

TUGAS AKHIR

ANALISIS PERBANDINGAN REPRESENTASI FITUR TF-IDF, *FASTTEXT*, DAN BERT PADA MODEL *SUPPORT VECTOR* *MACHINE* UNTUK KLASIFIKASI SENTIMEN KOMENTAR INSTAGRAM FOLKATIVE



RYSTIANA LAURA DIAN PRATIWI
2210651017

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2026**

TUGAS AKHIR

ANALISIS PERBANDINGAN REPRESENTASI FITUR TF-IDF, *FASTTEXT*, DAN BERT PADA MODEL *SUPPORT VECTOR MACHINE* UNTUK KLASIFIKASI SENTIMEN KOMENTAR INSTAGRAM FOLKATIVE

Diajukan sebagai salah satu syarat untuk kelulusan Strata Satu (S1) Program Studi
Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember



RYSTIANA LAURA DIAN PRATIWI
2210651017

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2026**

HALAMAN PERSETUJUAN

**ANALISIS PERBANDINGAN REPRESENTASI FITUR TF-IDF,
FASTTEXT, DAN BERT PADA MODEL *SUPPORT VECTOR MACHINE*
UNTUK KLASIFIKASI SENTIMEN KOMENTAR INSTAGRAM
FOLKATIVE**

Oleh:

Rystiana Laura Dian Pratiwi

2210651017

Telah disetujui Laporan Tugas Akhir ini untuk diajukan pada sidang Tugas Akhir
sebagai salah satu syarat kelulusan dan mendapatkan gelar

Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh,

Pembimbing I

Moh. Dasuki., M.Kom.
NIDN. 0722109103

Pembimbing II

Dr. Reni Umilasari, S.Pd., M.Si.
NIDN. 0728079101

HALAMAN PENGESAHAN

**ANALISIS PERBANDINGAN REPRESENTASI FITUR TF-IDF,
FASTTEXT, DAN BERT PADA MODEL *SUPPORT VECTOR MACHINE*
UNTUK KLASIFIKASI SENTIMEN KOMENTAR INSTAGRAM
FOLKATIVE**

Oleh:

Rystiana Laura Dian Pratiwi

2210651017

Telah mempertanggung jawabkan Laporan Tugas Akhir pada sidang Tugas Akhir
tanggal 24 Juli 2026 sebagai salah satu syarat kelulusan dan mendapatkan gelar
Sarjana Komputer (S.Kom) di Universitas Muhamadiyah Jember
Disetujui oleh,

**Dosen Penguji:
Penguji I**

Miftahur Rahman, S.Kom., M.Kom.
NIDN. 0724039201

**Dosen Pembimbing:
Pembimbing I**

Moh. Dasuki., M.Kom.
NIDN. 0722109103

Dosen Penguji II

Darvanto, S.Kom., M.Kom.
NIDN. 0707077203

Dosen Pembimbing II

Dr. Reni Umilasari, S.Pd., M.Si.
NIDN. 0728079101

Mengesahkan,
Dean Fakultas Teknik

Prof. Dr. I. Muhtar, S.T., MT., IPM.
NIDN. 0010067301

Mengetahui,
Ketua Program Studi
Teknik Informatika

Rosita Samudra, S.Kom., M.Cs.
NIDN. 0629018601

HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini:

Nama : Rystiana Laura Dian Pratiwi

NIM : 2210651017

Program Studi : Teknik Informatika

Judul Tugas Akhir : Analisis Perbandingan Representasi Fitur TF-IDF, *FastText*, dan BERT pada Model *Support Vector Machine* Untuk Klasifikasi Sentimen Komentar Instagram Folkative

Dengan ini menyatakan bahwa karya ilmiah yang saya susun adalah murni hasil pemikiran, penelitian, dan penyusunan saya sendiri. Segala sumber informasi yang berasal atau dikutip dari karya orang lain, baik berupa publikasi, artikel, buku, maupun sumber lainnya, telah disebutkan dengan jelas di dalam daftar pustaka sesuai dengan kaidah penulisan ilmiah.

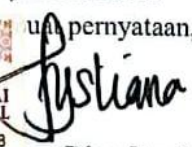
Saya menyadari bahwa dalam proses penyusunan Tugas Akhir ini, saya memanfaatkan bantuan teknologi kecerdasan buatan (AI) seperti *ChatGPT/Grammarly/Alat AI lain* hanya sebatas untuk:

1. Membantu dalam penyusunan tata bahasa, parafrasa, atau perbaikan redaksi.
2. Memberikan inspirasi ide awal atau kerangka berpikir, yang selanjutnya saya kembangkan secara mandiri.
3. Membantu dalam pengecekan konsistensi format, ejaan, dan tata tulis.

Saya menegaskan bahwa tidak ada bagian dari karya ilmiah ini yang seluruhnya dibuat oleh *AI* tanpa keterlibatan pemikiran kritis saya sendiri. Tanggung jawab penuh atas isi, keaslian, dan kebenaran karya ilmiah ini ada pada diri saya.

Apabila di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima segala konsekuensi sesuai dengan peraturan yang berlaku di Universitas Muhammadiyah Jember.

Jember, 24 Juli 2026

dan pernyataan,

Rystiana
Laura Dian Pratiwi
NIM.2210651017



MOTTO

وَأَنْ لَّيْسَ لِلْإِنْسَانِ إِلَّا مَا سَعَى

“Dan bahwa manusia hanya memperoleh apa yang telah diusahakannya.”

(QS. An-Najm: 39)

“Libatkan Allah SWT. disetiap perjalanan dan proses yang ditempuh. Hadapi kesulitan dengan doa, terima keberhasilan dengan sujud syukur”

(Penulis)

“Today is gonna be the day that they're gonna throw it back to you.”

(Oasis-Wonderwall)



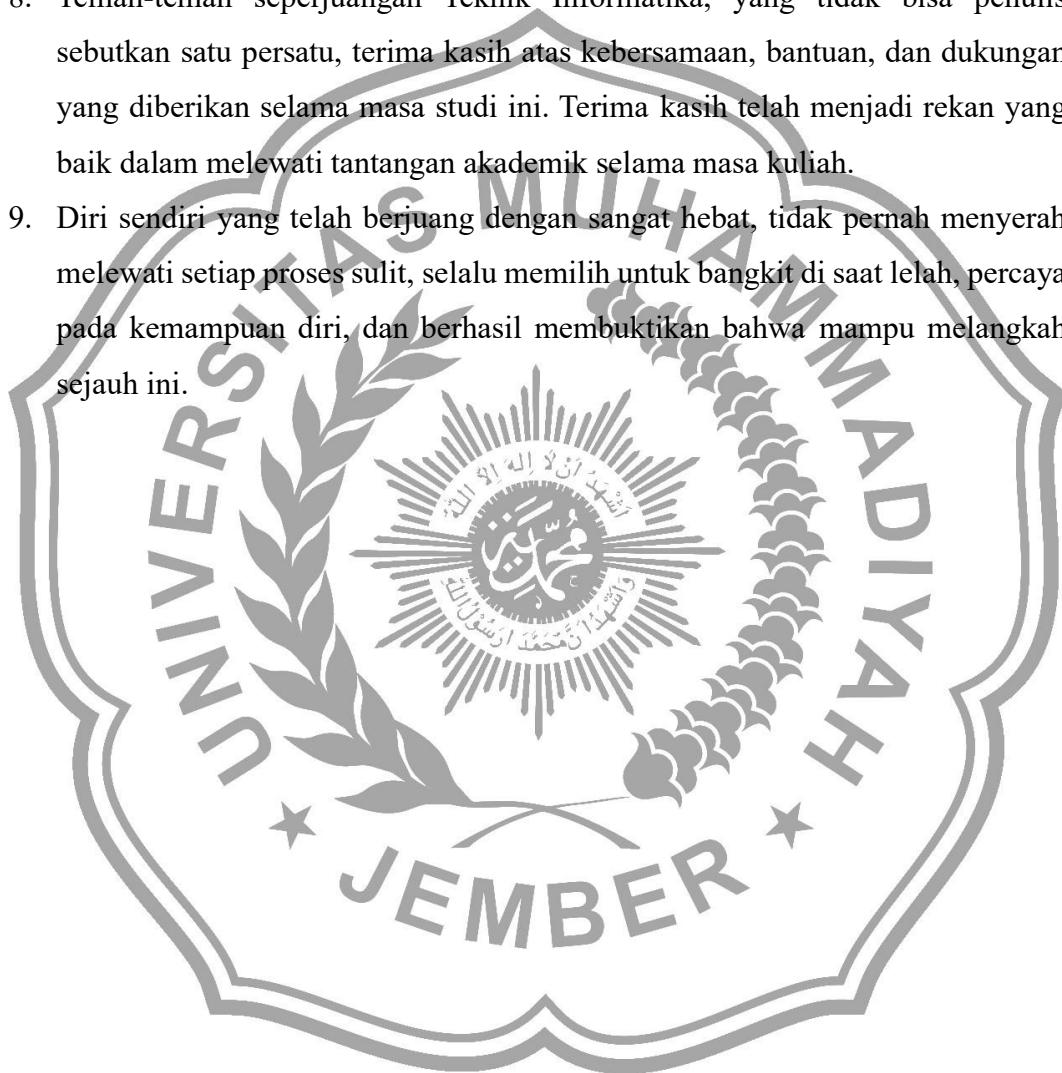
HALAMAN PERSEMBAHAN

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT. yang senantiasa melimpahkan rahmat, ketabahan, serta ilmu kepada penulis selama menyelesaikan tugas akhir ini. Segala tantangan dan rintangan yang menyertai perjalanan ini akhirnya dapat dilalui berkat kasih sayang dan kehendak-Nya. Sebagai wujud terima kasih atas kasih sayang dan dukungan tiada henti yang telah diberikan, dengan penuh ketulusan hati penulis mempersembahkan karya ini kepada:

1. Allah SWT, atas segala rahmat, hidayah, serta kekuatan yang selalu menyertai perjalanan penulis dalam menyelesaikan tugas akhir ini.
2. Ayah dan Ibu tercinta, terima kasih yang tak terhingga atas kasih sayang, pengorbanan, dan segala upaya yang diberikan agar penulis bisa menempuh pendidikan setinggi ini. Terima kasih sudah menjadi rumah yang selalu menenangkan, tempat penulis pulang saat lelah. Segala tetes keringat, doa-doa yang tulus, serta dukungan tanpa syarat yang ayah dan ibu berikan adalah kekuatan terbesar penulis hingga berdiri sampai titik ini. Karya ini penulis persembahkan sebagai tanda terima kasih atas cinta dan dukungan yang tak pernah putus sejak penulis lahir hingga saat ini.
3. Saudara tersayang, Krisna, Sabrina, dan kakak sepupu Karunia, yang telah menjadi teman tumbuh, pendengar yang sabar, dan penyemangat utama bagi penulis. Kehadiran kalian sebagai tempat berbagi tawa dan keluh kesah, mewarnai perjalanan penulis di masa-masa sulit.
4. Bapak Moh. Dasuki, M.Kom., selaku Dosen Pembimbing I, Ibu Dr. Reni Umilasari, S.Pd., M.Si., selaku Dosen Pembimbing II, yang dengan penuh kesabaran, memberikan banyak waktu, pengetahuan dan motivasi kepada penulis dalam menyelesaikan penelitian ini.
5. Bapak Miftahur Rahman, S.Kom., M.Kom selaku Dosen Penguji I, dan Bapak Daryanto, S.Kom., M.Kom., selaku Dosen Penguji II, atas saran konstruktif dan perspektif baru yang sangat membantu dalam perbaikan penelitian ini.
6. Seluruh dosen dan staf karyawan Universitas Muhammadiyah Jember, khususnya Program Studi Teknik Informatika, atas dedikasi dalam berbagi ilmu serta pelayanan yang luar biasa. Terima kasih telah menjadi bagian dari

perjalanan akademik penulis dan menyediakan lingkungan belajar yang sangat mendukung selama masa studi ini.

7. Teman-teman terkasih, Anis, Dilla, Sisil, Vania, Novita, Okta, Tasya, Indah, Putri, Elya, dan Arina, yang telah menjadi penyemangat dan selalu hadir untuk membantu dalam segala kesulitan. Selalu mendengarkan serta kebersamaan penulis di setiap proses ini dengan tawa.
8. Teman-teman seperjuangan Teknik Informatika, yang tidak bisa penulis sebutkan satu persatu, terima kasih atas kebersamaan, bantuan, dan dukungan yang diberikan selama masa studi ini. Terima kasih telah menjadi rekan yang baik dalam melewati tantangan akademik selama masa kuliah.
9. Diri sendiri yang telah berjuang dengan sangat hebat, tidak pernah menyerah melewati setiap proses sulit, selalu memilih untuk bangkit di saat lelah, percaya pada kemampuan diri, dan berhasil membuktikan bahwa mampu melangkah sejauh ini.



KATA PENGANTAR

Puji Syukur penulis panjatkan ke hadirat Allah SWT, karena atas rahmat, hidayah, dan karunia-Nya, penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir yang berjudul “**Analisis Perbandingan Representasi Fitur TF-IDF, FastText, Dan BERT Pada Model Support Vector Machine Untuk Klasifikasi Sentimen Instagram Folkative**” ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember. Penulis menyadari bahwa penyusunan Tugas Akhir ini tidak dapat berjalan lancar tanpa adanya bantuan, bimbingan, petunjuk, serta dukungan materiil maupun moral dari berbagai pihak selama masa perkuliahan hingga selesainya penelitian ini.

Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Moh. Dasuki, M.Kom., selaku Dosen Pembimbing I, dan Ibu Dr. Reni Umilasari, S.Pd, M.Si., selaku Dosen Pembimbing II, yang telah banyak memberikan waktu, pengetahuan, dan motivasi kepada penulis selama proses bimbingan berlangsung. Ucapan terima kasih juga penulis sampaikan kepada Bapak Miftahur Rahman, S.Kom., M.Kom., selaku Dosen Penguji I, dan Bapak Daryanto, S.Kom., M.Kom., selaku Dosen Penguji II, atas saran dan kritik konstruktif yang sangat membantu dalam perbaikan laporan ini.

Penulis juga menyampaikan apresiasi kepada seluruh dosen dan staf di lingkungan Fakultas Teknik Universitas Muhammadiyah Jember atas segala dukungan yang diberikan selama masa studi. Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna, baik dalam penulisan maupun isi yang disampaikan. Oleh karena itu, saran dan masukan yang membangun sangat diharapkan demi penyempurnaan di masa mendatang. Semoga karya ini dapat memberikan manfaat dan menjadi referensi yang berguna bagi pembaca serta pengembangan ilmu di bidang informatika.

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN ORISINALITAS.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
MOTTO.....	vii
HALAMAN PERSEMBAHAN.....	viii
KATA PENGANTAR	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	4
BAB 2 TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu	6
2.2 Landasan Teori.....	9
2.2.1 Analisis Sentimen.....	9
2.2.2 <i>Text Mining</i>	10
2.2.3 Media Sosial Instagram.....	10
2.2.4 Akun Instagram Folkative.....	11
2.2.5 <i>Python</i>	11
2.2.6 <i>Google Colab</i>	12
2.2.7 <i>Pre-processing</i>	12
2.2.8 <i>Word Embedding</i>	14
2.2.9 TF-IDF	14
2.2.10 <i>Bidirectional Encoder Representations from Transformers</i>	15
2.2.11 <i>FastText</i>	16

2.2.12	<i>Support Vector Machine</i>	17
2.2.13	<i>K-Fold Cross Validation</i>	19
2.2.14	<i>Confusion Matrix</i>	20
BAB 3	METODOLOGI PENELITIAN	23
3.1	Tahapan Penelitian	23
3.2	Pengumpulan Data	24
3.3	Pengolahan Data Awal	24
3.3.1	<i>Pre-Processing</i>	24
3.3.2	Pelabelan Data	28
3.4	Representasi Fitur	29
3.4.1	Pembobotan Kata menggunakan TF-IDF	29
3.4.2	<i>FastText</i>	30
3.4.3	BERT (<i>Bidirectional Encoder Representations from Transformers</i>)	32
3.4.4	Klasifikasi model <i>Support Vector Machine</i> (SVM)	33
3.5	Validasi dan Evaluasi	34
3.5.1	Validasi Model	34
3.5.2	Evaluasi Model	35
BAB 4	HASIL DAN PEMBAHASAN	37
4.1	Pengumpulan Data	37
4.2	<i>Pre-Processing</i>	37
4.2.1	<i>Cleaning</i>	38
4.2.2	<i>Case Folding</i>	38
4.2.3	<i>Filtering Bahasa dan Filtering Kata</i>	39
4.2.4	<i>Normalization</i>	39
4.2.5	<i>Tokenizing</i>	40
4.2.6	<i>Stopword Removal</i>	41
4.2.7	<i>Stemming</i>	42
4.3	Pelabelan Data	42
4.3.1	Pelabelan Otomatis (<i>lexicon-based</i>)	43
4.3.2	Pemeriksaan dan Penghapusan	44
4.3.3	Validasi Pelabelan Pakar	44
4.3.4	Hasil Distribusi Label	45
4.4	Representasi Fitur	47
4.4.1	Representasi Fitur Menggunakan TF-IDF	47

4.4.2	Representasi Fitur Menggunakan <i>FastText</i>	48
4.4.3	Representasi Fitur Menggunakan BERT	51
4.5	Pemodelan Support Vector Machine (SVM)	53
4.5.1	Pengujian Menggunakan Parameter Default.....	53
4.5.2	Pengujian <i>Hyperparameter</i> Model <i>Support Vector Machine</i> (SVM).....	54
4.6	Validasi Model.....	55
4.6.1	Validasi TF-IDF	56
4.6.2	Validasi <i>FastText</i>	57
4.6.3	Validasi BERT	59
4.7	Evaluasi Model.....	60
4.7.1	Evaluasi TF-IDF dengan <i>Support Vector Machine</i> (SVM)	60
4.7.2	Evaluasi <i>FastText</i> dengan <i>Support Vector Machine</i> (SVM).....	62
4.7.3	Evaluasi BERT dengan <i>Support Vector Machine</i> (SVM).....	64
4.7.4	Perbandingan Hasil Evaluasi Model	65
BAB 5	PENUTUP.....	69
5.1	Kesimpulan	69
5.2	Saran.....	69
DAFTAR PUSTAKA		70
LAMPIRAN.....		75

DAFTAR GAMBAR

Gambar 2.1 Data Instagram 2025	11
Gambar 2.2 Ilustrasi SVM	18
Gambar 2.3 Simulasi K-Fold Cross Validation.....	20
Gambar 2.4 Tabel prediksi confusion matrix	21
Gambar 3.1 Diagram Alur Penelitian.....	23
Gambar 3.2 Data awal.....	24
Gambar 3.3 Tahapan Pre-processing.....	25
Gambar 4.1 Data komentar unggahan.....	37
Gambar 4.2 Hasil Proses Filtering	39
Gambar 4.3 Data label <i>lexicon-based</i>	43
Gambar 4.4 Distribusi label data (<i>lexicon based</i>).....	44
Gambar 4.5 Data label pakar bahasa Indonesia	45
Gambar 4.6 Distribusi label data (pakar)	46
Gambar 4.7 Implementasi parameter model SVM	54
Gambar 4.8 Implementasi <i>hyperparameter GridSearchCV</i>	55
Gambar 4.9 Implementasi <i>Stratified K-fold Cross Validation</i>	56
Gambar 4.10 Implementasi validasi model TF-IDF	56
Gambar 4.11 Akurasi <i>cross validation</i> tiap fold.....	57
Gambar 4.12 Perbandingan rata-rata akurasi tiap kernel SVM	57
Gambar 4.13 Implementasi validasi model <i>FastText</i>	57
Gambar 4.14 Akurasi <i>cross validation</i> tiap fold.....	58
Gambar 4.15 Perbandingan rata-rata akurasi kernel SVM	58
Gambar 4.16 Implementasi validasi model BERT.....	59
Gambar 4.17 Akurasi <i>cross validation</i> tiap fold	60
Gambar 4.18 Perbandingan rata-rata akurasi kernel SVM	60
Gambar 4.19 Implementasi evaluasi TF-IDF.....	61
Gambar 4.20 Visualisasi confusion matrix TF-IDF + SVM	61
Gambar 4.21 Implementasi evaluasi FastText	62
Gambar 4.22 Visualisasi confusion matrix <i>FastText</i> + SVM.....	63
Gambar 4.23 Implementasi evaluasi BERT	64
Gambar 4.24 Visualisasi confusion matrix BERT + SVM.....	64
Gambar 4.25 Perbandingan evaluasi kelas negatif	66
Gambar 4.26 Perbandingan evaluasi kelas positif	66
Gambar 4.27 Perbandingan akurasi	67

DAFTAR TABEL

Tabel 3.1 Proses <i>Cleaning</i>	25
Tabel 3.2 Proses <i>Case Folding</i>	25
Tabel 3.3 Proses <i>Normalization</i>	26
Tabel 3.4 Proses <i>Tokenization</i>	27
Tabel 3.5 Proses <i>Stopword Removal</i>	27
Tabel 3.6 Proses <i>Stemming</i>	28
Tabel 3.7 Hasil pelabelan data	29
Tabel 3.8 Tabel TF-IDF	30
Tabel 3.9 Vektor setiap Subword.....	31
Tabel 3.10 Vektor Kalimat	32
Tabel 3.11 Pengklasifikasian komentar.....	33
Tabel 3.12 Tabel prediksi confusion matrix	35
Tabel 4.1 Hasil proses <i>cleaning</i>	38
Tabel 4.2 Hasil proses <i>case folding</i>	38
Tabel 4.3 Hasil proses <i>normalization</i>	40
Tabel 4.4 Hasil proses <i>tokenizing</i>	40
Tabel 4.5 Hasil proses <i>stopword removal</i>	41
Tabel 4.6 Hasil proses <i>stemming</i>	42
Tabel 4.7 Hasil distribusi label.....	45
Tabel 4.8 Jumlah label.....	46
Tabel 4.9 Penyajian Data.....	47
Tabel 4.10 Hasil pembobotan TF-IDF	48
Tabel 4.11 Representasi vektor numerik	49
Tabel 4.12 Hasil vektor dokumen	50
Tabel 4.13 Representasi <i>FastText</i>	50
Tabel 4.14 Representasi hasil <i>embedding</i>	52
Tabel 4.15 Representasi hasil BERT	52
Tabel 4.16 Hasil pengujian awal parameter default	54
Tabel 4.17 Hasil kombinasi <i>parameter</i>	55
Tabel 4.18 Hasil validasi tiap kernel	56
Tabel 4.19 Hasil validasi tiap kernel	58
Tabel 4.20 Hasil validasi tiap kernel	59
Tabel 4.21 Hasil evaluasi TF-IDF dengan SVM.....	62
Tabel 4.22 Evaluasi model <i>FastText</i> + SVM.....	63
Tabel 4.23 Evaluasi model BERT + SVM	65
Tabel 4.24 Perbandingan evaluasi model kelas negatif	65
Tabel 4.25 Perbandingan evaluasi model kelas positif	66
Tabel 4.26 Perbandingan akurasi model	66