



TABLE OF CONTENTS

UNDERGRADUATE THESIS	I
UNDERGRADUATE THESIS	II
APPROVAL PAGE	III
STATEMENT	V
FOREWORD	VII
TABLE OF CONTENTS	VIII
TABLE OF FIGURES	XII
TABLE OF TABLES	XV
ABSTRACT	XVII
CHAPTER I INTRODUCTION	1
1.1 Research Background	1
1.2 Research Problem	3
1.3 Research Scope	4
1.4 Research Objectives	4
1.5 Research Benefits	4
CHAPTER II LITERATURE REVIEW	5
CHAPTER III THEORETICAL BASIS	10
3.1 Clickbait	10
3.2 Preprocessing	11
3.2.1 Special Characters Removal	12
3.2.2 Case Folding	12



3.2.3 Stop-Words Removal	13
3.2.4 Tokenization	13
3.3 Feature Engineering	14
3.4 Feature Extraction	14
3.4.1 Part of Speech Tagging	14
3.4.2 Unigram and Bigrams	15
3.4.3 Coleman-Liau Score	16
3.4.4 RIX and LIX Indices	16
3.4.5 Formality Measure	18
3.4.6 Cosine Similarity	19
3.4.7 TF-IDF (Term Frequency – Inverse Document Frequency)	20
3.5 Machine Learning	21
3.5.1 Logistic Regression	21
3.5.2 Support Vector Machine (SVM)	22
3.6 Performance Evaluation	23
3.6.1 Accuracy	23
3.6.2 Precision	23
3.6.3 Recall	24
3.6.4 F-1 Score	24
3.6.5 K-Fold Cross Validation	24
CHAPTER IV RESEARCH METHODOLOGY	26
4.1 Research Description	26
4.2 Tools and Material	26
4.2.1 Tools	26
4.2.2 Materials	27
4.3 Research Phase	28
4.4 System Design	29
4.5 Dataset	30
4.6 Preprocessing	31
4.7 Feature Engineering	31
4.7.1 Headline	31



4.7.2 Article	32
4.7.3. Similarity	33
4.8 Feature Extraction	35
4.8.1 Headline	35
4.8.2 Article	38
4.8.3 Similarity	40
4.9 SVM	42
4.10 Logistic Regression	42
4.11 Evaluation Design	44
CHAPTER V IMPLEMENTATION	45
5.1 Dataset	45
5.1.1 Joining Dataset	47
5.2 Preprocessing	48
5.2.1 Combine Data	52
5.2.2 Selecting articles with > 100 words	53
5.3 Feature Engineering and Feature Extraction	54
5.3.1 Headline	55
5.3.2 Article	64
5.3.3 Similarity	69
5.4 Machine Learning Model	71
5.4.1 Headline	73
5.4.2 Headline + Article	74
5.4.3 Headline + Article + Similarity	75
5.6 Evaluation Measure	76
CHAPTER VI RESULT AND DISCUSSION	79
6.1 Dataset Overview	79
6.2 Feature Analysis	80
6.3 Classification	100



UNIVERSITAS
GADJAH MADA

Clickbait Detection for News Article in Bahasa Indonesia using Article Informality
HILDA, Sigit Priyanta, S.Si., M.Kom., Dr; Yunita Sari, S.Kom., M.Sc., Ph.D
Universitas Gadjah Mada, 2020 | Diunduh dari <http://etd.repository.ugm.ac.id/>

CHAPTER VII CONCLUSION	106
7.1 Conclusion	106
7.2 Future Works	107
REFERENCES	108
APPENDICES	114



UNIVERSITAS
GADJAH MADA

Clickbait Detection for News Article in Bahasa Indonesia using Article Informality
HILDA, Sigit Priyanta, S.Si., M.Kom., Dr; Yunita Sari, S.Kom., M.Sc., Ph.D

Universitas Gadjah Mada, 2020 | Diunduh dari <http://etd.repository.ugm.ac.id/>