

**PERBANDINGAN PERFORMANSI METODE SUPPORT
VECTOR MACHINE DAN RANDOM FOREST PADA
ANALISIS SENTIMEN ULASAN APLIKASI PELAYANAN
PUBLIK SIPRAJA DENGAN PENERAPAN TEKNIK SMOTE**

SKRIPSI



Oleh :

AISYAH PERTIWI

20081010083

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAWA TIMUR
2024**

LEMBAR PENGESAHAN
SKRIPSI

Judul : PERBANDINGAN PERFORMANSI METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST PADA ANALISIS SENTIMEN ULASAN APLIKASI PELAYANAN PUBLIK SIPRAJA DENGAN PENERAPAN TEKNIK SMOTE

Oleh : Aisyah Pertiwi

NPM : 20081010083

Telah Diseminarkan Dalam Ujian Skripsi Pada :

Hari Selasa, Tanggal 21 Mei 2024

Mengetahui

Dosen Pembimbing

1.



**Eva Yulia Puspaningrum, S.Kom.,
M.Kom.**

NIP : 19890705 2021212 002

2.

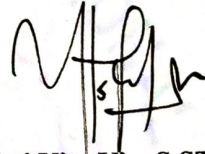


Hendra Maulana, S.Kom., M.Kom.

NPT : 201198 31 223248

Dosen Penguji

1.



Yisti Vita Via, S.ST., M.Kom.

NIP : 19860425 2021212 001

2.



Henni Endah Wahanani, S.T., M.Kom.

NIP : 19780922 2021212 005

Menyetujui

**Dekan
Fakultas Ilmu Komputer**



Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP : 19681126 199403 2 001

**Koordinator Program Studi
Teknik Informatika**



Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP : 19820211 2021212 005

SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Aisyah Pertiwi

NPM : 20081010083

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

“PERBANDINGAN PERFORMANSI METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST PADA ANALISIS SENTIMEN ULASAN APLIKASI PELAYANAN PUBLIK SIPRAJA DENGAN PENERAPAN TEKNIK SMOTE”

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dan juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.



Surabaya, 31 Mei 2024



NPM. 20081010083

PERBANDINGAN PERFORMAN SI METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST PADA ANALISIS SENTIMEN ULASAN APLIKASI PELAYANAN PUBLIK SIPRAJA DENGAN PENERAPAN TEKNIK SMOTE

Nama : Aisyah Pertiwi

NPM : 20081010083

Jurusan : Informatika

Pembimbing : Eva Yulia Puspaningrum, S.Kom., M.Kom.

Hendra Maulana, S.Kom., M.Kom.

Abstrak

Pemerintah sebagai penyelenggara layanan publik memiliki kewajiban untuk memberikan pelayanan yang terbaik kepada publik. Namun, dalam mencapai hal tersebut, pemerintah dihadapkan pada berbagai tantangan yang menuntut mereka untuk terus menerus mengembangkan ide-ide baru dan berinovasi untuk mengatasinya. Salah satunya adalah tantangan global dari perubahan dan kemajuan teknologi yang pesat. Dalam menghadapi tantangan tersebut, Pemerintah Kabupaten Sidoarjo telah memperkenalkan Aplikasi Sistem Pelayanan Rakyat Sidoarjo (Sipraja). Pada penelitian ini akan dilakukan analisis sentimen ulasan aplikasi Sipraja dengan menggunakan algoritma Support Vector Machines (SVM) dan Random Forest dengan menerapkan teknik SMOTE. Data yang digunakan adalah sebanyak 1137 data. Tujuan utama dari penelitian ini adalah untuk membandingkan performa antara metode SVM dan RF dalam menganalisis sentimen. Hasil penelitian menunjukkan bahwa implementasi metode SMOTE berhasil meningkatkan nilai akurasi pada kedua algoritma. Pada algoritma SVM, nilai akurasi meningkat dari 87% menjadi 95%, precision dari 82% menjadi 95%, recall dari 78% menjadi 95%, dan F1-Score dari 80% menjadi 95%. Sementara itu, pada algoritma Random Forest, akurasi meningkat dari 85% menjadi 90%, precision dari 79% menjadi 90%, recall dari 73% menjadi 90%, dan F1-Score dari 75% menjadi 90%. Berdasarkan hasil pengujian, algoritma SVM menunjukkan performa yang lebih baik dibandingkan dengan algoritma RF dalam analisis sentimen aplikasi Sipraja, baik dengan maupun tanpa penerapan SMOTE,

dengan akurasi akhir mencapai 95.03% dibandingkan dengan 90.12% pada algoritma RF.

Kata Kunci: Analisis Sentimen, Support Vector Machine, Random Forest, SMOTE

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala keberkahan dan kesehatan yang diberikan, sehingga penulis dapat menyelesaikan penelitian yang berjudul "Perbandingan Performansi Metode Support Vector Machine dan Random Forest pada Analisis Sentimen Ulasan Aplikasi Pelayanan Publik Sipraja dengan Penerapan Teknik SMOTE".

Laporan akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana (S1) di Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur. Penulis berharap laporan akhir ini dapat memberikan kontribusi ilmiah dan bermanfaat bagi semua pihak yang membacanya.

Penulis menyadari bahwa penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak yang telah memberikan bimbingan, motivasi, dan dukungan, sehingga laporan ini dapat diselesaikan dengan baik. Penulis juga menyadari bahwa laporan akhir ini masih jauh dari sempurna, dengan banyaknya kekurangan yang ada karena keterbatasan pengetahuan dan kemampuan penulis. Oleh karena itu, penulis sangat mengharapkan masukan, kritik, dan saran yang konstruktif dari semua pihak untuk penyempurnaan laporan ini di masa mendatang.

Surabaya, 21 Mei 2024

Penulis

UCAPAN TERIMA KASIH

Puji syukur kami panjatkan ke hadirat Allah SWT atas segala limpahan berkat, rahmat, dan inayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Penulis menyadari bahwa pencapaian ini tidak akan terwujud tanpa bimbingan, saran, motivasi, dan bantuan dari berbagai pihak yang telah memberikan dukungan yang luar biasa. Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua yang selalu memberikan dukungan moral dan spiritual serta doa yang tiada henti, yang menjadi sumber semangat dan inspirasi penulis untuk terus berjuang menyelesaikan skripsi ini.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT. selaku Rektor Universitas Pembangunan Nasional Veteran Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer
4. Ibu Fetty Try Anggraeny, S.Kom., M.Kom. selaku Koordinator Program Studi Informatika
5. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom. selaku Dosen Pembimbing I, yang telah memberikan bimbingan, saran, dan dukungan dalam setiap tahapan penulisan skripsi ini.
6. Bapak Hendra Maulana, S.Kom., M.Kom. selaku Dosen Pembimbing II, yang telah memberikan bimbingan dan nasihat dalam penulisan skripsi ini.
7. Seluruh teman-teman yang telah memberikan dukungan dan bantuan selama penulis menyelesaikan skripsi ini. Tanpa dukungan dan kerja sama kalian, penulis tidak akan mampu mencapai hasil seperti sekarang.

Ucapan terima kasih ini hanyalah sebagian kecil dari penghargaan penulis atas segala bantuan dan dukungan yang diberikan. Semoga kebaikan dan bantuan yang telah diberikan mendapatkan balasan yang setimpal dari Allah SWT.

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
SURAT PERNYATAAN BEBAS DARI PLAGIASI.....	ii
Abstrak	iii
KATA PENGANTAR	v
UCAPAN TERIMA KASIH.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiv
BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan.....	4
1.4 Manfaat.....	4
1.5 Batasan Masalah.....	5
BAB II.....	6
2.1 Penelitian Terdahulu.....	6
2.2 Analisis Sentimen.....	8
2.3 Sipraja.....	8
2.4 <i>Text Preprocessing</i>	8
2.4.1 <i>Case Folding</i>	9
2.4.2 <i>Cleansing</i>	9
2.4.3 <i>Tokenizing</i>	9
2.4.4 <i>Stopword Removal</i>	9
2.4.5 <i>Stemming</i>	9

2.5 TF-IDF.....	10
2.6 SMOTE	11
2.7 SVM	12
2.8 Random Forest	15
2.9 K-Fold Cross Validation	17
2.10 <i>Grid Search</i>	18
2.11 <i>Confusion Matrix</i>	19
BAB III	21
3.1 Tahapan Penelitian	21
3.2 Pengumpulan data	22
3.3 Eksplorasi Data.....	23
3.4 Preprocessing Data	23
3.4.1 Emoji Handling	23
3.3.2 Case Folding.....	24
3.3.3 Cleansing	24
3.3.4 Normalization.....	25
3.3.5 Tokenizing.....	26
3.3.6 Stopword Removal	26
3.3.7 Stemming.....	27
3.5 Pelabelan Data	28
3.6 Pembobotan Kata	29
3.7 SMOTE	32
3.8 <i>Hyperparameter Tuning dengan Grid Search</i>	35
3.9 Klasifikasi.....	36
3.9.1 SVM	36
3.8.2 Random Forest	44

3.8 Evaluasi Model	48
3.9 Skenario Pengujian	48
BAB IV	49
4.1 Pengumpulan data	49
4.2 Eksplorasi Data.....	51
4.3 Preprocessing Data	51
4.3.1 Emoji Handling	52
4.3.2 Case Folding.....	53
4.3.3 Cleansing	53
4.3.4 Normalization	55
4.3.5 Tokenizing.....	56
4.3.6 Stopword Removal	57
4.3.7 Stemming.....	58
4.4 Pelabelan data.....	59
4.5 Pembobotan Kata	61
4.6 Synthetic Minority Over Sampling Technique (SMOTE)	62
4.7 Split Data.....	64
4.8 Klasifikasi Data	64
4.8.1 Support Vector Machine	69
4.8.2 Hyperparameter Tuning SVM.....	69
4.8.3 Random Forest	72
4.8.4 Hyperparameter Tuning Random Forest.....	73
4.9 Evaluasi Model.....	74
4.10 Evaluasi Model Skenario Pengujian Support Vector Machine	75
4.11 Evaluasi Model Skenario Pengujian SVM dengan SMOTE.....	77
4.12 Evaluasi Model Skenario <i>Hyperparameter Tuning SVM</i>	79

4.13 Evaluasi Model Skenario Pengujian <i>Hyperparameter Tuning</i> SVM dengan SMOTE	82
4.14 Evaluasi Model Skenario Pengujian Random Forest	84
4.15 Evaluasi Model Skenario Pengujian Random Forest dengan SMOTE	86
4.16 Evaluasi Model Skenario Pengujian <i>Hyperparameter Tuning</i> Random Forest	88
4.17 Evaluasi Model Skenario Pengujian <i>Hyperparameter Tuning</i> Random Forest dengan SMOTE	91
BAB V	96
5.1 Kesimpulan	96
5.2. Saran	96
DAFTAR PUSTAKA	97

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Grid Search.....	18
Gambar 3.1 Tahapan Penelitian	21
Gambar 3.2 Contoh Data Ulasan Sipraja	22
Gambar 3.3 Diagram Alur Pelabelan Data	28
Gambar 3.4 Diagram Alur Pembobotan Kata	29
Gambar 3.5 Diagram Alur SMOTE.....	32
Gambar 3.6 Diagram Alur Grid Search	35
Gambar 3.7 Diagram Alur Klasifikasi SVM.....	36
Gambar 3.8 Diagram Alur Klasifikasi Random Forest.....	44
Gambar 3.9 Contoh Pohon Keputusan.....	47
Gambar 4.1 Contoh Kamus Normalisasi	51
Gambar 4.2 Hasil Implementasi <i>Emoji Handling</i>	52
Gambar 4.3 Hasil Implementasi <i>Case Folding</i>	53
Gambar 4.4 Hasil Implementasi <i>Cleansing</i>	55
Gambar 4.5 Hasil Implementasi <i>Normalization</i>	56
Gambar 4.6 Hasil Implementasi <i>Tokenizing</i>	57
Gambar 4.7 Hasil Implementasi <i>Stopword Removal</i>	58
Gambar 4.8 Hasil Implementasi <i>Stemming</i>	59
Gambar 4.9 Hasil Translasi.....	59
Gambar 4.10 Hasil Pembobotan Kata.....	62
Gambar 4.11 Visualisasi Data Sebelum SMOTE	63
Gambar 4.12 Visualisasi Data Sesudah SMOTE	63
Gambar 4.13 Data Hasil SMOTE	64
Gambar 4.14 Akurasi Skenario SVM	75
Gambar 4.15 <i>Confusion Matrix</i> Skenario SVM.....	76

Gambar 4.16 <i>Classification Report</i> Skenario SVM.....	76
Gambar 4.17 Akurasi Skenario SVM dengan SMOTE	77
Gambar 4.18 <i>Counfusion Matrix</i> Skenario SVM dengan SMOTE.....	78
Gambar 4.19 <i>Classification Report</i> Skenario SVM dengan SMOTE.....	78
Gambar 4.20 Hasil <i>Hyperparameter Tuning</i> SVM.....	79
Gambar 4.21 Akurasi Skenario <i>Hyperparameter Tuning</i> SVM	80
Gambar 4.22 <i>Counfusion Matrix</i> Skenario <i>Hyperparameter Tuning</i> SVM.....	80
Gambar 4.23 <i>Classification Report</i> Skenario <i>Hyperparameter Tuning</i> SVM.....	81
Gambar 4.24 Hasil <i>Hyperparameter Tuning</i> SVM dengan data hasil SMOTE ...	82
Gambar 4.25 Akurasi Skenario <i>Hyperparameter Tuning</i> SVM dengan SMOTE	82
Gambar 4.26 <i>Counfusion Matrix</i> Skenario <i>Hyperparameter Tuning</i> SVM dengan SMOTE	83
Gambar 4.27 <i>Classification Report</i> Skenario <i>Hyperparameter Tuning</i> SVM dengan SMOTE.....	83
Gambar 4.28 Akurasi Skenario Pengujian Random Forest.....	84
Gambar 4.29 <i>Confusion Matrix</i> Skenario Pengujian Random Forest.....	85
Gambar 4.30 <i>Classification Report</i> Skenario Pengujian Random Forest.....	85
Gambar 4.31 Akurasi Skenario Pengujian Random Forest dengan SMOTE	86
Gambar 4.32 <i>Confusion Matrix</i> Skenario Pengujian Random Forest dengan SMOTE	87
Gambar 4.33 <i>Classification Report</i> Skenario Pengujian Random Forest dengan SMOTE	87
Gambar 4.34 Hasil <i>Hyperparameter Tuning</i> Random Forest.....	88
Gambar 4.35 Akurasi Skenario Pengujian <i>Hyperparameter Tuning</i> Random Forest.....	89
Gambar 4.36 <i>Confusion Matrix</i> Skenario Pengujian <i>Hyperparameter Tuning</i> Random Forest	89

Gambar 4.37 <i>Classification Report</i> Skenario Pengujian <i>Hyperparameter Tuning</i> Random Forest	90
Gambar 4.38 Hasil <i>Hyperparameter Tuning</i> Random Forest dengan data hasil SMOTE	91
Gambar 4.39 Akurasi Skenario Pengujian <i>Hyperparameter Tuning</i> Random Forest dengan SMOTE.....	91
Gambar 4.40 <i>Confusion Matrix</i> Skenario Pengujian <i>Hyperparameter Tuning</i> Random Forest dengan SMOTE	92
Gambar 4.41 <i>Classification Report</i> Skenario Pengujian <i>Hyperparameter Tuning</i> Random Forest dengan SMOTE	93

DAFTAR TABEL

Tabel 2.1 Tabel <i>Confusion Matrix</i>	20
Tabel 3.1 Hasil <i>emoji handling</i>	23
Tabel 3.2 Hasil <i>case folding</i>	24
Tabel 3.3 Hasil <i>cleansing</i>	25
Tabel 3.4 Hasil <i>normalization</i>	25
Tabel 3.5 Hasil <i>tokenizing</i>	26
Tabel 3.6 Hasil stopword removal	26
Tabel 3.7 Hasil <i>stemming</i>	27
Tabel 3.8 Contoh pelabelan data	29
Tabel 3.9 Jumlah data tiap sentimen	29
Tabel 3.10 Contoh data ulasan	30
Tabel 3.11 Contoh hasil TF, DF, IDF	31
Tabel 3.12 Contoh hasil TF-IDF	32
Tabel 3.13 Data hasil TF-IDF	33
Tabel 3.14 Data hasil TF-IDF untuk perhitungan SVM	37
Tabel 3.15 Contoh pasangan matriks kernel	39
Tabel 3.16 Contoh hasil perhitungan kernelisasi data latih	39
Tabel 3.17 TF data Uji	43
Tabel 3.18 Hasil kernelisasi data uji	43
Tabel 3.19 Data hasil pengambilan sampel acak	45
Tabel 3.20 Contoh hasil perhitungan informasi gain node pertama	46
Tabel 3.21 Contoh hasil perhitungan informasi gain node kedua	47
Tabel 3.22 Contoh hasil perhitungan informasi gain node pertama	48
Tabel 4.1 Hasil pengumpulan data sipraja	50
Tabel 4.2 Hasil pelabelan data	61

Tabel 4.3 Perbandingan Jumlah Data Hasil SMOTE.....	63
Tabel 4.4 Hasil Pengujian Nilai K	64
Tabel 4.5 Perbandingan Hasil Skenario Pengujian	93