A STUDY ON DATA MINING ALGORITHMS FOR CHURN PREDICTION IN AN ORGANIZATION

By

KEVIN KURNIAWAN
11302010

BACHELOR'S DEGREE
in

INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

July 2017

Revision after the Thesis Defense on 20th July 2017

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.

Kevin Kurniawan
Student — Date

Approved by:

Alva Erwin, ST, M.Sc., MTI
Thesis Advisor — Date

Dipl. –Inf. Kho I Eng
Thesis Co-Advisor — Date

Dr. Ir. Gembong Baskoro, M.Sc
Dean — Date

ABSTRACT

A STUDY ON DATA MINING ALGORITHMS
FOR CHURN PREDICTION IN AN ORGANIZATION

By

Kevin Kurniawan
Alva Erwin, ST, M.Sc., MTI, Advisor
Dipl. –Inf. Kho I Eng, Co-Advisor

SWISS GERMAN UNIVERSITY

Employee churn and customer churn are two common problems for an organization as it indicates organization loss of resources and profits. Hence, implementing churn prediction helps to minimize those problems by detect potential churn employees or customers and apply targeted strategies to prevent them churning. By minimizing churn rate could maximize organizations revenue and productivity. This research presents churn prediction for human resource as a tool to prevent loss of valuable employees and for marketing as a tool to prevent customers' loss in telecom industry. The author compared three common data mining algorithms on both datasets in predicting churn. The experiment result shows the overview prediction performance of each data mining algorithms in different unit of analysis. Overall, the best prediction was performed by Random Forest with PCC score 85.9% in employee churn and 95.4% in customer churn, F-Measure score 26.2% in employee churn and 82.2% in customer churn. However, Random Forest's F-Measure score in employee was not the highest as Neural Network has F-Measure score 47.1%. From this research, it can be concluded that Random Forest is a good data mining algorithm to perform churn prediction in different industries. Moreover, more comparison is recommended for real implementation in order to achieve best prediction model.

Keywords: customer churn, employee churn, data mining, churn prediction, telecom, Random Forest

DEDICATION

I dedicate this works for my beloved family, friends, and country: Indonesia.

ACKNOWLEDGEMENTS

As completion of this thesis work, I would like to express my gratitude for several parties involved:

Swanly Kurniawan & Lilywaty Utama - as parents, they motivated me throughout the process of making. In addition, they supported me financially to make this thesis work done.

Alva Erwin and Kho I Eng - as the advisor and co-advisor of my thesis work. They helped to shape the ideas of the thesis. From the starting point, they had confidence on me to accomplish the thesis topics. Besides, they supported me to analyzed problems encountered during the process, problem solving, and directed me to think critically.

Swiss German University - as the organization who provided me this opportunity to create thesis work and arranged all the needs in administration and thesis defense.

Finally, I also appreciated the support from my colleagues as they had being supportive and giving critics and ideas throughout the process.

TABLE OF CONTENTS

Page