

PERBANDINGAN TINGKAT AKURASI ALGORITMA *NAIVE BAYES*, *DECISION TREE* DAN *RANDOM FOREST* DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI SAMSAT DIGITAL NASIONAL

SKRIPSI



Oleh:

Hilmi Ammar

2003015162

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH PROF. DR. HAMKA
JAKARTA
2026**

PERBANDINGAN TINGKAT AKURASI ALGORITMA *NAIVE BAYES*, *DECISION TREE* DAN *RANDOM FOREST* DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI SAMSAT DIGITAL NASIONAL

SKRIPSI

Dibuat untuk Memenuhi Persyaratan Kelulusan Sarjana Teknik Informatika



Oleh:

Hilmi Ammar

2003015162

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH PROF. DR. HAMKA
JAKARTA
2026**

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

PERBANDINGAN TINGKAT AKURASI ALGORITMA NAIVE BAYES, DECISION TREE DAN RANDOM FOREST DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI SAMSAT DIGITAL NASIONAL

SKRIPSI

Dibuat untuk Memenuhi Persyaratan Kelulusan Sarjana Teknik

Oleh:
Hilmi Ammar
2003015162

Telah diperiksa dan disetujui untuk diajukan ke Sidang Ujian Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Industri dan
Informatika UHAMKA
Tanggal, 5 Januari 2025

Pembimbing-1



Ir. Ade Davy Wiranata, S.Kom., M.Kom.
NIDN. 0325119302

Mengetahui,
Ketua Program Studi Teknik Informatika



Mia Kamayani, S.T., M.T.
NIDN. 0312028704

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

PERBANDINGAN TINGKAT AKURASI ALGORITMA NAIVE BAYES, DECISION TREE DAN RANDOM FOREST DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI SAMSAT DIGITAL NASIONAL

SKRIPSI

Oleh:
Hilmi Ammar
2003015162

Telah diuji dan dinyatakan lulus dalam Sidang Ujian Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Industri dan Informatika
Universitas Muhammadiyah Prof. Dr. Hamka
Tanggal, 24 April 2026

Pembimbing-1

Ir. Ade Davy Wiranata, S.Kom, M.Kom
NIDN. 0325119302

Penguji-1

Firman Noor Hasan, S.Kom., M.TI.
NIDN. 0301088305

Penguji-2

Estu Sinduningrum, S.T., M.T.
NIDN. 0314098403

Mengesahkan,
Dekan
Fakultas Teknologi Industri dan
Informatika

Dr. Ir. Dan Mugisidi, S.T., M.Si., IPP.
NIDN. 0301126901

Mengetahui,
Ketua Program Studi
Teknik Informatika

Mia Kamayani, S.T., M.T.
NIDN. 0312028704

PERNYATAAN KEASLIAN

PERNYATAAN KEASLIAN

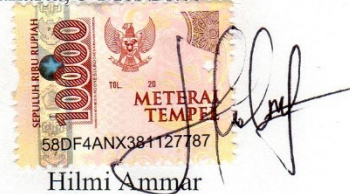
Saya, yang membuat pernyataan

Nama : Hilmi Ammar
NIM : 2003015162
Judul skripsi : Perbandingan Tingkat Akurasi Algoritma *Naive Bayes*,
Decision Tree dan *Random Forest* Dalam Analisis Sentimen
Pengguna Aplikasi Samsat Digital Nasional

Menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI) dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi mana pun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, KECUALI yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Referensi.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Jakarta, 5 Mei 2026



10000
SEPULUH RIBU RUPIAH
TEL. 20
METERAL
TEMPEL
58DF4ANX381127787
Hilmi Ammar

KATA PENGANTAR

Assallamu'alaikum wa rohmatullahi wa barokaatuh.

Segala puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi berjudul **“Perbandingan Tingkat Akurasi Algoritma *Naive Bayes*, *Decision Tree* dan *Random Forest* dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional”**. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Universitas Muhammadiyah Prof. Dr. HAMKA.

Dalam proses penyusunannya, penulis mendapat banyak bantuan dan dukungan. Ucapan terima kasih penulis sampaikan kepada:

1. Bapak Dr. Ir. Dan Mugisidi, S.T., M.Si.,IPP selaku dekan dan Ibu Mia Kamayani Sulaeman, S.T., M.T. selaku Ketua Program Studi Fakultas Teknologi Industri dan Informatika Universitas Muhammadiyah Prof. Dr. Hamka.
2. Bapak Ir. Ade Davy Wiranata, S.Kom., M.Kom, selaku dosen pembimbing skripsi yang telah dengan sabar membimbing, memberi masukan, dan motivasi.
3. Bapak Faldy Irwiensyah, S.Kom., M.TI, selaku dosen pembimbing akademik, yang telah mendukung sejak awal perkuliahan hingga akhir masa studi.
4. Kedua orang tua penulis, Bapak Darsono dan Ibu Eli Sumarni yang telah memberikan dukungan berupa materi, nasehat serta doa.
5. Seluruh keluarga dan teman-teman tersayang yang senantiasa mendoakan serta memberikan semangat dalam menyelesaikan skripsi.

Penulis berharap skripsi ini bermanfaat bagi pembaca serta dapat menjadi referensi bagi penelitian selanjutnya dengan topik serupa. Penulis juga membuka diri terhadap saran dan masukan demi perbaikan penelitian ini di masa mendatang.

PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Prof. DR. HAMKA (UHAMKA), saya yang bertandatangan di bawah ini:

Nama : Hilmi Ammar
NIM : 2003015162
Program Studi : Teknik Informatika

Menyetujui, memberikan Hak Bebas Royalti Noneksklusif (*non-exclusive royalty free right*) kepada Universitas Muhammadiyah Prof. DR. HAMKA (UHAMKA) atas karya ilmiah saya beserta perangkat yang ada (jika diperlukan) yang berjudul:

Perbandingan Tingkat Akurasi Algoritma *Naive Bayes*, *Decision Tree* dan *Random Forest* Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional

Hak Bebas Royalti Noneksklusif ini Universitas Muhammadiyah Prof. DR. HAMKA berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Jakarta, 5 Mei 2026



Hilmi Ammar

ABSTRAK

Perbandingan Tingkat Akurasi Algoritma *Naïve Bayes*, *Decision Tree* dan *Random Forest* Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional

Hilmi Ammar

Penelitian ini bertujuan menganalisis sentimen pengguna terhadap aplikasi Signal – Samsat Digital Nasional menggunakan 2.000 ulasan, yang setelah melalui proses *filtering* dan *preprocessing* menjadi 1.743 data, kemudian dibagi menjadi 80% data *training* (1.394) dan 20% data *testing* (349). Data dianalisis menggunakan tiga algoritma klasifikasi, yaitu *Naïve Bayes*, *Decision Tree*, dan *Random Forest*. Hasil menunjukkan bahwa *Naïve Bayes* memperoleh akurasi 89.4% dengan kinerja sangat baik pada sentimen positif namun kurang optimal pada sentimen negatif, sedangkan *Decision Tree* mencapai akurasi 86.8% dengan hasil yang lebih seimbang. *Random Forest* menunjukkan performa terbaik dengan akurasi 92.9% serta kemampuan klasifikasi yang lebih seimbang antara sentimen positif dan negatif. Secara keseluruhan, *Random Forest* merupakan algoritma paling optimal dalam penelitian ini. Visualisasi data juga menunjukkan mayoritas ulasan berupa rating bintang 5, dengan distribusi sentimen 82,9% positif dan 17,1% negatif, sehingga dapat disimpulkan pengalaman pengguna aplikasi *Signal* adalah positif.

Kata Kunci: Analisis Sentimen, Samsat, *Naïve Bayes*, *Decision Tree*, *Random Forest*

Comparison of the Accuracy Levels of Naïve Bayes, Decision Tree, and Random Forest Algorithms in Sentiment Analysis of National Digital Samsat Application Users

Hilmi Ammar

This research aims to analyze user sentiment towards the Signal – Samsat Digital Nasional application using 2,000 reviews, which after going through a filtering and preprocessing process became 1,743 data, then divided into 80% training data (1,394) and 20% testing data (349). The data were analyzed using three classification algorithms, namely Naïve Bayes, Decision Tree, and Random Forest. The results show that Naïve Bayes obtained an accuracy of 89.4% with very good performance on positive sentiment but less optimal on negative sentiment, while Decision Tree achieved an accuracy of 86.8% with more balanced results. Random Forest showed the best performance with an accuracy of 92.9% and a more balanced classification ability between positive and negative sentiment. Overall, Random Forest is the most optimal algorithm in this study. Data visualization also shows that the majority of reviews are in the form of 5-star ratings, with a sentiment distribution of 82.9% positive and 17.1% negative, so it can be concluded that the user experience of the Signal application is positive.

Keywords: Sentiment Analysis, Samsat, *Naïve Bayes*, *Decision Tree*, *Random Forest*

DAFTAR ISI

HALAMAN PERSETUJUAN	i
HALAMAN PENGESAHAN	ii
PERNYATAAN KEASLIAN	iii
KATA PENGANTAR	iv
PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	v
ABSTRAK	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR NOTASI	xi
DAFTAR LAMPIRAN	xii
BAB.1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB.2 DASAR TEORI	8
2.1 Kerangka Teori	8
2.2 Penelitian Terdahulu	8
2.3 Jenis Referensi	11
2.3.1 Analisis Sentimen	11
2.3.2 Signal – Samsat Digital Nasional	11
2.3.3 <i>Scraping</i> Data	12
2.3.4 <i>Naïve Bayes</i>	12
2.3.5 <i>Decision Tree</i>	13
2.3.6 <i>Random Forest</i>	13
2.3.7 Pra-Pemrosesan Teks	13
2.3.8 <i>TF-IDF</i>	14
2.3.9 <i>Confusion Matrix</i>	15
BAB.3 METODOLOGI PENELITIAN	16
3.1 Alur Penelitian	16
3.1.1 Pengumpulan Data	16
3.1.2 <i>Filter</i> Data	17
3.1.3 <i>Text Processing</i>	17
3.1.4 Pelabelan Data	20

3.1.5 <i>Split</i> Data.....	21
3.1.6 Pembobotan <i>TF-IDF</i>	21
3.1.7 Klasifikasi Model.....	22
3.1.8 Evaluasi Algoritma.....	22
3.1.9 Visualisasi Data.....	23
3.2 Identifikasi Kebutuhan	23
3.3 Jadwal Penelitian.....	24
BAB.4 HASIL DAN PEMBAHASAN	25
4.1 Pengumpulan Data	25
4.2 Filter Data.....	26
4.3 <i>Text Processing</i>	28
4.4 Pelabelan Data.....	29
4.5 <i>Split</i> Data.....	31
4.6 Pembobotan <i>TF-IDF</i>	32
4.7 Klasifikasi Model.....	34
4.8 Evaluasi Algoritma.....	35
4.9 Hasil Perbandingan Akurasi Algoritma	40
4.10 Visualisasi Data.....	41
BAB.5 KESIMPULAN DAN SARAN	44
5.1 Kesimpulan	44
5.2 Saran.....	45
DATAR PUSTAKA	46
LAMPIRAN.....	52

DAFTAR TABEL

Tabel 2. 1. Penelitian Terdahulu	8
Tabel 2.2. Nilai Aktual <i>Confusion Matrix</i>	15
Tabel 3.1. Spesifikasi <i>Hardware</i>	23
Tabel 3.2. Spesifikasi <i>Software</i>	24
Tabel 3.3. Jadwal Rencana Kegiatan Penelitian	24
Tabel 4.1. Ilustrasi Hasil <i>Web Scraping</i>	26
Tabel 4.2. Ilustrasi Hasil <i>Text Processing</i>	29
Tabel 4.3. Ilustrasi Hasil Pelabelan Sentimen.....	30
Tabel 4.4. Ilustrasi Hasil Bobot Kata TF-IDF.....	33
Tabel 4.5. Hasil Perbandingan Evaluasi Algoritma	40

DAFTAR GAMBAR

Gambar 2.1. Kerangka Teori.....	8
Gambar 3.1. Alur Penelitian.....	16
Gambar 4.1. Alur Pengumpulan Data	25
Gambar 4.2. Hasil Pengumpulan Data Melalui <i>Web Scraping</i>	25
Gambar 4.3. <i>Filtering</i> Kolom Utama Penelitian.....	27
Gambar 4.4. Hasil Proses <i>Filtering</i> Data	28
Gambar 4.5. Hasil Proses <i>Text Processing</i>	28
Gambar 4.6. Hasil Pelabelan Data Sentimen	30
Gambar 4.7. Hasil <i>Split</i> Data	32
Gambar 4.8. Hasil Pembobotan <i>TF-IDF</i>	34
Gambar 4.9. Hasil Klasifikasi Model.....	35
Gambar 4.10. Hasil <i>Confusion Matrix Heatmap Naïve Bayes</i>	36
Gambar 4.11. Hasil <i>Confusion Matrix Heatmap Decision Tree</i>	37
Gambar 4.12. Hasil <i>Confusion Matrix Heatmap Random Forest</i>	39
Gambar 4.13. Hasil Distribusi Sentimen Dengan Kamus Lexicon.....	41
Gambar 4.14. Hasil Jumlah Pengguna Memberikan Rating	42
Gambar 4.15. Hasil <i>Word Cloud</i> Kata	43

DAFTAR NOTASI

No.	Notasi	Keterangan
1	TP (<i>True Positive</i>)	Jumlah data dengan nilai aktual kelas positif yang juga diprediksi sebagai kelas positif.
2	TN (<i>True Negative</i>)	Jumlah data dengan nilai aktual kelas negatif yang juga diprediksi sebagai kelas negatif.
3	FP (<i>False Positive</i>)	Jumlah data dengan nilai aktual kelas positif yang salah diprediksi sebagai kelas negatif.
4	FN (<i>False Negative</i>)	Jumlah data dengan nilai aktual kelas positif yang salah diprediksi sebagai kelas negatif.

DAFTAR LAMPIRAN

Lampiran 1. LoA.....	52
Lampiran 2. Editor Decision.....	53
Lampiran 3. Sertifikat Publisher	54
Lampiran 4. Lembar Bimbingan Skripsi.....	55
Lampiran 5. Turnitin	56
Lampiran 6. Lembar Revisi Penguji 1	57
Lampiran 7. Lembar Revisi Penguji 2	58
Lampiran 8. Jurnal	74

BAB.1 PENDAHULUAN

1.1 Latar Belakang

Adopsi teknologi digital telah membawa dampak transformatif terhadap model layanan publik di Indonesia, khususnya melalui pemanfaatan teknologi dalam sektor pemerintahan. Transformasi digital menjadi instrumen penting dalam menciptakan layanan yang lebih efisien, transparan, serta mampu meningkatkan partisipasi masyarakat. Seiring dengan perkembangan teknologi yang pesat, pemerintah terus mendorong inovasi digital guna meningkatkan kualitas dan efektivitas tata kelola pelayanan publik (Ulfa & Frinaldi, 2025).

Sistem Administrasi Manunggal Satu Atap (*SAMSAT*) merupakan salah satu bukti konkret transformasi tersebut, dengan dasar operasional yang merujuk pada ketentuan UU Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan. Sistem ini dirancang sebagai mekanisme administrasi terpadu untuk mengelola perpajakan kendaraan bermotor dengan tujuan meningkatkan kepatuhan wajib pajak serta menyediakan layanan yang efektif dan terintegrasi (Wijaya dkk., 2024). Melalui *SAMSAT*, pemerintah dapat memantau dan mengoptimalkan penerimaan pajak kendaraan sebagai salah satu sumber pembiayaan pembangunan infrastruktur dan layanan publik lainnya (Afifudin dkk., 2020).

Kolaborasi antara Korlantas Polri, Kementerian Dalam Negeri, dan Jasa Raharja melahirkan aplikasi *SIGNAL - Samsat Digital Nasional* sebagai wujud digitalisasi layanan *SAMSAT* untuk meningkatkan standar pelayanan kepada masyarakat. Aplikasi ini memungkinkan masyarakat untuk melakukan pembayaran *PKB* dan pengesahan *STNK* secara *online*. (Suriyani BB dkk., 2024). Kehadiran aplikasi *SIGNAL* bertujuan untuk menyederhanakan proses birokrasi, mengurangi antrean di kantor *SAMSAT*, meningkatkan kepatuhan wajib pajak, serta meminimalkan potensi praktik pungutan liar melalui sistem yang lebih transparan (Afandi Afandi dkk., 2025).

Akan tetapi, implementasi *SIGNAL* di lapangan masih dihadapkan oleh berbagai kendala teknis. Meskipun aplikasi *SIGNAL* dinilai membantu dari sisi