

**ANALISIS SENTIMEN LAYANAN SHOPEEFOOD
PADA MEDIA SOSIAL TWITTER MENGGUNAKAN
ALGORITMA *NAÏVE BAYES CLASSIFIER***

Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Komputer**



*Mencerdaskan dan
Memartabatkan Bangsa*

**Nilatil Moena
1313618002**





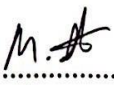


**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA
2023**

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

ANALISIS SENTIMEN LAYANAN SHOPEEFOOD PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER*

Nama : Nilatil Moena

No. Registrasi : 1313618002

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: <u>Prof. Dr. Muktiningsih N, M.Si.</u> NIP. 196405111989032001		3 / 03 / 2023
Wakil Penanggung Jawab			
Wakil Dekan I	: <u>Dr. Esmar Budi, S.Si., MT.</u> NIP. 197207281999031002		3 / 03 / 2023
Ketua	: <u>Med Irzal, M.Kom.</u> NIP. 197706152003121001		14 / 02 / 2023
Sekretaris	: <u>Ari Hendarno, S.Pd., M.Kom.</u> NIP. 198811022022031002		20 / 02 / 2023
Penguji	: <u>Muhammad Eka Suryana, M.Kom.</u> NIP. 198512232012121002		20 / 02 / 2023
Pembimbing I	: <u>Dr. Ria Arafiyah, M.Si.</u> NIP. 197511212005012004		20 / 2 / 2023
Pembimbing II	: <u>Drs. Mulyono, M.Kom.</u> NIP. 196605171994031003		20 / 02 / 2023

Dinyatakan lulus ujian skripsi tanggal : 10 Januari 2023

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul “**Analisis Sentimen Layanan ShopeeFood pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes Classifier**” yang disusun sebagai syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Ilmu Komputer Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika dikemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, 16 Januari 2023



Nilatil Moena



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Nilatil Moena
NIM : 1313618002
Fakultas/Prodi : FMIPA / Ilmu Komputer
Alamat email : nilamoena4@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

Analisis Sentimen Layanan ShopeeFood pada Media Sosial Twitter

Menggunakan Algoritma Naive Bayes Classifier

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 2 Maret 2023

Penulis

(Nilatil Moena)
nama dan tanda tangan

HALAMAN PERSEMBAHAN



*Teruntuk ibu, bapak tercinta yang tak pernah berhenti
mendoakan dan memotivasi.*

ABSTRAK

NILATIL MOENA. Analisis Sentimen Layanan ShopeeFood pada Media Sosial Twitter Menggunakan Algoritma *Naïve Bayes Classifier*. Skripsi, Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. Januari 2023.

Shopee mulai memperkenalkan layanan pesan antar makanan bernama ShopeeFood di Indonesia sejak April 2020. Setiap layanan pasti memiliki kelebihan dan kekurangan, karena setiap pelanggan memiliki pengalaman yang berbeda selama menggunakan layanan. Opini pengguna terkait layanan biasanya disampaikan melalui media sosial, seperti Twitter. Berdasarkan penjabaran tersebut, maka penelitian ini dilakukan untuk menganalisis sentimen pengguna layanan ShopeeFood dengan metode *Naïve Bayes Classifier*. Proses pengumpulan data dilakukan dengan cara *crawling* berupa data opini publik pada media sosial Twitter atau disebut juga dengan *tweet*. *Tweet* yang diambil dalam penelitian ini adalah *tweet* yang menyertakan '@shopeefood_id' atau 'shopeefood' sebagai kata kunci, mulai tanggal 15 Juli 2022 hingga 22 Juli 2022, dan 16 September 2022 hingga 23 September 2022. *Dataset* yang terkumpul bersifat tidak seimbang, sehingga diterapkan metode SMOTE (*Synthetic Minority Oversampling Technique*). Hasilnya penerapan SMOTE + *Naïve Bayes* terbukti mampu menangani ketidakseimbangan kelas pada *dataset* yang digunakan dengan menghasilkan nilai akurasi yang lebih tinggi dibandingkan dengan *Naïve Bayes* saja. Hal tersebut membuktikan bahwa metode SMOTE efektif meningkatkan performa klasifikasi data tidak seimbang.

Kata kunci. Analisis Sentimen, *tweet*, *Naïve Bayes*, SMOTE.

ABSTRACT

NILATIL MOENA. Sentiment Analysis of ShopeeFood Service on Social Media Twitter Using the Naïve Bayes Classifier Algorithm. Thesis, Computer Science, Faculty of Mathematics and Natural Sciences, State University of Jakarta. January 2023.

Shopee started introducing a food delivery service called ShopeeFood in Indonesia since April 2020. Each service definitely has advantages and disadvantages, because each customer has a different experience while using the service. User opinions regarding services are usually conveyed through social media, such as Twitter. Based on this description, this research was conducted to analyze the sentiments of ShopeeFood service users using the Naïve Bayes Classifier algorithm. The data is collected through crawling public opinion data on Twitter. Tweets captured in this research are tweets that include '@shopeefood_id' or 'shopeefood' as keywords, from July 15, 2022 to July 22, 2022, and September 16, 2022 to September 23, 2022. The collected dataset is imbalanced, so the SMOTE (Synthetic Minority Oversampling Technique) method is applied to handle class imbalances. The result of the research shows that the SMOTE method, when combined with Naïve Bayes, produces higher accuracy compared to Naïve Bayes only. This proves that the SMOTE method is effective in improving the performance of unbalanced data classification.

Keywords. *Sentiment Analysis, tweet, Naïve Bayes, SMOTE.*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT, atas segala berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Layanan ShopeeFood pada Media Sosial Twitter Menggunakan Algoritma *Naïve Bayes Classifier*” tepat pada waktunya. Adapun tujuan dari penulisan skripsi ini dibuat untuk memenuhi persyaratan memperoleh gelar sarjana komputer.

Pada kesempatan ini, penulis hendak menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan sehingga skripsi ini dapat selesai. Ucapan terima kasih ini penulis tujukan kepada:

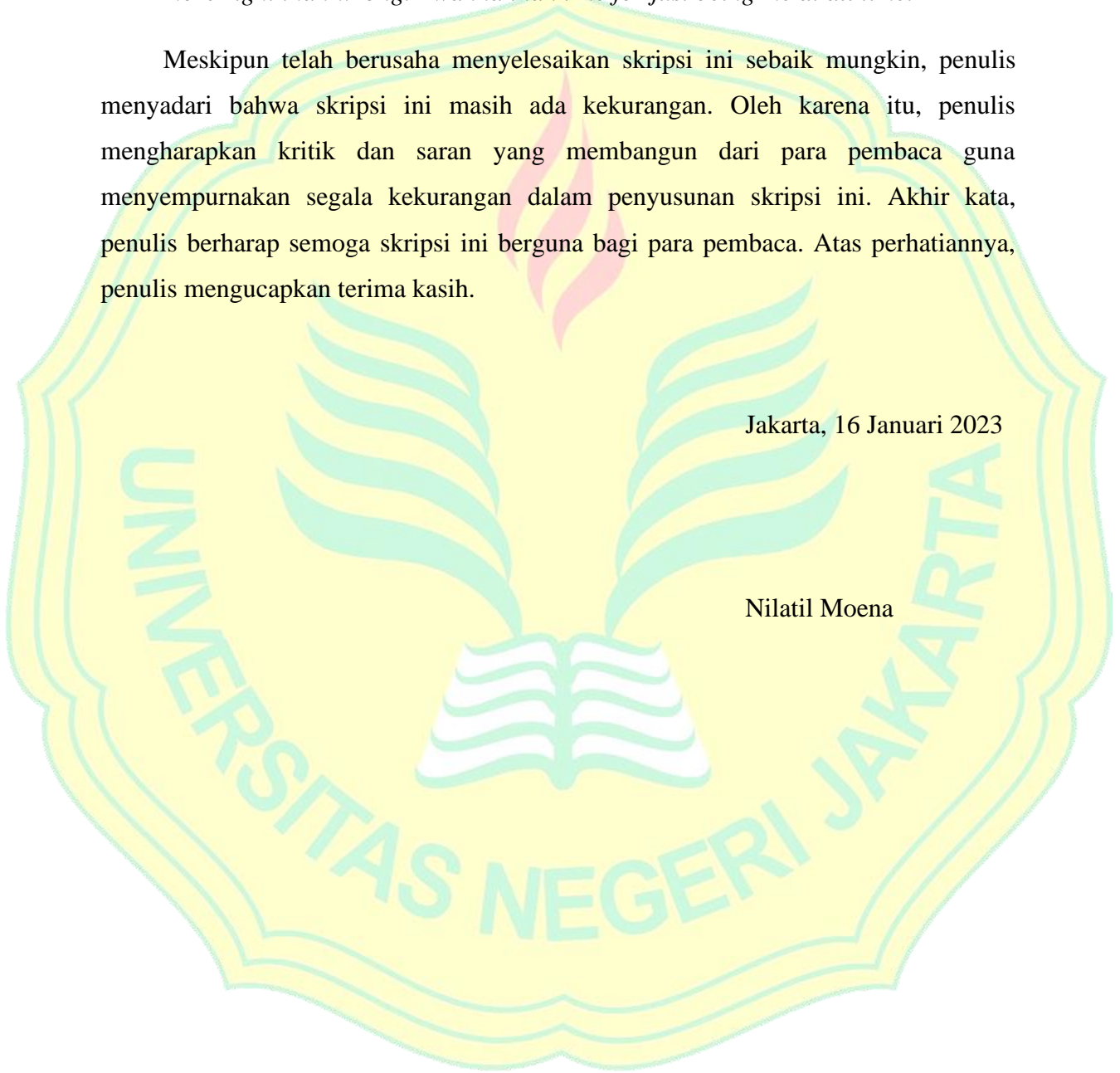
1. Kedua orang tua serta keluarga tercinta yang selalu memberikan doa, dorongan dan semangat selama penyusunan skripsi ini hingga akhir.
2. Ibu Dr. Ria Arafiyah, M.Si., selaku Dosen Pembimbing I, dan Bapak Drs. Mulyono, M.Kom., selaku Dosen Pembimbing II, yang telah rela meluangkan waktu untuk membimbing dan memberikan pengarahan kepada penulis.
3. Ibu Ir. Fariani Hermin Indiyah, M.T. selaku Koordinator Program Studi yang telah memberikan arahan dan informasi serta mendukung mahasiswa selama berkuliah di Ilmu Komputer, FMIPA UNJ.
4. Bapak/Ibu dosen Program Studi Ilmu Komputer, FMIPA UNJ atas ilmu-ilmu yang bermanfaat selama perkuliahan, serta seluruh jajaran pegawai di program studi yang telah membantu proses belajar mengajar.
5. Teman-teman seperjuangan di Ilmu Komputer yang telah kebersamai dan saling memberikan dorongan moral.
6. Sylviana, teman baik penulis yang selalu memberikan semangat dan bantuan selama penyusunan skripsi.
7. Semua pihak yang tidak bisa disebutkan satu per satu yang sudah terlibat untuk membantu penulis, baik dalam bentuk saran maupun dukungan selama proses pengerjaan skripsi ini sehingga dapat terselesaikan dengan baik.

8. *Last but not least, I wanna thank me. I wanna thank me for believing in me. I wanna thank me for doing all this hard work. I wanna thank me for having no days off. I wanna thank me for never quitting. I wanna thank me for trying to do more right than wrong. I wanna thank me for just being me at all time.*

Meskipun telah berusaha menyelesaikan skripsi ini sebaik mungkin, penulis menyadari bahwa skripsi ini masih ada kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari para pembaca guna menyempurnakan segala kekurangan dalam penyusunan skripsi ini. Akhir kata, penulis berharap semoga skripsi ini berguna bagi para pembaca. Atas perhatiannya, penulis mengucapkan terima kasih.

Jakarta, 16 Januari 2023

Nilatil Moena



DAFTAR ISI

LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN	iii
HALAMAN PERSEMBAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Perumusan Masalah	3
C. Tujuan Penelitian	3
D. Batasan Masalah	4
E. Manfaat Penelitian	4
BAB II KAJIAN PUSTAKA	5
A. <i>Text Mining</i>	5
B. Analisis Sentimen	6
C. Twitter	6
D. ShopeeFood	7
E. Python	7
F. <i>Crawling Data</i>	8
G. <i>Fleiss Kappa</i>	9
H. <i>Pre-Processing Data</i>	10
1. <i>Case Folding</i>	11
2. <i>Data Cleaning</i>	11
3. <i>Tokenization</i>	11

4.	<i>Normalization</i>	12
5.	<i>Stopwords Removal</i>	12
6.	<i>Stemming</i>	12
I.	Klasifikasi Menggunakan Algoritma <i>Naïve Bayes</i>	12
J.	<i>Synthetic Minority Oversampling Technique (SMOTE)</i>	17
K.	Evaluasi.....	17
BAB III METODOLOGI PENELITIAN		21
A.	Identifikasi Masalah.....	22
B.	Studi Literatur	22
C.	Pengumpulan Data	22
D.	<i>Filtering Data</i>	23
E.	Pelabelan Data	24
F.	<i>Pre-Processing Data</i>	25
1.	<i>Case Folding</i>	25
2.	<i>Data Cleaning</i>	26
3.	<i>Tokenization</i>	26
4.	<i>Normalization</i>	27
5.	<i>Stopwords Removal</i>	28
6.	<i>Stemming</i>	28
G.	Klasifikasi Menggunakan Algoritma <i>Naïve Bayes</i>	29
H.	Hasil dan Evaluasi	30
I.	Dokumentasi	31
J.	Perangkat Penelitian	31
K.	Jadwal Penelitian	32
BAB IV HASIL DAN PEMBAHASAN		33
A.	Pengumpulan Data	33
B.	<i>Filtering Data</i>	34
C.	Pelabelan Data	35
D.	<i>Pre-Processing Data</i>	40
1.	<i>Case Folding</i>	41
2.	<i>Data Cleaning</i>	42
3.	<i>Tokenization</i>	43
4.	<i>Normalization</i>	44

5. <i>Stopwords Removal</i>	45
6. <i>Stemming</i>	46
E. Pembobotan Kata.....	48
F. Klasifikasi Menggunakan Algoritma <i>Naïve Bayes</i>	51
1. Proses Pelatihan (<i>Training</i>)	51
2. Proses Pengujian (<i>Testing</i>).....	55
G. Hasil dan Evaluasi	58
BAB V KESIMPULAN DAN SARAN.....	67
A. Kesimpulan	67
B. Saran	68
DAFTAR PUSTAKA	69
LAMPIRAN.....	74
Lampiran 1 Data Penelitian.....	75
Lampiran 2 Kamus <i>Slang Word</i>	90
Lampiran 3 Surat Pernyataan <i>Labelling Data</i>	94
Lampiran 4 <i>Source Code Crawling Data</i>	97
Lampiran 5 <i>Source Code Pre-Processing Data</i>	98
Lampiran 6 <i>Source Code</i> Klasifikasi	101
Lampiran 7 <i>Source Code</i> SMOTE	105
RIWAYAT HIDUP.....	106

DAFTAR TABEL

Tabel 2.1 Interpretasi <i>Kappa</i>	10
Tabel 2.2 <i>Confusion Matrix</i>	18
Tabel 3.1 Contoh <i>Tweet</i> Sentimen Positif dan Negatif.....	24
Tabel 3.2 Contoh Proses <i>Case Folding</i>	26
Tabel 3.3 Contoh Proses <i>Data Cleaning</i>	26
Tabel 3.4 Contoh Proses <i>Tokenization</i>	27
Tabel 3.5 Contoh Proses <i>Normalization</i>	27
Tabel 3.6 Contoh Proses <i>Stopwords Removal</i>	28
Tabel 3.7 Contoh Proses <i>Stemming</i>	29
Tabel 3.8 Jadwal Penelitian.....	32
Tabel 4.1 Contoh <i>Tweet</i> pada Proses <i>Filtering</i>	34
Tabel 4.2 Contoh Pelabelan Data.....	35
Tabel 4.3 Sampel Data untuk <i>Kappa Value</i>	37
Tabel 4.4 Sampel Data Mentah.....	40
Tabel 4.5 Hasil Sebelum dan Sesudah Proses <i>Case Folding</i>	41
Tabel 4.6 Hasil Sebelum dan Sesudah Proses <i>Data Cleaning</i>	43
Tabel 4.7 Hasil Sebelum dan Sesudah Proses <i>Tokenization</i>	43
Tabel 4.8 Hasil Sebelum dan Sesudah Proses <i>Normalization</i>	45
Tabel 4.9 Hasil Sebelum dan Sesudah Proses <i>Stopwords Removal</i>	46
Tabel 4.10 Hasil Sebelum dan Sesudah Proses <i>Stemming</i>	47
Tabel 4.11 Sampel Data untuk Perhitungan <i>Bag of Words</i>	48
Tabel 4.12 Hasil Perhitungan <i>Bag of Words</i>	49
Tabel 4.13 Sampel Data Latih (<i>training</i>).....	51
Tabel 4.14 Nilai <i>Bag of Words</i> Data Latih (<i>training</i>).....	52
Tabel 4.15 Nilai Probabilitas Data Latih (<i>training</i>).....	54
Tabel 4.16 Sampel Data Uji (<i>testing</i>).....	55
Tabel 4.17 <i>Pre-processing</i> Data Uji (<i>testing</i>)	55
Tabel 4.18 Nilai <i>Bag of Words</i> Data Uji (<i>testing</i>).....	56
Tabel 4.19 Perbandingan Jumlah Data Latih (<i>training</i>)	62

DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian	21
Gambar 3.2 Data Hasil <i>Crawling</i>	23
Gambar 3.3 Data dalam Bentuk Tabel.....	23
Gambar 3.4 Proses <i>Case Folding</i>	25
Gambar 3.5 Proses <i>Data Cleaning</i>	26
Gambar 3.6 Proses <i>Tokenization</i>	27
Gambar 3.7 Proses <i>Normalization</i>	27
Gambar 3.8 Proses <i>Stopwords Removal</i>	28
Gambar 3.9 Proses <i>Stemming</i>	29
Gambar 3.10 Contoh Visualisasi <i>Word Cloud</i> (Sumber: Luvsandorj, 2020)	31
Gambar 4.1 <i>Code Crawling Data</i>	33
Gambar 4.2 Hasil <i>Crawling Data</i>	34
Gambar 4.3 Visualisasi Perbandingan Data.....	36
Gambar 4.4 <i>Code Read Dataset</i>	40
Gambar 4.5 <i>Code Case Folding</i>	41
Gambar 4.6 <i>Code Data Cleaning</i>	42
Gambar 4.7 <i>Code Tokenization</i>	44
Gambar 4.8 <i>Code Normalization</i>	44
Gambar 4.9 <i>Code Normalization</i>	46
Gambar 4.10 <i>Code Stemming</i>	47
Gambar 4.11 <i>Code Bag of Words</i>	50
Gambar 4.12 <i>Code Split Data</i>	51
Gambar 4.13 <i>Code Naive Bayes</i> untuk <i>Training</i>	54
Gambar 4.14 <i>Code Naive Bayes</i> untuk <i>Prediksi</i>	58
Gambar 4.15 <i>Code Confusion Matrix</i>	58
Gambar 4.16 <i>Confusion Matrix</i> Hasil Evaluasi	59
Gambar 4.17 <i>Confusion Matrix</i> Hasil Evaluasi dengan SMOTE	62
Gambar 4.18 Perbandingan Hasil Evaluasi.....	64
Gambar 4.19 <i>Word Cloud</i> Sentimen Positif Layanan ShopeeFood.....	65
Gambar 4.20 <i>Word Cloud</i> Sentimen Negatif Layanan ShopeeFood	65

DAFTAR LAMPIRAN

Lampiran 1 Data Penelitian.....	75
Lampiran 2 Kamus <i>Slang Word</i>	90
Lampiran 3 Surat Pernyataan <i>Labelling Data</i>	94
Lampiran 4 <i>Source Code Crawling Data</i>	97
Lampiran 5 <i>Source Code Pre-Processing Data</i>	98
Lampiran 6 <i>Source Code</i> Klasifikasi	101
Lampiran 7 <i>Source Code</i> SMOTE.....	105

