

**DATA MINING BASED USER PROFILING FOR ADVERTISEMENT SERVING**

By

Wisely Liu Dennis

12112006

BACHELOR'S DEGREE

In

INFORMATION TECHNOLOGY

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY

EduTown BSD City

Tangerang 15339

Indonesia

August 2016

**Revision after the Thesis Defense on 21 July 2016**

## 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.

Wisely Liu Dennis

\_\_\_\_\_  
Student

\_\_\_\_\_  
Date

Approved By:

Alva Erwin, ST, M.Sc., MTI

\_\_\_\_\_  
Thesis Advisor

\_\_\_\_\_  
Date

Maulahikmah Galinium, S.Kom, M.Sc, PhD

\_\_\_\_\_  
Thesis Co - Advisor

\_\_\_\_\_  
Date

Dr. Ir. Gembong Baskoro, M.Sc

\_\_\_\_\_  
Dean

\_\_\_\_\_  
Date



## ABSTRACT

### DATA MINING BASED USER PROFILING FOR ADVERTISEMENT SERVING

By

Wisely Liu Dennis  
Alva Erwin, ST, M.Sc., MTI, Advisor  
Maulahikmah Galinium, S.Kom, M.Sc, PhD, Co-Advisor

SWISS GERMAN UNIVERISTY

Advertisement serving on website is a prosperous business with huge market and millions of dollar prospect. By placing right advertisement at right time and place to right people, advertiser can increase their revenue by huge margin. The question is how advertiser and broker can push the right advertisement to the right user. User profiling can be used to analyze user's behavior and predict what kind of advertisement should be served to the website user. Data mining approach can be harnessed to help with user profiling process. With data mining technique, user's trace can be used as data source for behavior analysis. This research is trying to do user profiling based on their browsing history stored on proxy server. Their browsing history will serve as the basis of content crawling for content analysis using Multinomial Naïve Bayes classifier. The result of profiling then will be used as the basis for serving advertisement to user. The result will be validated by asking user's preferences and comparing it with profile generated by classifier engine. The result of the advertisement serving will be determined by user questionnaire method. The final result of this research is 70% accuracy of user profile generated by Multinomial Naïve Bayes. This accuracy value prove that user profile generated by using Multinomial Naïve Bayes can be used on advertisement serving

*Keywords: Advertisement Serving, Data Mining, Multinomial Naïve Bayes, Text Classification, Machine Learning.*





## DEDICATION

I dedicate this thesis work for my parents that always supporting me.

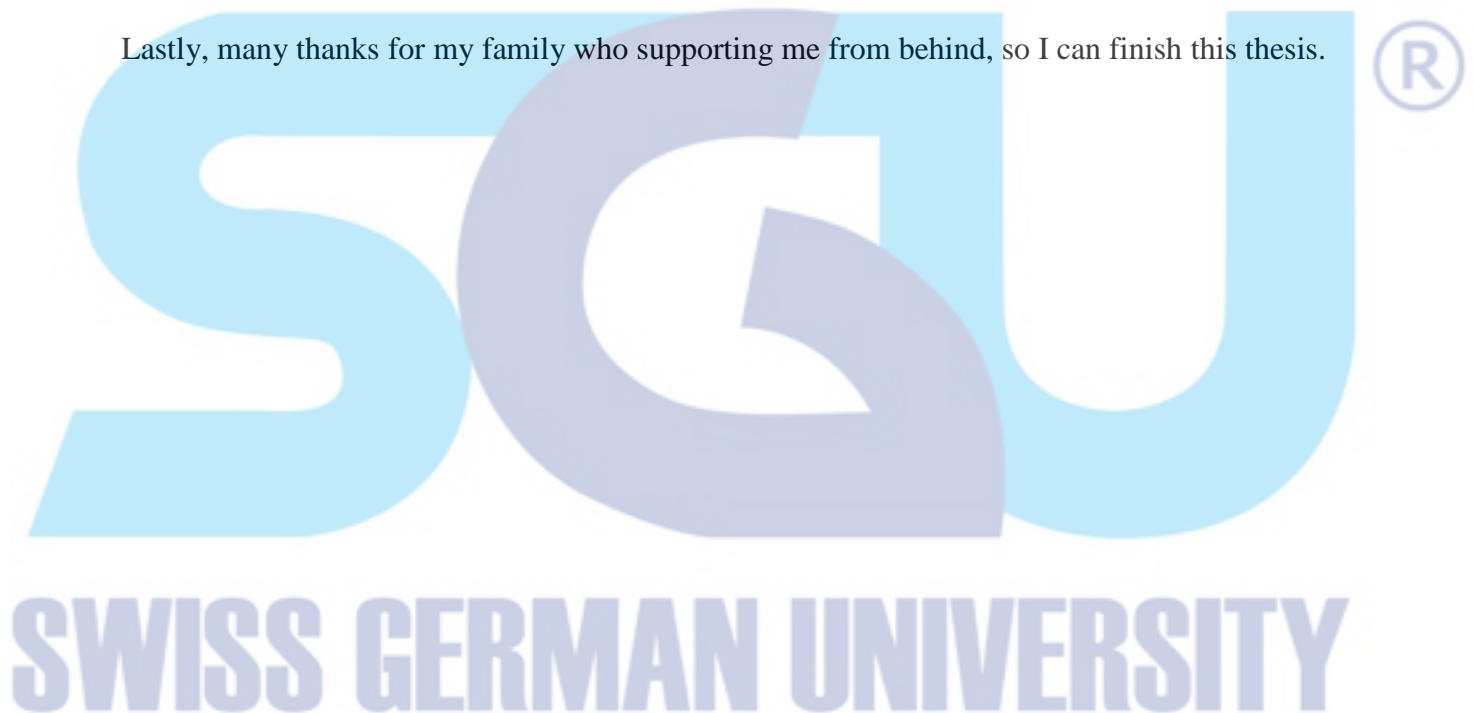


## ACKNOWLEDGEMENTS

First, I would like to thank God for all his blessing so I can finish this thesis. Also, for Mr Alva Erwin, and Mr Maulahikmah Galinium as my thesis Advisor and Co-Advisor. Also, thanks Mr James Purnama for his guidance so I can do this thesis in the right direction.

I am grateful the company X for the support on data gathering and all of my colleagues there for their cooperation for helping and validating my thesis result.

Lastly, many thanks for my family who supporting me from behind, so I can finish this thesis.





## TABLE OF CONTENTS

<b>STATEMENT BY THE AUTHOR.....</b>	<b>2</b>
<b>ABSTRACT.....</b>	<b>4</b>
<b>DEDICATION.....</b>	<b>7</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>8</b>
<b>TABLE OF CONTENTS .....</b>	<b>9</b>
<b>LIST OF FIGURES .....</b>	<b>12</b>
<b>1. CHAPTER 1 – INTRODUCTION.....</b>	<b>13</b>
1.1. Background.....	13
1.2. Research Problem .....	14
1.3. Research Objectives.....	14
1.4. Significance of Study.....	14
1.5. Research Questions.....	15
1.6. Hypothesis.....	15
1.7. Scope.....	15
1.8. Limitations .....	15
<b>2. CHAPTER 2 - LITERATURE REVIEW .....</b>	<b>16</b>
2.1. Related Works.....	16
2.2. Ad Serving .....	17
2.2.1. Ad Server Architecture.....	17
2.2.2. Revive Ad Server .....	18
2.3. Behavioral targeting.....	19
2.3.1. Browsing History Analysis .....	19
2.3.2. Website Impressiveness Analysis .....	20
2.3.3. User profiling .....	20
2.4. Data Mining .....	21
2.4.1. Recommender System.....	22
2.4.2. Naïve Bayes Classification.....	22
2.4.3. Naïve Bayes Multinomial.....	23
2.4.4. Text Classification.....	24
2.4.5. TF-IDF (Term Frequency – Inverse Document Frequency) .....	25
2.4.6. Keyword Generation .....	26

<b>3.</b>	<b>CHAPTER 3 - RESEARCH METHODOLOGY .....</b>	<b>27</b>
3.1.	Methodology Overview .....	27
3.2.	Data Collection .....	29
3.2.1.	Web Log Analysis Data .....	31
3.2.2.	Web Crawling Result Data.....	33
3.2.3.	Data Mining Result Data.....	35
3.3.	Data analysis .....	37
3.3.1.	First Training Set Generation.....	37
3.3.2.	Second Training Set Generation .....	38
3.3.3.	User Profiling.....	40
3.4.	Data Verification.....	42
3.4.1.	First Training Set Generation.....	43
3.4.2.	Second Training Set Generation .....	45
3.4.3.	User Profiling .....	45
<b>4.</b>	<b>CHAPTER 4 - RESULT ANALYSIS .....</b>	<b>47</b>
4.1.	Data Collection .....	47
4.1.1.	Web Log Data .....	47
4.1.2.	Web Crawling Result Data.....	49
4.2.	Data Analysis .....	50
4.2.1.	First Training Set Generation.....	50
4.2.2.	Second Training Set Generation .....	52
4.2.3.	User Profiling.....	53
4.3.	Data Verification.....	55
4.3.1.	First Training Set Generation.....	55
4.3.2.	Second Training Set Generation .....	56
4.3.3.	User Profiling.....	57
4.4.	Advertisement Serving.....	59
<b>5.</b>	<b>CONCLUSION AND FUTURE WORKS .....</b>	<b>60</b>
5.1.	Conclusion .....	60
5.2.	Future Works .....	60
<b>6.</b>	<b>REFERENCES.....</b>	<b>61</b>
	<b>APPENDIX 1.....</b>	<b>63</b>

**CURRICULUM VITAE..... 64**

