REPUTATION SCORING FAKE NEWS USING TEXT MINING

By

Ahmad Firdaus
2-1551-015

MASTER'S DEGREE
In

INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGY

SWISS GERMAN UNIVERSITY
Prominence Office Tower
Tangerang 15143
Indonesia

Revision after Thesis Deffense 27th 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.

Ahmad Firdaus

_____ _____

Student Date

Approved by:

Dr. Maulahikmah Galinium, M.Sc

_____ _____

Thesis Advisor Date

Alva Erwin, M.Sc, MTI, ST

_____ _____

Thesis Co-Advisor Date

Dr. Ir. Gembong Baskoro, M.Sc

_____ _____

Dean Date

ABSTRACT

REPUTATION SCORING FAKE NEWS USING TEXT MINING

By

Ahmad Firdaus

Dr. Maulahikmah Galinium, M.Sc, S.Kom Advisor

Alva Erwin, M.Sc, MTI, ST Co-Advisor

SWISS GERMAN UNIVERSITY

The classification of hoax news or news with incorrect information is one of the text categorization applications. Like text-based categorization of machine applications in general, this system consists of pre-processing and execution of classification models. In this study, experiments were conducted to select the best technique in each sub-process by using 1200 articles hoax and 600 article no hoax collected manually. This research Tried experimenting to determine the best preprocessing stages between stop removal and stemming and showing the results of the deception Tree alghoritma achieving an accuracy of 100% concluded above naive bayes more stable level of accuracy in the number of datasets used in all candidates . Information gain, TFIDF and GGA based on using Naive Bayes algorithm, supporting Vector Machine and Decision Tree no significant percentage change occurred on all candidates. But after using GGA (Optimize Generation) feature selection there is an increase of accuracy level The results of a comparison of classification algorithms between Naive Bayes, decision trees and Support Vector machines combined with the GGA feature selection method for classifying the best result is generated by the selection of GGA + Decission Tree feature on candidate 2 (Paslon2) 100% and in the selection of the Information Gain + Decission Tree Feature selection with the lowest accuracy Candidate 3 at 36.67%, but overall improvement of accuracy Occurred on all alghoritma after using feature selection and Naive bayes are faster in processing time and Decision Tree is the longest for processing time and Naive bayes more stable level of accuracy in the number of datasets used in all candidates.

Keyword Classification, Pre-processing, Feature Selection, Accuracy

COPYRIGHT

DEDICATION

I dedicated this thesis to my lovely family who have prayed for me every day, my parents for all the support and encouragement, colleagues in MIT Swiss German University batch 17.

ACKNOWLEDGEMENTS

Gratitude and thanks to Allah SWT matchless so I could finish this thesis on time. In addition also to all the colleagues and interviewee and expert panel who provide support and time either in the interview or discussion are:

1. Ika Suryani and Arsya – My lovely Wife and Son
2. All friend Batch 17 Swiss German University

Thanks are expressed Dr. Maulahikmah Galinium, M.Sc as my Advisor Alva Erwin, M.Sc as Co-Advisor.

Contents

List of Figures