DAFTAR ISI

ABSTRAK i

ABSTRACT ii

LEMBAR PENGESAHAN iii

LEMBAR PERNYATAAN ORISINALITAS iv

KATA PENGANTAR v

DAFTAR ISI vi

DAFTAR GAMBAR x

DAFTAR TABEL xii

DAFTAR SIMBOL xiii

DAFTAR ISTILAH xiv

BAB I PENDAHULUAN 1

I.1 Latar Belakang 1

I.2 Perumusan Masalah 7

I.3 Tujuan Penelitian 7

I.4 Batasan Penelitian 7

I.5 Manfaat Penelitian 8

I.6 Sistematika Penulisan 8

BAB II TINJAUAN PUSTAKA 10

II.1 Pemilu Presiden dan Wakil Presiden 10

II.2 YouTube 11

II.3 *Data Mining* 11

II.4 *Text Mining* 12

II.5 *Natural Language Processing* 13

II.6 Analisis Sentimen 14

II.7 Analisis Sentimen Berbasis *Lexicon* 14
II.8 *Machine Learning* 15
II.9 Data Preprocessing 17
II.9.1 *Case Folding* 18
II.9.2 *Unwanted Characters Removal* 19
II.9.3 *Tokenization* 20
II.9.4 *Normalization* 20
II.9.5 *Stop word Removal* 20
II.9.6 *Stemming* dan *Lemmatization* 21
II.10 *Term Frequency-Inverse Document Frequency* (TF-IDF) 22
II.11 *Support Vector Machine* (SVM) 24
II.12 *Confusion Matrix* 26
II.13 Kurva ROC-AUC 27
II.14 *Co-occurrence Network* 28
II.15 Perbandingan Algoritma Klasifikasi 30
II.16 Penelitian Terdahulu 32
BAB III METODOLOGI PENELITIAN 37
III.1 Kerangka Berpikir 37
III.2 Sistematika Penyelesaian Masalah 38
BAB IV ANALISIS DAN PERANCANGAN 42
IV.1 *Initialization* 42
IV.2 *Data Selection* 43
IV.3 *Data Preprocessing* 44
IV.3.1 *Case Folding* 44
IV.3.2 Penghapusan Angka 45
IV.3.3 Penghapusan URL 47

IV.3.4 Penghapusan Simbol 48
IV.3.5 Penghapusan *Whitespace* 49
IV.3.6 *Tokenization* 51
IV.3.7 *Normalization* 52
IV.3.8 Penghapusan *Stop Word* 54
IV.3.9 *Stemming* 56
IV.3.10 *Labelling* 57
IV.4 *Data Transformation* 64
IV.4.1 *Term Frequency – Inverse Document Frequency* (TF-IDF) 64
IV.4.2 *Splitting* Data 69
IV.5 Klasifikasi Menggunakan *Support Vector Machine* (SVM) 70
IV.6 *Interpretation*/*Evaluation* 73
IV.6.1 *Confusion Matrix* 73
IV.6.2 Classification Report 74
IV.6.3 Kurva ROC dan AUC 75
IV.6.4 Interpretasi dengan *Co-occurrence Network* 76
BAB V IMPLEMENTASI DAN PENGUJIAN 78
V.1 Model Klasifikasi Algoritma SVM 78
V.1.1 Model Klasifikasi SVM *Dataset* Anies – Muhaimin 78
V.1.2 Model Klasifikasi SVM *Dataset* Prabowo - Gibran 80
V.1.3 Model Klasifikasi SVM *Dataset* Ganjar – Mahfud 81
V.2 Evaluasi Performa Model 83
V.2.1 *Confusion Matrix* 83
V.2.1.1 *Confusion Matrix* Pelabelan Manual 84
V.2.1.2 *Confusion Matrix* Pelabelan *Lexicon* 86
V.2.1.3 Rangkuman *Confusion Matrix* Pelabelan Manual 89

V.2.1.4 Rangkuman *Confusion Matrix* Pelabelan *Lexicon* 91
V.2.2 *Classification Report*..... 92
V.2.2.1 *Classification Report* Pelabelan Manual..... 92
V.2.2.2 *Classification Report* Pelabelan *Lexicon* 94
V.2.3 Kurva ROC-AUC..... 95
V.2.3.1 Kurva ROC-AUC Pelabelan Manual..... 95
V.2.3.2 Kurva ROC-AUC Pelabelan *Lexicon* 98
V.2.4 Analisis Performa Model 101
V.3 Visualisasi *Co-occurrence Network*..... 103
BAB VI KESIMPULAN DAN SARAN 107
VI.1 Kesimpulan..... 107
VI.2 Saran 108
DAFTAR PUSTAKA 109