

**ANALISIS SENTIMEN KASUS FERDY SAMBO PADA
TWITTER MENGGUNAKAN ALGORITMA NAIVE
BAYES DENGAN SELEKSI FITUR**

SKRIPSI

**Oleh:
Arief Priyadi Pratama
201910225037**



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2023**

LEMBAR PERSETUJUAN PEMBIMBING

LEMBAR PERSETUJUAN PEMBIMBING

(DIGUNAKAN UNTUK TUGAS AKHIR)

Judul Tugas Akhir : Analisis Sentimen Kasus Ferdy Sambo pada *Twitter*

Menggunakan Algoritma Naive Bayes dengan

Seleksi Fitur

Nama Mahasiswa : ARIEF PRIYADI PRATAMA

Nomor Pokok Mahasiswa : 201910225037

Program Studi/Fakultas : Informatika/Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 01 Agustus 2023

Jakarta, 08 Agustus 2023

MENYETUJUI

Pembimbing I

Pembimbing II


Prima Dina Atika, S.Kom., M.Kom.
NIDN: 0311037107


Siti Setiawati, S.Pd., M.Pd.
NIDN: 0313107904

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Judul Skripsi : ANALISIS SENTIMEN KASUS FERDY SAMBO
PADA *TWITTER* MENGGUNAKAN ALGORITMA
NAIVE BAYES DENGAN SELEKSI FITUR

Nama Mahasiswa : ARIEF PRIYADI PRATAMA

Nomor Induk Mahasiswa : 201910225037

Program Studi/Fakultas : Informatika/Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 01 Agustus 2023

Jakarta, 08 Agustus 2023

MENGESAHKAN

Ketua Tim Penguji : Rakhmat Purnomo, S.Pd., S.Kom., M.Kom.
NIDN: 0322108201

Penguji II : Aida Fitriyani, S.Kom., M.M.S.I.
NIDN: 0302078508

Penguji III : Prima Dina Atika, S.Kom., M.Kom.
NIDN: 0311037107

MENGETAHUI,

Ketua
Program Studi Informatika

Ahmad Fathurrozi, S.E., M.M.S.I.
NIP: 2012486

Dekan
Fakultas Ilmu Komputer

Dr. Dra. Tyastuti Sri Lestari, M.M.
NIP: 1408206

LEMBAR PERNYATAAN BUKAN PLAGIASI



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Arief Priyadi Pratama
NPM : 201910225037
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Analisis Sentimen Kasus Ferdy Sambo pada *Twitter*
Menggunakan Algoritma *Naive Bayes* dengan Seleksi Fitur

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Jakarta, 08 Agustus 2023
Penulis



Arief Priyadi Pratama

ABSTRAK

Arief Priyadi Pratama. 201910225037. Analisis Sentimen Kasus Ferdy Sambo pada *Twitter* Menggunakan Algoritma *Naive Bayes* dengan Seleksi Fitur. Bekasi: Fakultas Ilmu Komputer. Universitas Bhayangkara Jakarta Raya. 2023.

Kepolisian Negara Republik Indonesia (Polri) merupakan salah satu lembaga pemerintah yang berintegritas sebagai pengayom keamanan masyarakat. Namun, Polri terkena malu ketika sebuah kasus pembunuhan berencana oleh Kadiv Propam Irjen Ferdy Sambo menjadi viral. Kasus ini menarik antusiasme masyarakat yang tinggi sehingga mencuatkan beragam opini di *Twitter*. Kompleksitas dalam mengklasifikasikan opini menjadi sebuah tantangan karena belum adanya standar pengklasifikasian opini. Untuk mengatasi tantangan tersebut, diperlukan sebuah teknik analisis sentimen yang dapat mengklasifikasikan opini dengan akurat. Oleh karena itu, penelitian ini bertujuan untuk menganalisis opini masyarakat di *Twitter* dengan menggunakan algoritma *Naive Bayes* yang dipadukan dengan metode seleksi fitur *Chi Square* dan *Information Gain* menggunakan metodologi CRISP-DM. Sebanyak 1100 data digunakan dalam penelitian ini, dengan 715 data berlabel positif dan 385 data berlabel negatif, yang telah diberi label oleh pakar bahasa. Dalam upaya meningkatkan akurasi, tiga model dibangun dan diuji serta hasil akurasi terbaik dilakukan uji prediksi teks. Hasil penelitian menunjukkan bahwa model *Naive Bayes* mampu mencapai akurasi sebesar 73%. *Chi Square* dikombinasikan dengan *Naive Bayes*, akurasi meningkat menjadi 75%. Selain itu, *Information Gain* juga menghasilkan akurasi sebesar 74%. Percobaan dilakukan dengan tiga rasio *split* data, yaitu 90:10, 80:20, dan 70:30. Dapat disimpulkan bahwa model *Naive Bayes* dengan seleksi fitur *Chi Square* dan rasio *split* data 80:20 mengalami peningkatan akurasi sebesar 2% sebagai hasil akurasi terbaik, yaitu sebesar 75%.

Kata Kunci: Analisis Sentimen, Kasus Viral, Ferdy Sambo, CRISP-DM, Naive Bayes, Chi Square, Information Gain, Twitter

ABSTRACT

Arief Priyadi Pratama. 201910225037. Sentiment Analysis of Ferdy Sambo's Case on Twitter Using the Naive Bayes Algorithm with Feature Selection. Bekasi: Faculty of Computer Science. Jakarta Bhayangkara University. 2023.

The Indonesian National Police (Polri) is a government institution with integrity as a protector of public security. However, the National Police were embarrassed when a case of premeditated murder by the Head of the Propam Division, Inspector General Ferdy Sambo, went viral. This case attracted high public enthusiasm so that various opinions emerged on Twitter. The complexity in classifying opinions is a challenge because there is no standard for classifying opinions. To overcome these challenges, we need a sentiment analysis technique that can classify opinions accurately. Therefore, this study aims to analyze public opinion on Twitter using the Naive Bayes algorithm combined with the Chi Square feature selection method and Information Gain using the CRISP-DM methodology. A total of 1100 data were used in this study, with 715 data labeled positive and 385 data labeled negative, which were labeled by linguists. In an effort to improve accuracy, three models were built and tested and the best accuracy results were tested by predicting text. The results showed that the Naive Bayes model was able to achieve an accuracy of 73%. Chi Square combined with Naive Bayes, accuracy increased to 75%. In addition, Information Gain also produces an accuracy of 74%. The experiment was carried out with three data split ratios, namely 90:10, 80:20, and 70:30. It can be concluded that the Naive Bayes model with Chi Square feature selection and a data split ratio of 80:20 has an accuracy increase of 2% as the best accuracy result, which is equal to 75%.

Keywords: Sentiment Analysis, Viral Case, Ferdy Sambo, CRISP-DM, Naive Bayes, Chi Square, Information Gain, Twitter

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Arief Priyadi Pratama
NPM : 201910225037
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalty-Free Right)**, atas karya ilmiah saya yang berjudul :

Analisis Sentimen Kasus Ferdy Sambo pada *Twitter* Menggunakan Algoritma *Naive Bayes* dengan Seleksi Fitur

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 08 Agustus 2023
Yang Menyatakan



Arief Priyadi Pratama

KATA PENGANTAR

Puji syukur peneliti panjatkan kepada Allah SWT. Tuhan Yang Maha Esa yang telah melimpahkan karunia-Nya sehingga peneliti dapat menyelesaikan skripsi yang disusun sebagai salah satu syarat untuk mencapai gelar sarjana Strata satu (S1) pada program studi Informatika Universitas Bhayangkara Jakarta Raya. Penelitian ini tidaklah dapat terwujud dengan baik tanpa adanya bimbingan, dorongan, saran-saran dan bantuan dari berbagai pihak. Pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada kedua Orang Tua yang selalu memberikan semangat dan dukungan serta kepada:

1. Bapak Irjen Pol (Purn) Prof. Dr. Drs. Bambang Karsono, S.H., M.M. selaku Rektor Universitas Bhayangkara Jakarta Raya;
2. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M. selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya;
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. selaku Ketua Program Studi Informatika;
4. Bapak Prio Kustanto, S.T., M.Kom. selaku dosen pembimbing akademik kelas A1;
5. Ibu Prima Dina Atika, S.Kom., M.Kom. selaku dosen pembimbing satu penyusunan laporan skripsi;
6. Ibu Siti Setiawati, S.Pd., M.Pd. selaku dosen pembimbing dua penyusunan laporan skripsi;

7. Ibu Hafizah, S.S., M.Pd. selaku validator instrumen dalam penentuan palabelan data dalam penyusunan laporan skripsi;
8. Bapak Muhammad Yasir, S.Si., M.Kom. selaku dosen mentor analisis data pemrograman dalam penyusunan laporan skripsi;
9. Teman-teman ANGK (Erik, Lutfi, Jamal, Eka, Dhea, Dila) yang selalu menemani berproses bersama mulai dari awal kuliah;
10. Teman-teman seperjuangan pada program studi Informatika Universitas Bhayangkara Jakarta Raya terutama kelas A1 yang selalu memberikan dukungan dan selalu membantu dalam menyelesaikan skripsi ini.

Peneliti menyadari bahwa penelitian ini jauh dari kesempurnaan, untuk itu peneliti mohon maaf apabila terdapat kekurangan dan kesalahan baik dari penuturan penulisan. Oleh karena itu, peneliti mengharapkan kritik dan saran yang membangun. Terimakasih banyak.

Bekasi, 17 Juli 2023

Arief Priyadi Pratama

NPM: 201910225037

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iii
ABSTRAK	iv
ABSTRACT	v
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	4
1.3. Rumusan Masalah	5
1.4. Batasan Masalah.....	5
1.5. Tujuan Penelitian.....	5
1.6. Manfaat Penelitian.....	6
1.7. Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 Penelitian Terdahulu.....	9
2.1.1 Tabel <i>Related Research / State of The Art</i>	12
2.2 Sentimen.....	16
2.3 Media Sosial.....	16
2.4 <i>Twitter</i>	18
2.5 <i>Text Mining</i>	19
2.6 <i>Data Mining</i>	20
2.6.1 Seleksi Fitur <i>Chi Square</i>	23
2.6.2 Seleksi Fitur <i>Information Gain</i>	26
2.6.3 Algoritma <i>Naive Bayes</i>	28
2.6.4 <i>Confusion Matrix</i>	35
2.7 Analisis Sentimen.....	38
2.8 CRISP-DM	39

2.8.1	<i>Text Pre-processing</i>	44
2.9	Pemrograman <i>Python</i>	47
BAB III	METODOLOGI PENELITIAN	48
3.1	Kerangka Penelitian	48
3.1.1	<i>Business Understanding Phase</i>	49
3.1.2	<i>Data Understanding Phase</i>	49
3.1.3	<i>Data Preparation Phase</i>	50
3.1.4	<i>Modelling Phase</i>	51
3.1.5	<i>Evaluation Phase</i>	51
3.1.6	<i>Result</i>	51
3.2	Instrumen Penelitian	52
3.3	Analisis Kebutuhan <i>Hardware</i> dan <i>Software</i>	53
3.1.1	Spesifikasi Perangkat Keras	53
3.1.2	Spesifikasi Perangkat Lunak	53
BAB IV	HASIL DAN PEMBAHASAN	54
4.1	<i>Business Understanding Phase</i>	54
4.2	<i>Data Understanding Phase</i>	54
4.3	<i>Data Preparation Phase</i>	58
4.3.1	<i>Label Encoder</i>	58
4.3.2	<i>Case Folding</i>	60
4.3.3	<i>Cleansing</i>	62
4.3.4	<i>Tokenizing</i>	64
4.3.5	<i>Normalizing</i>	65
4.3.6	<i>Stopwording</i>	68
4.3.7	<i>Stemming</i>	70
4.3.8	<i>Final Dataset</i>	71
4.3.9	<i>Splitting Dataset</i>	73
4.3.10	<i>Transforming Dataset</i>	74
4.3.11	Seleksi Fitur <i>Chi Square</i>	75
4.3.12	Seleksi Fitur <i>Information Gain</i>	76
4.3.13	Algoritma <i>Naive Bayes</i>	76
4.4	<i>Modelling Phase</i>	80
4.4.1	Model 1: <i>Naive Bayes (NB)</i>	81
4.4.2	Model 2: <i>Naive Bayes dan Chi Square (NB+CHI)</i>	82

4.4.3	Model 3: <i>Naive Bayes</i> dan <i>Information Gain</i> (NB+IG)	82
4.5	<i>Evaluation Phase</i>	83
4.5.1	<i>Evaluation Model 1: Naive Bayes</i> (NB).....	83
4.5.2	<i>Evaluation Model 2: Naive Bayes</i> dan <i>Chi Square</i> (NB+CHI).....	85
4.5.3	<i>Evaluation Model 3: Naive Bayes</i> dan <i>Information Gain</i> (NB+IG).....	87
4.6	<i>Result</i>	89
4.6.1	Percobaan Rasio <i>Split Data</i>	90
4.6.2	<i>WordCloud</i>	92
4.6.3	Uji Prediksi Model	94
BAB V	PENUTUP	95
5.1	Kesimpulan.....	95
5.2	Saran.....	96
DAFTAR PUSTAKA		97
LAMPIRAN		101



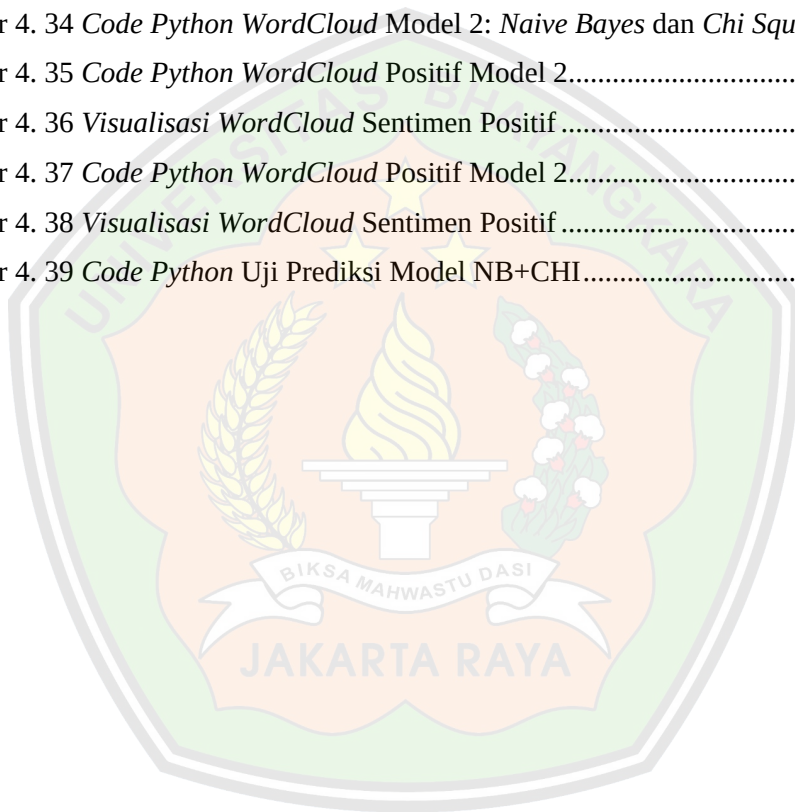
DAFTAR TABEL

Tabel 2. 1 <i>State of The Art</i>	12
Tabel 2. 2 Kajian Obat Nyeri pada Lansia.....	30
Tabel 2. 3 Distribusi probabilitas ML bersyarat untuk Pengobatan yang diberikan targetnya Nyeri.....	32
Tabel 2. 4 Distribusi probabilitas beta bersyarat untuk Pengobatan	32
Tabel 2. 5 Distribusi probabilitas bersyarat untuk Jenis Kelamin yang diberikan targetnya Nyeri	33
Tabel 2. 6 Distribusi probabilitas bersyarat untuk Umur yang diberikan targetnya Nyeri 33	
Tabel 2. 7 Distribusi probabilitas bersyarat untuk Durasi yang diberikan targetnya Nyeri	33
Tabel 2. 8 <i>Confusion Matrix</i>	35
Tabel 2. 9 <i>pre-processing text convert emoticon</i>	45
Tabel 4. 1 <i>Pre-dataset</i> tipe kategorik.....	59
Tabel 4. 2 <i>Pre-dataset</i> tipe numerik.....	59
Tabel 4. 3 Tahapan <i>Case Folding</i>	60
Tabel 4. 4 Tahapan <i>Cleansing</i>	62
Tabel 4. 5 Tahapan <i>Tokenizing</i>	64
Tabel 4. 6 Daftar Kata <i>Slang</i> Tambahan.....	66
Tabel 4. 7 Tahapan <i>Normalizing</i>	66
Tabel 4. 8 Tahapan <i>Stopwording</i>	68
Tabel 4. 9 Tahapan <i>Stemming</i>	70
Tabel 4. 10 Pembagian Jumlah Data Percobaan Rasio <i>Split Data</i>	73
Tabel 4. 11 <i>Dataset</i> Perhitungan Manual <i>Naive Bayes</i>	77
Tabel 4. 12 Jumlah Kemunculan Kata Data Uji <i>Naive Bayes</i>	78
Tabel 4. 13 Probabilitas Kemunculan Kata Data Uji <i>Naive Bayes</i>	78
Tabel 4. 14 <i>Confusion Matrix</i> Model 1: <i>Naive Bayes</i>	84
Tabel 4. 15 <i>Confusion Matrix</i> Model 2: <i>Naive Bayes</i> dan <i>Chi Square</i>	86
Tabel 4. 16 <i>Confusion Matrix</i> Model 3: <i>Naive Bayes</i> dan <i>Information Gain</i>	88
Tabel 4. 17 Hasil Uji Coba Rasio <i>Split Data</i>	90
Tabel 4. 18 Hasil <i>Evaluation Phase</i> Penelitian Sebelumnya	91

DAFTAR GAMBAR

Gambar 2. 1 Fungsi tujuan utama <i>data mining</i>	21
Gambar 2. 2 Model Naive Bayes untuk Studi Pengobatan Nyeri	31
Gambar 2. 3 CRISP-DM <i>Phases</i>	39
Gambar 2. 4 Grafik Survey <i>Kdnuggets</i>	42
Gambar 2. 5 Grafik <i>survey datascience-pm.com</i>	43
Gambar 2. 6 Grafik <i>survey Google Searches</i>	43
Gambar 3. 1 Tahapan Alur Kerangka Penelitian	49
Gambar 4. 1 <i>Node.js Twitter Crawler</i>	55
Gambar 4. 2 <i>command prompt scraping tweet harvest</i>	56
Gambar 4. 3 <i>Raw Dataset</i>	56
Gambar 4. 4 <i>Pre-dataset</i>	57
Gambar 4. 5 Diagram Batang <i>pre-dataset</i>	58
Gambar 4. 6 <i>Code Python Label Encoder</i>	60
Gambar 4. 7 <i>Code Python Case Folding</i>	61
Gambar 4. 8 <i>Code Python Cleansing</i>	63
Gambar 4. 9 <i>Code Python Tokenizing</i>	65
Gambar 4. 10 <i>Code Python Normalizing</i>	67
Gambar 4. 11 <i>Code Python Stopwording</i>	69
Gambar 4. 12 <i>Code Python Stemming</i>	71
Gambar 4. 13 <i>Code Python Final Dataset</i>	72
Gambar 4. 14 <i>Final Dataset</i>	72
Gambar 4. 15 <i>Code Python Splitting Dataset</i>	73
Gambar 4. 16 <i>Code Python Transforming Dataset CountVectorizer</i>	74
Gambar 4. 17 <i>Code Python Seleksi Fitur Chi Square</i>	75
Gambar 4. 18 <i>Code Python Seleksi Fitur Information Gain</i>	76
Gambar 4. 19 Perhitungan Manual <i>Excel Naive Bayes</i> Sentimen Positif dan Negatif	80
Gambar 4. 20 <i>Code Python Visualisasi Split Data 80:20</i>	80
Gambar 4. 21 Visualisasi Distribusi Data dengan <i>Split Data 80:20</i>	81
Gambar 4. 22 <i>Code Python Model 1 Naive Bayes</i>	81
Gambar 4. 23 <i>Code Python Model 2 Naive Bayes dan Chi Square</i>	82
Gambar 4. 24 <i>Code Python Model 2 Naive Bayes dan Information Gain</i>	83
Gambar 4. 25 <i>Code Python Evaluation Model 1 Naive Bayes</i>	84
Gambar 4. 26 <i>Confusion Matrix Model 1: Naive Bayes</i>	84

Gambar 4. 27 <i>Code Python metrics classification report Model 1: Naive Bayes (NB)</i> ...	85
Gambar 4. 28 <i>Code Python Evaluation Model 2 Naive Bayes dan Chi Square</i>	86
Gambar 4. 29 <i>Confusion Matrix Model 2: Naive Bayes dan Chi Square</i>	86
Gambar 4. 30 <i>Code Python metrics classification report Model 2: Naive Bayes dan Chi Square (NB+CHI)</i>	87
Gambar 4. 31 <i>Code Python Evaluation Model 3 Naive Bayes dan Information Gain</i>	88
Gambar 4. 32 <i>Confusion Matrix Model 3: Naive Bayes dan Information Gain</i>	88
Gambar 4. 33 <i>Code Python metrics classification report Model 3: Naive Bayes dan Information Gain (NB+IG)</i>	89
Gambar 4. 34 <i>Code Python WordCloud Model 2: Naive Bayes dan Chi Square</i>	92
Gambar 4. 35 <i>Code Python WordCloud Positif Model 2</i>	93
Gambar 4. 36 <i>Visualisasi WordCloud Sentimen Positif</i>	93
Gambar 4. 37 <i>Code Python WordCloud Positif Model 2</i>	93
Gambar 4. 38 <i>Visualisasi WordCloud Sentimen Positif</i>	94
Gambar 4. 39 <i>Code Python Uji Prediksi Model NB+CHI</i>	94



DAFTAR LAMPIRAN

Lampiran 1 Uji Plagiarisme	101
Lampiran 2 Biodata Peneliti	102
Lampiran 3 Kartu Bimbingan Tugas Akhir Dosen Pembimbing I	104
Lampiran 4 Kartu Bimbingan Tugas Akhir Dosen Pembimbing II	105
Lampiran 5 Surat Rekomendasi Dosen Pembimbing untuk Tugas Akhir	106
Lampiran 6 Pengajuan Lembar Validasi Instrumen.....	107
Lampiran 7 Lembar Validasi Instrumen Revisi.....	110
Lampiran 8 Surat Pernyataan Validitas Instrumen	113
Lampiran 9 Jurnal ICT: <i>Information Communication & Technology</i>	114

