



SKRIPSI

**KOMPARASI ALGORITMA TOPIC MODELING
DALAM *MULTI-ASPECT* SENTIMENT
ANALYSIS ULASAN APLIKASI SEABANK
MENGUNAKAN RANDOM FOREST**

KEYSYA ALIFIA ZABINA

NPM 22082010187

DOSEN PEMBIMBING

Reisa Permatasari, S.T., M.Kom.

Tri Luhur Indayanti Sugata, S.ST., M.IM

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2026**

LEMBAR PENGESAHAN
**KOMPARASI ALGORITMA TOPIC MODELING DALAM MULTI-
ASPECT SENTIMENT ANALYSIS ULASAN APLIKASI SEABANK
MENGUNAKAN RANDOM FOREST**

Oleh :
KEYSYA ALIFIA ZABINA
NPM. 22082010187

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 28 Agustus 2026.

Reisa Permatasari, S.T., M.Kom
NIP. 19920514 2022032 007



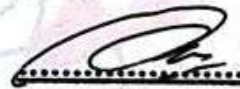
(Pembimbing I)

Tri Luhur Indayanti Sugata, S.ST., M.IM
NIP. 19920616 2024062 001



(Pembimbing II)

Nur Cahyo Wibowo, S.Kom., M.Kom.
NIP. 19790317 2021211 002



(Ketua Penguji)

Nambi Sembilu, S.Kom., M.Kom.
NIP. 19900516 2024061 003



(Anggota Penguji II)

Muhammad Reza Pahlawan, S.Kom, M.Kom.
NIP. 19980516 2025061 005



(Anggota Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**KOMPARASI ALGORITMA TOPIC MODELING DALAM *MULTI-ASPECT* SENTIMENT ANALYSIS ULASAN APLIKASI SEABANK
MENGUNAKAN RANDOM FOREST**

Oleh :

KEYSYA ALIFIA ZABINA

NPM. 22082010187

Menyetujui,

Koordinator Program Studi Sistem Informasi

Fakultas Ilmu Komputer



Siti Mukaromah, S.Kom., M.Kom.

NIP. 198107042021212011

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Keysya Alifia Zabina

NPM : 22082010187

Program : Sarjana (S1)

Program Studi : Sistem Informasi

Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, September 2026

Yang Membuat Pernyataan,



Keysya Alifia Zabina

NPM. 22082010187

ABSTRAK

Nama Mahasiswa / NPM : Keysya Alifia Zabina / 22082010187

Judul Skripsi : Komparasi Algoritma Topic Modeling Dalam Multi Aspect Sentiment Analysis Ulasan Aplikasi SeaBank Menggunakan Random Forest

Dosen Pembimbing : 1. Reisa Permatasari, S.T., M.Kom.
2. Tri Luhur Indayanti Sugata, S.ST., M.IM

Topik modeling merupakan salah satu pendekatan yang dapat digunakan untuk mengidentifikasi aspek atau topik yang dimiliki dari beberapa teks. Perkembangan metode topic modeling menghasilkan pendekatan tradisional berbasis statistik dan aljabar linear, seperti *Latent Dirichlet Allocation* (LDA) dan *Non-Negative Matrix Factorization* (NMF), serta pendekatan modern seperti BERTopic yang memanfaatkan *embedding* dan *transformer*. Penelitian ini bertujuan untuk melakukan komparasi LDA, NMF, dan BERTopic dalam mengidentifikasi aspek pada ulasan aplikasi SeaBank serta melakukan analisis sentimen pada setiap aspek yang diperoleh. Evaluasi topic modeling dilakukan menggunakan *coherence score* dan *topic diversity* untuk mengetahui kualitas dan keberagaman topik yang dihasilkan. Analisis sentimen dilakukan menggunakan algoritma *Random Forest* dengan representasi fitur TF-IDF dan Word2Vec. Untuk mengatasi permasalahan ketidakseimbangan kelas, penelitian ini membandingkan pendekatan *baseline*, SMOTETomek, dan *Class Weighted*. Hasil penelitian menunjukkan bahwa metode BERTopic menghasilkan kualitas topik terbaik berdasarkan nilai coherence score sebesar 0.691327 dan topic diversity sebesar 0.95. Pada tahap analisis sentimen, Skenario 5 dengan menggunakan TF-IDF dan SMOTETomek menghasilkan performa terbaik dengan nilai *Accuracy* sebesar 0.912 dan F1 Score sebesar 0.576. Hasil tersebut menunjukkan bahwa pendekatan topic modeling yang dipilih mampu menghasilkan aspek yang relevan dan dapat digunakan sebagai dasar dalam analisis sentimen multi-aspek pada ulasan aplikasi perbankan digital.

Kata kunci: topic modeling, multi-aspect sentiment analysis, LDA, NMF, BERTopic, Random Forest, SMOTETomek, Class Weighted.

ABSTRACT

Student Name / NPM : Keysya Alifia Zabina / 22082010187
Thesis Tittle : A Comparison of Topic Modeling Algorithms in
Multi-Aspect Sentiment Analysis of SeaBank App
Reviews Using Random Forest
Advisor : 1. Reisa Permatasari, S.T., M.Kom.
2. Tri Luhur Indayanti Sugata, S.ST., M.IM

Topic modeling is one approach that can be used to identify aspects or topics present in multiple texts. Advances in topic modeling methods have led to traditional approaches based on statistics and linear algebra, such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), as well as modern approaches like BERTopic, which utilize embeddings and transformers. This study aims to compare LDA, NMF, and BERTopic in identifying aspects in reviews of the SeaBank app and to perform sentiment analysis on each identified aspect. Topic modeling was evaluated using coherence scores and topic diversity to assess the quality and diversity of the generated topics. Sentiment analysis was performed using the Random Forest algorithm with TF-IDF and Word2Vec feature representations. To address the issue of class imbalance, this study compared the baseline, SMOTETomek, and Class Weighted approaches. The results show that the BERTopic method produced the best topic quality, with a coherence score of 0.691327 and a topic diversity score of 0.95. In the sentiment analysis stage, Scenario 5, which uses TF-IDF and SMOTETomek, delivered the best performance with an accuracy of 0.912 and an F1 score of 0.576. These results indicate that the selected topic modeling approach is capable of identifying relevant aspects and can serve as a foundation for multi-aspect sentiment analysis in reviews of digital banking apps.

Keywords: topic modeling, multi-aspect sentiment analysis, LDA, NMF, BERTopic, Random Forest, SMOTETomek, Class Weighted.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul *"Komparasi Algoritma Topic Modeling Dalam Multi-Aspect Sentiment Analysis Ulasan Aplikasi SeaBank Menggunakan Random Forest"*. Penulisan skripsi ini merupakan tahapan akhir yang harus ditempuh untuk memperoleh gelar Sarjana Ilmu Komputer pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Penyusunan skripsi ini memerlukan waktu, tenaga, dan komitmen, serta mendapat dukungan dari berbagai pihak. Untuk itu, penulis menyampaikan terima kasih kepada:

1. Almarhumah Mama tercinta, yang telah berpulang jauh sebelum penulis memulai perjalanan perkuliahan. Meskipun Mama tidak dapat melihat secara langsung perjalanan penulis hingga menyelesaikan pendidikan ini, kasih sayang, pengorbanan, dan kenangan yang ditinggalkan selalu menjadi kekuatan bagi penulis untuk terus melangkah. Skripsi ini penulis persembahkan untuk Mama sebagai bentuk cinta dan penghormatan, serta sebagai bukti bahwa segala doa dan harapan yang pernah diberikan akan selalu penulis bawa dalam setiap langkah kehidupan.
2. Ayah tercinta, yang selalu mendampingi penulis, memberikan support, doa dan dukungan yang besar bagi penulis selama menjalani masa perkuliahan hingga menyelesaikan skripsi ini. Terima kasih atas pengorbanan dan perjuangan yang telah diberikan kepada penulis, serta menjadi tempat pulang untuk menguatkan diri dalam setiap proses yang dilalui.
3. Ibu Reisa Permatasari S.T., M.Kom, selaku dosen pembimbing pertama dan Ibu Tri Luhur Indayanti Sugata, S.SI., M.IM, selaku dosen pembimbing kedua yang telah meluangkan waktu, memberikan bimbingan, arahan, masukan, serta saran kepada penulis selama proses penyusunan hingga penyelesaian skripsi ini. Terima kasih atas kesabaran dan dukungan yang diberikan selama proses penelitian, sehingga penulis dapat menyelesaikan skripsi ini dengan sebaik-baiknya.

4. Seluruh Bapak dan Ibu Dosen Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang telah memberikan ilmu, pengetahuan, pengalaman, serta pembelajaran selama penulis menempuh pendidikan hingga dapat menyelesaikan skripsi ini.
5. Keluarga besar, terutama Tante Nur dan Sepupu penulis, Yuseva yang selalu memberikan dukungan, perhatian, serta menjadi tempat bagi penulis untuk berbagi cerita setiap kali merasa *burnout* selama proses perkuliahan dan penyusunan skripsi. Terima kasih telah selalu mendengarkan, memberikan semangat, dan menemani penulis dalam melewati berbagai proses hingga skripsi ini dapat diselesaikan.
6. Teman seperkuliahan penulis D. Febry dan Vera, yang telah menemani penulis selama menjalani masa perkuliahan hingga proses penyusunan skripsi. Terima kasih telah menjadi teman untuk berbagi cerita, bertukar pikiran, mengerjakan berbagai hal bersama, serta saling memberikan semangat di tengah kesibukan dan berbagai tantangan selama perkuliahan. Terima kasih telah menjadi bagian dari perjalanan penulis hingga dapat menyelesaikan pendidikan ini.
7. Teman-teman masa kecil penulis, khususnya jajaran Kopi Takbir, yang telah bersedia menjadi tempat untuk berbagi keluh kesah, cerita, dan berbagai hal selama proses penyusunan skripsi ini. Terima kasih atas dukungan, semangat, dan kebersamaan yang selalu diberikan kepada penulis.
8. Diri sendiri, yang telah bertahan dan terus berusaha dalam menyelesaikan skripsi ini. Terima kasih telah tetap melangkah meskipun menghadapi berbagai tantangan, rasa lelah, dan keraguan selama proses penyusunan skripsi. Terima kasih telah memilih untuk terus berusaha hingga akhirnya dapat menyelesaikan pendidikan ini.
9. Pihak-pihak lain yang telah membantu penulis dalam menyelesaikan penelitian ini, khususnya annotator yang telah meluangkan waktu dan memberikan bantuan dalam proses pelabelan data, sehingga penelitian ini dapat diselesaikan dengan baik.

Surabaya, 28 Agustus 2026

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN.....	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI	vii
ABSTRAK.....	ix
ABSTRACT	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI	xv
DAFTAR TABEL	xviii
DAFTAR GAMBAR.....	xxi
DAFTAR LAMPIRAN	xxv
DAFTAR SINGKATAN, ISTILAH DAN SIMBOL	xxvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	7
1.3 Batasan Masalah	7
1.4 Tujuan Penelitian	8
1.5 Manfaat Penelitian	8
1.6 Sistematika Penulisan.....	8
BAB II TINJAUAN PUSTAKA	11
2.1 Dasar Teori	11
2.1.1 Bank Digital	11
2.1.2 SeaBank	12
2.1.3 Ulasan Pengguna	13
2.1.4 Text Mining.....	13
2.1.5 Multi-Aspect Sentiment Analysis	15
2.1.6 Latent Dirichlet Allocation	15
2.1.7 Non-Negative Matrix Factorization (NMF).....	16
2.1.8 BERTopic	17
2.1.9 Evaluasi Topic Modeling.....	18

2.1.10 Fleiss Kappa	20
2.1.11 Random Forest	21
2.1.12 Pembagian Data <i>Holdout</i>	22
2.1.13 Term Frequency–Inverse Document Frequency (TF-IDF).....	22
2.1.14 Word2Vec	23
2.1.15 Teknik Imbalance Data.....	24
2.1.16 Evaluasi Model.....	26
2.1.17 Hamming Loss	28
2.1.18 <i>Flask</i>	28
2.2 Penelitian Terdahulu	29
BAB III METODOLOGI PENELITIAN	37
3.1 Studi Literatur	37
3.2 Analisis Kebutuhan	38
3.2.1 Kebutuhan Data.....	38
3.2.2 Kebutuhan Software dan Hardware.....	38
3.3 Pengumpulan Data	39
3.4 Explorasi Data.....	41
3.5 Pre-Processing Data	41
3.5.1 Data Cleaning.....	42
3.5.2 Filtering Data	42
3.5.3 Case Folding	43
3.5.4 Normalization.....	43
3.5.5 Tokenization.....	43
3.5.6 Stopwords Removal.....	43
3.5.7 Stemming	44
3.6 Topic Modeling.....	44
3.7 Evaluasi Hasil Topic Modeling	45
3.8 Pelabelan Data	46
3.9 Pembagian Data	46
3.10 Perancangan Model Klasifikasi	47
3.10.1 Pembagian Holdout	47
3.10.2 Representasi Fitur Teks	48

3.10.3 Skenario Model	49
3.10.4 Evaluasi Model.....	52
3.11 Penyimpanan Model.....	52
3.12 Deployment Model.....	52
3.13 Validasi Model.....	54
BAB IV HASIL DAN PEMBAHASAN.....	55
4.1 Hasil	55
4.1.1 Studi Literature.....	55
4.1.2 Analisis Kebutuhan	55
4.1.3 Pengumpulan Data.....	56
4.1.4 Eskplorasi Data	59
4.1.5 Pre-Processing Data.....	61
4.1.5 Topic Modeling	77
4.1.6 Evaluasi Topic Modeling.....	100
4.1.7 Pelabelan Data.....	108
4.1.8 Pembagian Data.....	116
4.1.9 Perancangan Model Klasifikasi.....	118
4.1.10 Evaluasi Model Terbaik.....	139
4.1.11 Penyimpanan Model.....	145
4.1.12 Deployment Model.....	147
4.1.13 Validasi Model	150
4.2 Pembahasan	153
BAB V KESIMPULAN DAN SARAN.....	165
5.1 KESIMPULAN	165
5.2 SARAN.....	166
DAFTAR PUSTAKA.....	169
LAMPIRAN	175

DAFTAR TABEL

Tabel 2.1 Confusion Matrix [53]	26
Tabel 2.2 Penelitian Terdahulu	29
Tabel 3.1 Contoh data <i>Head</i>	40
Tabel 3.2 Contoh data <i>Tail</i>	40
Tabel 3.3 Skenario Model	49
Tabel 3.4 Parameter Random Forest	51
Tabel 4.1 Hasil <i>Cleaning</i> Data	62
Tabel 4.2 Contoh Ulasan Tidak Relevan	64
Tabel 4.3 Jumlah Data Hasil <i>Filtering</i>	64
Tabel 4.4 Hasil <i>Case Folding</i>	65
Tabel 4.5 Hasil Normalisasi	68
Tabel 4.6 Hasil <i>Tokenization</i>	69
Tabel 4.7 Frekuensi Kata Sering Muncul	72
Tabel 4.8 Hasil Stopwords Removal	73
Tabel 4.9 Hasil Data Kosong Sesudah <i>Stopword Removal</i>	75
Tabel 4.10 Total Data Setelah <i>Stopwords Removal</i>	75
Tabel 4.11 Hasil <i>Stemming</i>	76
Tabel 4.12 Hasil Pengujian <i>Coherence</i> LDA	79
Tabel 4.13 Hasil Interpretasi Topik Menggunakan LDA	80
Tabel 4.14 Hasil <i>Coherence</i> NMF	87
Tabel 4.15 Hasil Interpretasi Topik Menggunakan NMF	88
Tabel 4.16 Hasil <i>Coherence</i> BERTopic	96
Tabel 4.17 Hasil Interpretasi Topik Menggunakan BERTopic	97
Tabel 4.18 Hasil Topik Terbaik Menggunakan LDA	102
Tabel 4.19 Hasil Topik Terbaik Menggunakan NMF	104
Tabel 4.20 Hasil Topik Terbaik Menggunakan BERTopic	105
Tabel 4.21 Hasil Topik Terbaik	107
Tabel 4.22 Hasil Interpretasi Topik Dipilih	107
Tabel 4.23 Definisi Aspek	108
Tabel 4.24 Panduan Labeling	109

Tabel 4.25 Profile Annotator.....	110
Tabel 4.26 Hasil Fleiss' Kappa	112
Tabel 4.27 Hasil Pembagian Data	118
Tabel 4.28 Perancangan Skenario Model.....	118
Tabel 4.29 Tabel Skenario 1.....	120
Tabel 4.30 Tabel Skenario 2.....	121
Tabel 4.31 Tabel Skenario 3.....	122
Tabel 4.32 Tabel Skenario 4.....	123
Tabel 4.33 Tabel Skenario 5.....	124
Tabel 4.34 Tabel Skenario 6.....	125
Tabel 4.35 Tabel Skenario 7.....	126
Tabel 4.36 Tabel Skenario 8.....	128
Tabel 4.37 Tabel Skenario 10.....	130
Tabel 4.38 Tabel Skenario 11.....	131
Tabel 4.39 Tabel Skenario 12.....	132
Tabel 4.40 Tabel Skenario 13.....	133
Tabel 4.41 Tabel Skenario 14.....	134
Tabel 4.42 Tabel Skenario 15.....	135
Tabel 4.43 Tabel Skenario 16.....	136
Tabel 4.44 Tabel Skenario 17.....	138
Tabel 4.45 Tabel Skenario 18.....	139
Tabel 4.46 Hasil Rekapitulasi Skenario	142
Tabel 4.47 Perbandingan Skenario Terbaik	144
Tabel 4.48 Hasil Topik Terbaik.....	154
Tabel 4.49 Hasil Interpretasi Topik Dipilih.....	154
Tabel 4.50 Hasil Penerapan SMOTETomek.....	156
Tabel 4.51 Skenario Terbaik	157

DAFTAR GAMBAR

Gambar 1.1 Perkiraan Pengguna Aktif 2025 Bank Digital di Indonesia	2
Gambar 1.2 Rating Aplikasi SeaBank di Google Play Store dan App Store	3
Gambar 2.1 Halaman Utama Aplikasi SeaBank	12
Gambar 2.2 Ilustrasi Model LDA [30]	16
Gambar 2.3 Arsitektur <i>Skip-Gram</i> pada <i>Word2Vec</i> [47]	24
Gambar 3.1 Diagram Alur Penelitian	37
Gambar 3.2 Tahap <i>Pre-Processing</i>	42
Gambar 3.3 Perancangan Model	47
Gambar 3.4 <i>Wireframe</i> Single Input Ulasan	53
Gambar 3.5 <i>Wireframe</i> Upload Ulasan CSV	53
Gambar 4.1 Kode Scraping Ulasan SeaBank di APP Store	57
Gambar 4.2 Hasil Scraping Ulasan SeaBank di APP Store	57
Gambar 4.3 Kode Scraping Ulasan SeaBank di Google Play Store	57
Gambar 4.4 Hasil Scraping Ulasan SeaBank di Google Play Store	58
Gambar 4.5 Hasil Penggabungan Data	58
Gambar 4.6 Jumlah Masing-Masing Data	59
Gambar 4.7 Distribusi Jumlah dan Rating Ulasan	59
Gambar 4.8 Wordcloud Data Ulasan	59
Gambar 4.9 Distribusi Jumlah Kata per Ulasan	60
Gambar 4.10 Missing Value	61
Gambar 4.11 Kode <i>Cleaning</i> Data	62
Gambar 4.12 Kode Penghapusan >2 Kata	63
Gambar 4.13 Kode Implementasi <i>Case Folding</i>	65
Gambar 4.14 Kode Implementasi <i>Normalization</i>	67
Gambar 4.15 Kode implementasi <i>Tokenization</i>	69
Gambar 4.16 Kode Implementasi <i>Stopwords Removal</i>	72
Gambar 4.17 Kode Implementasi <i>Empty Row</i>	74
Gambar 4.18 Kode Implementasi <i>Stemming</i>	76
Gambar 4.19 Kode Implementasi LDA	78
Gambar 4.20 Kode Implementasi <i>Coherence</i> LDA	78

Gambar 4.21 Grafik Coherence LDA	79
Gambar 4.22 Kode Implementasi NMF	85
Gambar 4.23 Kode Implementasi <i>Coherence</i> NMF	86
Gambar 4.24 Grafik Coherence NMF	87
Gambar 4.25 Kode Implementasi BERTopic.....	94
Gambar 4.26 Kode Implementasi <i>Coherence</i> BERTopic.....	95
Gambar 4.27 Grafik <i>Coherence</i> BERTopic	96
Gambar 4.28 Grafik <i>Coherence Score</i> LDA, NMF, BERTopic	100
Gambar 4.29 Grafik <i>Topic Diversity</i> LDA, NMF, BERTopic	101
Gambar 4.30 Hasil PyLDAVIS untuk 9 Topik LDA	101
Gambar 4.31 Hasil Intertopic Distance map untuk 9 Topik NMF	103
Gambar 4.32 Hasil Intertopic Distance untuk 7 Topik BERTopic.....	105
Gambar 4.33 Hasil Pelabelan Manual Aspek dan Sentimen oleh Annotator 1...	111
Gambar 4.34 Kode Implementasi Fleiss' Kappa	112
Gambar 4.35 Majority Voting Pelabelan Manual Aspek dan Sentimen.....	114
Gambar 4.36 Hasil Distribusi Label per-Aspek	115
Gambar 4.37 Kode Pembagian Data.....	117
Gambar 4.38 Grafik Performa Seluruh Skenario	141
Gambar 4.39 Kode Implementasi Penyimpanan Data.....	147
Gambar 4.40 Kode Implementasi Deployment Model	148
Gambar 4.41 Gambar Tampilan Analisis Teks Ulasan	149
Gambar 4.42 Halaman Analisis File CSV	149
Gambar 4.43 Halaman Input Ulasan baru.....	150
Gambar 4.44 Halaman Hasil Prediksi Sistem Input Ulasan Baru	151
Gambar 4.45 Halaman Hasil Upload Sistem Input Analisis CSV	151
Gambar 4.46 Halaman Hasil Prediksi Sistem Input Ulasan CSV	152
Gambar 4.47 Halaman Hasil Prediksi Sentimen Ulasan.....	152
Gambar 4.48 Halaman Hasil Insight dari Data CSV	153
Gambar 4.49 Halaman Hasil Prediksi Seluruh Ulasan CSV	153
Gambar 4.50 Distribusi Kelas	155
Gambar 4.51 Confussion Matrix & Classification Report Aspek 1	159
Gambar 4.52 Confussion Matrix & Classification Report Aspek 2	160

Gambar 4.53 Confussion Matrix & Classification Report Aspek 3	160
Gambar 4.54 Confussion Matrix & Classification Report Aspek 4	161
Gambar 4.55 Confussion Matrix & Classification Report Aspek 5	162
Gambar 4.56 Confussion Matrix & Classification Report Aspek 6	162

DAFTAR LAMPIRAN

Lampiran 1. Hasil scraping data ulasan di Google Play Store	175
Lampiran 2. Hasil scraping data ulasan di Apple Store	175
Lampiran 3. Surat Izin Penelitian	176
Lampiran 4. Permohonan Izin Penelitian Kepada Pihak SeaBank	176

DAFTAR SINGKATAN, ISTILAH DAN SIMBOL

Notasi	Keterangan
D	Kumpulan dokumen (corpus)
d	Sebuah dokumen/ulasan
N	Jumlah kata dalam satu dokumen
w_n	Kata ke- n dalam dokumen
z_n	Topik laten untuk kata ke- n
k	Jumlah topik pada model LDA
α	Parameter Dirichlet untuk distribusi topik dokumen
β	Parameter distribusi kata pada setiap topik
θ_d	Distribusi topik pada dokumen d
$TF_{(t,d)}$	<i>term frequency</i>
$IDF_{(t)}$	<i>Inverse Document Frequency</i>
d	jumlah dimensi <i>neuron</i>
t	ukuran jendela kata
B	Jumlah pohon keputusan (<i>number of tree</i>) dalam <i>Random Forest</i>
b	Indeks keputusan pohon ke- b , dengan $b = 1, 2, \dots, B$
Z^*	Data bootstrap yang diambil acak dengan pengembalian data pelatihan
N	Jumlah dataset pada pelatihan
T_b	Pohon Keputusan ke- b dari bangun bootstrap
p	Jumlah total fitur atau variable
m	Jumlah fitur yang dipilih secara acak dari total p fitur setiap node
m_{min}	Jumlah minimum data pada node sebagai kriteria penghentian pembentukan pohon
x	Data atau sampel input yang akan diprediksi
$C_b(x)$	Hasil prediksi kelas dari pohon keputusan ke- b terhadap data x
$C_{rf}^B(x)$	Hasil prediksi akhir Random Forest untuk klasifikasi berdasarkan voting mayoritas
$f_{rf}^B(x)$	Hasil prediksi akhir Random Forest untuk klasifikasi berdasarkan voting mayoritas