

DAFTAR ISI

HALAMAN PENGESAHAN	iii
PRAKATA	v
ARTILAMBANG DAN SINGKATAN	vi
ABSTRACT	vii
INTISARI	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Keaslian Penelitian	3
1.4 Tujuan Penelitian	8
1.5 Batasan Penelitian	8
1.6 Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	9
2.1 Tinjauan Pustaka	9
2.2 Landasan Teori	14
2.2.1 Natural Language Processing	14
2.2.2 Hidden Markov Model Pos Tagger	15
2.2.3 Term Frequency-Inverse Document Frequency	20
2.2.4 Named Entity Recognition	21
2.2.5 Co-Training	24
2.2.6 Support Vector Machine	27
2.2.7 Naive Bayes	31
2.2.8 Logistic Regression	33
BAB III METODOLOGI	36
3.1 Alat dan Bahan	36
3.1.1 Alat	36
3.1.2 Bahan	37
3.2 Jalannya Penelitian	38
3.2.1 Persiapan Data	39
3.2.2 Pra-proses Teks	43

3.2.3	Klasifikasi Kalimat Kutipan	46
3.2.4	Ekstraksi Kata Unik dengan PoS Tagger	47
3.2.5	Klasifikasi Entitas Kata Unik dengan NER	51
3.3	Evaluasi Prediksi Kalimat Kutipan	55
BAB IV	HASIL DAN PEMBAHASAN	58
4.1	Persiapan Data	58
4.2	Pra-proses Teks	58
4.3	Hasil Klasifikasi Kutipan	61
4.4	Hasil Ekstraksi Kata Unik dengan PoS Tagger	62
4.5	Hasil Klasifikasi Entitas Kata Unik dengan NER	63
4.6	Hasil Evaluasi Prediksi Kalimat Kutipan	66
BAB V	KESIMPULAN DAN SARAN	69
5.1	Kesimpulan	69
5.2	Saran	69
DAFTAR PUSTAKA		71
LAMPIRAN		L-1
L.1	Algorithm	L-1
L.2	Confusion Matrix dari klasifikasi	L-2