPLEMENTATION ON SKK MIGAS SISTEM OPERASI TERPADU (SOT) DATA

By

Bobby Suryajaya
22012208

A thesis submitted to the Faculty of

ENGINEERING AND INFORMATION TECHNOLOGY

in partial fulfillment of the requirements
for the
MASTER'S DEGREE
in

INFORMATION TECHNOLOGY

SWISS GERMAN UNIVERSITY
EduTown BSDCity
Tangerang 15339
INDONESIA

JANUARY 2014
Revision after Thesis Defense on February 18, 2014

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.

Bobby Suryajaya

_____ _____
Student Date

Approved by:

Dr. Ir. Mohammad A. Amin Soetomo, M.Sc.

_____ _____
Thesis Advisor Date

Dr. Harya Widiputra

_____ _____
Thesis Co-Advisor Date

Alva Erwin, M.Sc.

_____ _____
Thesis Co-Advisor Date

Dr. Ir. Gembong Baskoro, M.Sc.

_____ _____
Dean Date

ABSTRACT

DATA QUALITY MINING IMPLEMENTATION FOR SKK MIGAS SISTEM OPERASI TERPADU (SOT) DATA

By
Bobby Suryajaya
Dr. Ir. Mohammad A. Amin Soetomo, M.Sc., Advisor
Dr. Harya Widiputra, Co-Advisor
Alva Erwin, M.Sc., Co-Advisor

SWISS GERMAN UNIVERSITY

SKK Migas implemented SOT (Sistem Operasi Terpadu) to allow SKK Migas system to retrieve specific information from PSC Contractor data source within integrated online system. It is expected that at the completion of SOT program, there will be huge transactional data that will be exchanged on daily basis, and without Data Mining, it will be extremely difficult to perform analysis. Meanwhile, this research will focus on poor data quality that may be resulted from poor data acquisition process, and this research objective is to examine whether data mining techniques can improve SOT Data quality or not by utilizing CRISP-DM methodology. This research found that SOT Data contain some quality issues including missing values, negative values, useless attributes, and outliers that may affect analysis result. Finally, this research proves that Data Quality Mining processes and algorithms can improve SOT Data quality, which expected to contribute to SKK Migas data analysis improvement.

Keywords: Data Quality Mining, DQM, SKK Migas, CRISP-DM, Missing Values, and Outliers

DEDICATION

I dedicate this thesis to my family, Faculty of Engineering and Information Technology in SGU (Swiss German University), and Indonesia Oil & Gas IT society.

Hope this will add more knowledge to Data Mining domain especially the application within Oil & Gas industry.

ACKNOWLEDGMENTS

I would like to thank Allah Jalla wa ‘Alaa for the abundance of grace and gift, so I can finish this work as planned and without experiencing significant obstacles.

I also wish to thank everyone for his or her full support and commitment on completion of my thesis, especially:

- My family for their patience and support they had given from the beginning until it is completed.
- My colleagues at SGU MIT Class 2012-2013: Frieda Putri Aryani, Umabala Devarakonda, Fenty Simanjuntak, Aldo Elam Majiah, and Jusak Rahardja.
- My Advisor Dr. Ir. Mohammad A. Amin Soetomo, M.Sc.
- My Co-Advisor Dr. Harya Widiputra.
- My Co-Advisor Alva Erwin, M.Sc.
- SKK Migas Deputy of Internal Control Bpk. Dr. Budi Ibrahim.
- SKK Migas Management Information System Division Head Bpk. Handoyo Budi Santoso.
- SKK Migas Strategic Data & Information Head Bpk. Syukri Waldi.
- SKK Migas Strategic Data Processing Head Ibu Sri Wahyuning Astuti and all her team members.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES