

TABLE OF CONTENTS

ENDORSEMENT PAGE	i
HALAMAN PERNYATAAN	i
REVISION TABLE	ii
REVISION TABLE 1	ii
REVISION TABLE 2	ii
REVISION TABLE 3	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vi
LIST OF FIGURES	vii
FOREWORD	viii
INTISARI.....	ix
ABSTRACT	x
CHAPTER I INTRODUCTION	1
1.1 Research Background.....	1
1.2 Research Problem.....	2
1.3 Research Scope	3
1.4 Research Objectives.....	3
1.5 Research Benefits	4
CHAPTER II LITERATURE REVIEW	5
CHAPTER III THEORETICAL BASIS	8
3.1 Semantic Similarity	8
3.2 Pre-Processes Text.....	9
3.3 Vectorization	10
3.3.1 <i>TF-IDF Matrix</i>	10
3.3.2 <i>Sentence Transformer</i>	12
3.3.3 <i>DistilBERT</i>	13
3.4 Similarity Measures	16
3.4.1 <i>Cosine Similarity</i>	16
CHAPTER IV RESEARCH METHOD	19
4.1 Research Description	19
4.2 Data Collection	19
4.3 System Architecture	20
4.4 System Data Process.....	21
4.4.1 <i>Pre-process Text</i>	21
4.4.2 <i>TF - IDF</i>	21
4.4.3 <i>Sentence Transformer</i>	22
4.4.4 <i>DistilBERT</i>	24
4.4.5 <i>Similarity metrics</i>	25
4.5 Performance Evaluation	25
CHAPTER V IMPLEMENTATION	27
5.1 Library Preparation.....	27
5.2 Data Preparation.....	27
5.2.1 <i>Data Preprocessing</i>	28

5.2.2 <i>Vectorization</i>	29
5.3 Cosine Similarity	31
5.4 Data Evaluation	31
CHAPTER VI RESEARCH RESULT	33
6.1 Data Result.....	33
CHAPTER VII CONCLUSION	37
LIST OF REFERENCE	38
ATTACHMENT 1	41
ATTACHMENT 2	42
ATTACHMENT 3	63