



PROJECT REPORT
ANALYSIS OF PERSONALITY ASPECT
CLASSIFICATION FROM THE RECRUITMENT
INTERVIEW TEST WITH LDA TOPIC
MODELING

Evander Reinhart Mulyono
16.K1.0007

Faculty of Computer Science
Soegijapranata Catholic University
2020

APPROVAL AND RATIFICATION PAGE

ANALYSIS OF PERSONALITY ASPECT CLASSIFICATION FROM THE
RECRUITMENT INTERVIEW TEST WITH LDA TOPIC MODELING

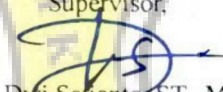
by

Evander Reinhart Mulyono – 16.K1.0007

This project report has been approved and ratified
by the Faculty of Computer Science on January, 9, 2020

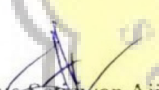
With approval,

Supervisor,



YB. Dwi Setianto, ST., M.Cs
NPP: 058.7.2017.021



Examiners,

1.)


Robertus Setiawan Aji Nugroho, Ph.D
NPP: 058.1.2004.264

2.)


Shinta Estri Wahyuningrum, S.Si., M.Cs
NPP: 058.1.2007.272


Dean of Faculty of Computer Science,

Robertus Setiawan Aji Nugroho, Ph.D
NPP: 058.1.2004.264

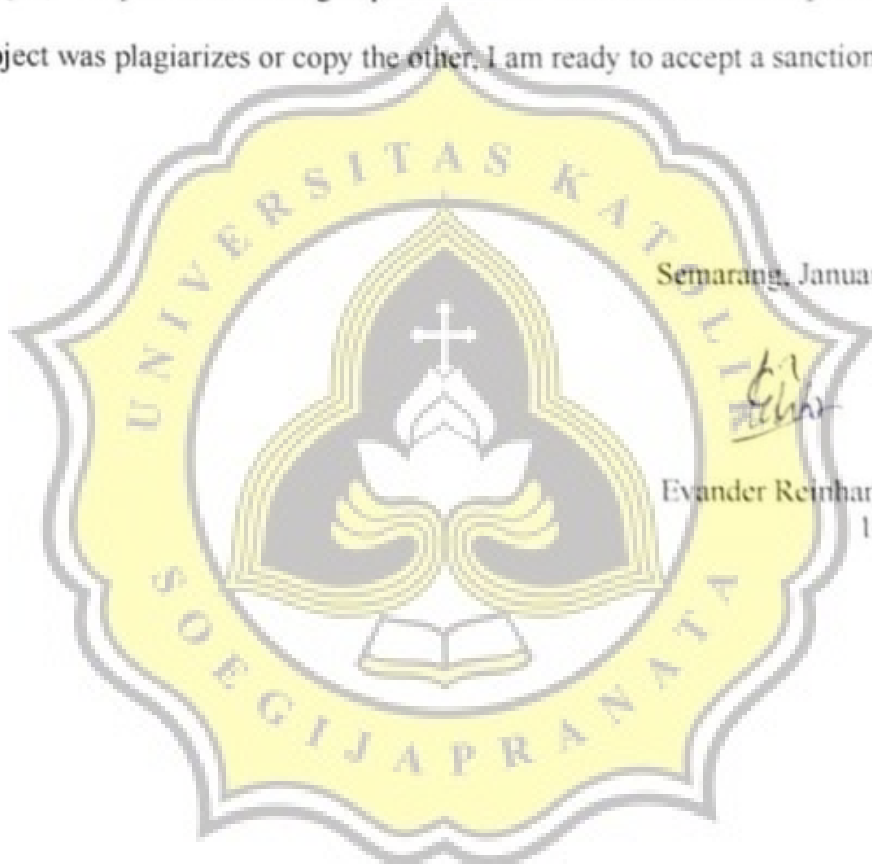
STATEMENT OF ORIGINALITY

I, the undersigned:

Name : Evander Reinhart Mulyono

ID : 16.K1.0007

Certify that this project was made by myself and not copy or plagiarize from other people, except that in writing expressed to the other article. If it is proven that this project was plagiarizes or copy the other, I am ready to accept a sanction.



Semarang, January, 9, 2020

Evander
Evander Reinhart Mulyono
16.K1.0007

ABSTRACT

The development of data processing and data storage techniques in this modern era has been growing rapidly. The size of data collected nowadays are very large and from these large scaled data we can get useful information. This information can be used for several things especially can be used for determining aspects of a person's personality from a recruitment interview. From the results, we can store those data and study it and we can get the patterns then those patterns will help classifying person's personality aspect from the Motivation, Work Enthusiasm, and Self-awareness.

To help classifying, LDA algorithm is used. But not the original version of LDA but the supervised variant of the LDA (L-LDA). There are two approaches that will be used and compared. The first one views that a text consists of common words and the second one views that a document does not consist of common words. From these approaches, will be compared and analyze which approach is more effective and better.

The result summaries that the second approach is more effective because the first approach depends on the training data used. If the dataset has varied words, it will be difficult to detect words that are common.

Keyword: Topic Model, Classification, L-LDA

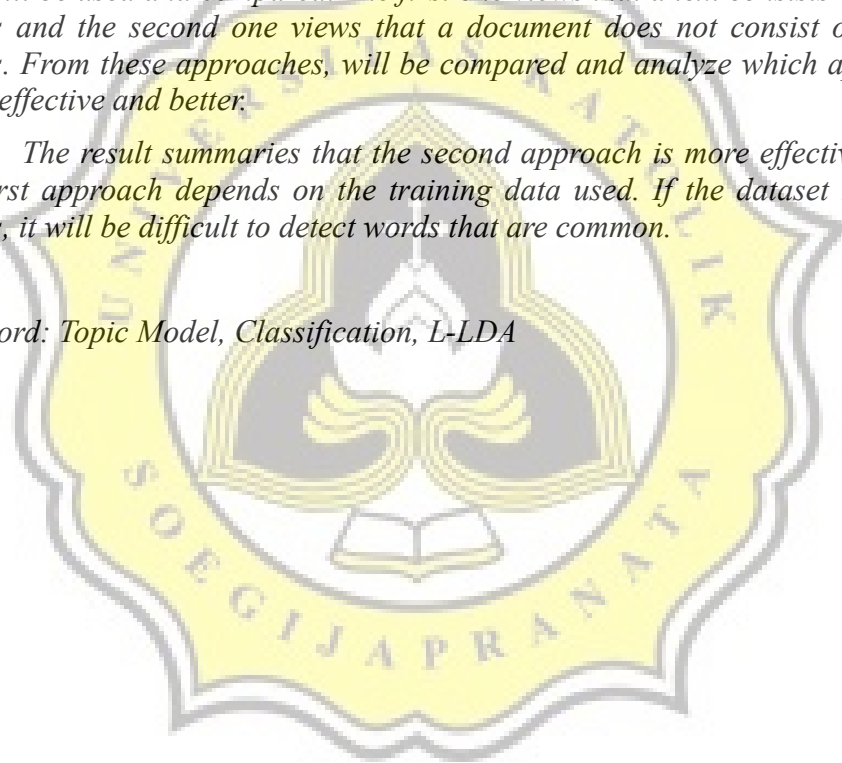


TABLE OF CONTENTS

Cover.....	i
APPROVAL AND RATIFICATION PAGE.....	ii
STATEMENT OF ORIGINALITY.....	iii
ABSTRACT.....	iv
TABLE OF CONTENTS.....	v
ILLUSTRATION INDEX.....	vi
INDEX OF TABLES.....	vii
CHAPTER 1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Problem Formulation.....	3
1.3 Scope.....	3
1.4 Objective.....	3
CHAPTER 2 LITERATURE STUDY.....	5
CHAPTER 3 RESEARCH METHODOLOGY.....	8
3.1 Data Collection.....	8
3.2 Data Preprocessing.....	8
3.3 Creating Classification Model.....	8
3.4 Predicting Test Data.....	9
CHAPTER 4 ANALYSIS AND DESIGN.....	10
4.1 Analysis.....	10
4.2 Desain.....	33
CHAPTER 5 IMPLEMENTATION AND TESTING.....	35
5.1 Implementation.....	35
5.2 Testing.....	42
CHAPTER 6 CONCLUSION.....	44
REFERENCES.....	
APPENDIX.....	A

ILLUSTRATION INDEX

Illustration 4.1: Flowchart.....	33
Illustration 5.1: Testing Classification Program.....	42



INDEX OF TABLES

Table 4.1: Interview Questions Table.....	10
Table 4.2: Answer Criteria for Motivation’s Interview Question.....	11
Table 4.3: Answer Criteria for Work Enthusiasm’s Interview Question.....	11
Table 4.4: Answer Criteria for Self-awareness’s Interview Question.....	12
Table 4.5: Answer Examples from Motivation's Interview Question.....	12
Table 4.6: Answer Examples from Work Enthusiasm’s Interview Question.....	12
Table 4.7: Answer Examples from Self-awareness’s Interview Question.....	13
Table 4.8: Topic Vocabulary.....	16
Table 4.9: Word or Term Vocabulary Example.....	16
Table 4.10: Corpus Examples.....	17
Table 4.11: Lambda Matrix Based on Table 4.10.....	18
Table 4.12: Lambda Matrix Example.....	18
Table 4.13: Topic Vocabulary with Common or Neutral Topic.....	19
Table 4.14: Lambda Matrix Based on Table 4.10 with Common Topic.....	19
Table 4.15: Example Corpus.....	20
Table 4.16: Topic Vocabulary Based on Table 4.15.....	20
Table 4.17: Word or Term Vocabulary Based on Table 4.15.....	20
Table 4.18: Lambda Based on Table 4.15.....	20
Table 4.19: Replace Each Words Inside Documents With Their Ids and Add a Topic in Every Word.....	21
Table 4.20: Document Topic Matrix Based on Table 4.19.....	21
Table 4.21: Topic Term Matrix Based on Table 4.19.....	21
Table 4.22: Current Condition of Document 1.....	23
Table 4.23: Current Condition of Document Topic Matrix.....	23
Table 4.24: Current Condition of Topic Term Matrix.....	23
Table 4.25: $\alpha(d)$ Based on the Example.....	23
Table 4.26: Formula Calculation (Left Side).....	24
Table 4.27: Formula Calculation (Right Side).....	24
Table 4.28: New Topic Probability Calculation For Word I.....	24
Table 4.29: Current Condition of Document 1 (Changed I Topic to Cat).....	25
Table 4.30: Updated Condition of Document Topic Matrix Based on Table 4.29.....	25
Table 4.31: Updated Condition of Topic Term Matrix Based on Table 4.29.....	25
Table 4.32: Final Document Topic Matrix Based on Table 4.15.....	26
Table 4.33: Final Topic Term Matrix Based on Table 4.15.....	26
Table 4.34: Distribution of Topics Inside Documents Based on Table 4.15.....	27
Table 4.35: Distribution of Words or Terms Inside Topics Based on Table 4.15.....	27
Table 4.36: Words Distributions in Motivation.....	29
Table 4.37: Words Distributions in Work Enthusiasm.....	29
Table 4.38: Words Distributions in Self-awareness.....	29
Table 4.39: Words Distributions in Motivation.....	30
Table 4.40: Words Distributions in Work Enthusiasm.....	30
Table 4.41: Words Distributions in Self-awareness.....	30

Table 4.42: Word Distributions inside Motivation Model that occurs in new document.....	31
Table 5.1: Result Table with Common Topic or Document Contains Neutral Words.....	43
Table 5.2: Result Table without Common Topic or Document Does not Contain Neutral Words.....	43
Table 6.1: Result Table with Common Topic or Document Contains Neutral Words.....	W
Table 6.2: Result Table without Common Topic or Document Does not Contain Neutral Words.....	Z

