# TEXT CLASSIFICATION TECHNIQUES USED TO FACILITATE CYBER TERRORISM INVESTIGATION

By

David Allister Simanjuntak

A Bachelor's Thesis
Submitted to the Faculty of

INFORMATION TECHNOLOGY

in partial fulfillment of the
requirements for the Degree of

BACHELOR OF SCIENCES
INFORMATION TECHNOLOGY

SWISS GERMAN UNIVERSITY
EduTown BSDCity
Tangerang – 15339
Island of Java, Indonesia
www.sgu.ac.id

July 2010

Revised after Thesis Defense on 27 July 2010

## 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, not material which to a substantial extent has been accepted for the award of may other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.

________________________________________ __________________

David Allister Simanjuntak Date

Approved by:

________________________________________ __________________

Anto Satriyo Nugroho, Dr.Eng Date

________________________________________ __________________

Charles Lim, Msc., ECSA, ECSP, ECIH, CEH, CEI Date

________________________________________ __________________

Chairman of the Examination Steering Committee Date

# ABSTRACT

## TEXT CLASSIFICATION TECHNIQUES USED TO FACILITATE CYBER TERRORISM INVESTIGATION

By

David Allister Simanjuntak

SWISS GERMAN UNIVERSITY
Bumi Serpong Damai

Anto Satriyo Nugroho, Major Advisor

The rising of computer violence, such as Distributed Denial of Service (DDoS), web vandalism, and cyber bullying become more serious issues when they are politically motivated and intentionally conducted to generate fear in society. These kinds of activities are categorized as cyber terrorism. As the number of such cases increase, the availability of information regarding these actions is required to facilitate experts in investigating cyber terrorism. Meanwhile, web mining is one of significant technologies applied to extract information from the Web. In this case, web mining facilitates data acquisition related to cyber terrorism information from the Web. This research aims to create text classification technique based upon number of occurrences of certain relevant words in the term of Cyber Terrorism. This research compared the result of accuracy of several algorithms including Naïve Bayes, Nearest Neighbor, Support Vector Machine (SVM), Decision Tree, and Multilayer Perceptron Neural Network. The result shows that SVM outperform by achieving 100% of accuracy. According to this result, it concludes the excellent performance of SVM in handling high dimensional of data.

*Keywords* – cyber terrorism, data mining, feature selection, text classification, web mining

## DEDICATION

I dedicate this thesis to my family who has supported me all the way.

## ACKNOWLEDGMENTS

The author wishes to give utmost gratitude to Jesus Christ, without His love and blessing all of the work and effort will be nothing, praise and glory only for His name.

The author would also give the best regard to Anto Satriyo Nugroho, Dr.Eng and Charles Lim, Msc for the guidance, encouragement, patience, and their assistance from the beginning until the end of this thesis research. So many thanks also addressed to all of the lecturers and staffs of Swiss German University.

Special regards also comes for my fellow students in Information Technology 2006, I gratitude you guys for these beautiful four years in my life, especially Endy Chen for being  such a collaborative partner in conducting this research.

In addition, the author would also like to give thanks to Armando, Irfan Suwadi, Ivan Firdausi, and Ryan Andhika Perdana for being such a mutual friends and very helpful to me.

Last but not least, the author would give special thanks to Alicia Item for being my perfect distraction but also my spirit booster.

# TABLE OF CONTENTS