

TUGAS AKHIR

**PENGARUH METODE *IMBALANCE DATA SAMPLING*
TERHADAP KINERJA *SUPPORT VECTOR MACHINE* UNTUK
SENTIMEN ANALISIS NATURALISASI PEMAIN
SEPAK BOLA**



Ferdy Ilham Maulana
2110651116

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2025**

TUGAS AKHIR

PENGARUH METODE *IMBALANCE DATA SAMPLING* TERHADAP KINERJA *SUPPORT VECTOR MACHINE* UNTUK SENTIMEN ANALISIS NATURALISASI PEMAIN SEPAK BOLA

Diajukan untuk Memenuhi Persyaratan Guna Meraih Gelar Sarjana Komputer
Teknik Informatika Universitas Muhammadiyah Jember



Ferdy Ilham Maulana
2110651116

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2025**

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini:

NAMA : FERDY ILHAM MAULANA

NIM : 21 1065 1116

INSTITUSI : Strata-1 Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember.

Menyatakan bahwa Tugas Akhir yang berjudul **“PENGARUH METODE *IMBALANCE DATA SAMPLING* TERHADAP KINERJA *SUPPORT VECTOR MACHINE* UNTUK SENTIMEN ANALISIS NATURALISASI PEMAIN SEPAK BOLA”** bukan merupakan Tugas Akhir orang lain baik sebagian maupun keseluruhan kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar penulis bersedia mendapatkan sanksi dari akademik.

Jember, 2025

Yang membuat pernyataan

A handwritten signature in blue ink is written over a yellow revenue stamp. The stamp features a soccer ball and the text 'PENGULANG KUTIPAN', '1000', 'METERAI TEMPEL', and the alphanumeric code 'I-QZY1625X426535400'.

Ferdy Ilham Maulana

NIM. 2110651116

HALAMAN PERSETUJUAN

PENGARUH METODE *IMBALANCE DATA SAMPLING* TERHADAP KINERJA *SUPPORT VECTOR MACHINE* UNTUK SENTIMEN ANALISIS NATURALISASI PEMAIN SEPAK BOLA

Ferdy Ilham Maulana

2110651116

Telah disetujui bahwa 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

Universitas Muhammadiyah Jember

Jember, 18 Juli 2025

Menyetujui,

Pembimbing I

Pembimbing II

Deni Arifianto, M.Kom.

NIDN. 0718068103

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

NIDN. 0728079101

HALAMAN PENGESAHAN

PENGARUH METODE *IMBALANCE DATA SAMPLING* TERHADAP KINERJA *SUPPORT VECTOR MACHINE* UNTUK SENTIMEN ANALISIS NATURALISASI PEMAIN SEPAK BOLA

Oleh:

Ferdy Ilham Maulana

2110651116

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada Sidang Tugas Akhir Tanggal 18 Bulan Juli 2025 sebagai salah satu kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh,

Dosen Penguji I



Dudi Irawan, S.T., M.Kom
NIDN. 0730037703

Dosen Pembimbing I



Deni Arifianto, M.Kom
NIDN. 0718068103

Dosen Penguji II



Luluk Handayani, S.Si., M.Si.
NIDN. 0725108003

Dosen Pembimbing II

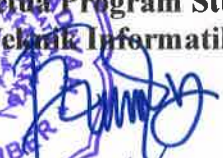


Dr. Reni Umilasari, S.Pd., M.Si
NIDN. 0728079101

Mengesahkan,
Dekan Fakultas Teknik


Dr. Ir. Muhtar, ST., M.T., IPM
NIDN. 0010067301

Mengetahui,
Ketua Program Studi
Teknik Informatika


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

MOTTO

“Telur Busuk Tidak Terlihat Dari Luar Jadi Pecahkan Jika Kamu Ingin Tahu
Dalamnya”

“Jika Lalat Jatuh Ke Air Putih, Air Putihnya Akan Dibuang Jika Lalat Jatuh Ke
Madu Maka Lalatnya Akan Dibuang”



KATA PENGANTAR

Dengan menyebut nama Allah Yang Maha Pengasih dan Maha Penyayang, puji syukur penulis panjatkan atas rahmat dan petunjuk-Nya sehingga tugas akhir yang berjudul “Pengaruh Metode *Imbalance Data Sampling* Terhadap Kinerja *Support Vector Machine* Untuk Sentimen Analisis Naturalisasi Pemain Sepak Bola”. Di balik terselesaikannya laporan ini, banyak pihak yang telah memberikan dukungan baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, atas segala nikmat, kekuatan, dan kelapangan waktu yang telah diberikan selama proses penyusunan laporan ini.
2. Para dosen Universitas Muhammadiyah Jember, khususnya dosen-dosen di Program Studi Teknik Informatika yang telah membekali penulis dengan ilmu, wawasan, dan bimbingan selama masa perkuliahan.
3. Bapak Deni Arifianto dan Ibu Reni Umilasari selaku dosen pembimbing yang dengan penuh kesabaran dan ketulusan telah membimbing, memberikan arahan, serta masukan yang sangat berarti dalam proses penyusunan laporan ini.
4. Keluarga tercinta, atas segala doa, kasih sayang, dukungan moral maupun materiil yang tidak pernah berhenti mengalir sepanjang perjalanan pendidikan ini.
5. Teman-teman mahasiswa angkatan 2021, yang selalu memberikan semangat, berbagi pengetahuan, dan menemani dalam proses belajar serta menyelesaikan tugas-tugas selama masa studi.
6. Seseorang yang telah setia mendampingi sejak masa SMA hingga saat ini, yang dengan caranya sendiri mampu memberi ketenangan, semangat, dan dukungan yang tulus dalam setiap langkah perjalanan ini.
7. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah membantu dan mendukung kegiatan serta penyusunan laporan ini dengan berbagai kontribusi yang sangat berarti.

Penulis menyadari masih adanya kekurangan dalam penyusunan laporan ini. Oleh karena itu, penulis terbuka terhadap kritik dan saran yang membangun untuk perbaikan ke depan. Semoga laporan ini bermanfaat, baik sebagai pembelajaran pribadi maupun referensi bagi pembaca.



DAFTAR ISI

HALAMAN JUDUL.....	i
SURAT PERNYATAAN	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
MOTTO.....	v
KATA PENGANTAR.....	vi
ABSTRAK	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
DAFTAR RUMUS.....	xiv
LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian	3
1.5 Batasan Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu.....	5
2.2 Analisis Sentimen.....	7
2.3 <i>Text mining</i>	7
2.3.1 <i>Text Preprocessing</i>	8
2.4 TF-IDF	9
2.5 Klasifikasi	10
2.5.1 <i>Support Vector Machine (SVM)</i>	10
2.6 <i>Imbalance Data</i>	12
2.7 <i>Imbalance Data Sampling</i>	12
2.7.1 <i>Borderline-SMOTE</i>	13
2.7.2 <i>SVM-SMOTE</i>	14
2.7.3 <i>ADASYN</i>	15
2.8 <i>K-fold Cross validation</i>	16
2.9 <i>Confusion Matrix</i>	16

2.10 X (<i>Twitter</i>)	18
2.11 Sepak Bola.....	19
BAB III METODE PENELITIAN.....	20
3.1 Tahapan penelitian.....	20
3.2 Pengumpulan Data	20
3.3 Pelabelan Data.....	21
3.4 <i>Preprocessing</i>	22
3.5 Pembobotan TF-IDF	26
3.6 Teknik Penanganan Ketidakseimbangan Data	31
3.6.1 <i>Borderline-SMOTE</i>	31
3.6.2 SVM-SMOTE	33
3.6.3 ADASYN	35
3.7 <i>Support Vector Machine (SVM)</i>	37
3.8 Evaluasi Model.....	41
BAB IV IMPLEMENTASI DAN HASIL	43
4.1 Pengumpulan Data	43
4.2 Pelabelan Data.....	43
4.3 <i>Preprocecing</i> Data.....	45
4.4 Pembobotan TF-IDF	47
4.5 Evaluasi Model Tanpa Penanganan Ketidakseimbangan Data	48
4.6 Evaluasi Model dengan Penanganan Ketidakseimbangan Data	52
4.6.1 <i>Borderline-SMOTE</i>	53
4.6.2 SVM-SMOTE	57
4.6.3 ADASYN	61
4.7 Evaluasi dan Perbandingan Kinerja Model.....	66
BAB V KESIMPULAN DAN SARAN.....	70
5.1 Kesimpulan	70
5.2 Saran.....	71
DAFTAR PUSTAKA	72
LAMPIRAN.....	75

DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian	20
Gambar 3.2 <i>Flowchart Preprocessing</i>	22
Gambar 3.3 Tahapan Perhitungan <i>Borderline-SMOTE</i>	31
Gambar 3.4 Tahapan Perhitungan SVM-SMOTE.....	33
Gambar 3.5 Tahapan Perhitungan ADASYN.....	35
Gambar 4.1 <i>Script</i> Proses <i>Scraping</i>	43
Gambar 4.2 Hasil <i>Scraping</i>	43
Gambar 4.3 <i>Script</i> Proses <i>Cleaning</i>	45
Gambar 4.4 Hasil Data <i>Cleaning</i>	45
Gambar 4.5 <i>Script</i> Proses <i>Case Folding</i>	45
Gambar 4.6 Hasil Data <i>Case Folding</i>	45
Gambar 4.7 <i>Script</i> Proses <i>Tokenizing</i>	46
Gambar 4.8 Hasil Data <i>Tokenizing</i>	46
Gambar 4.9 <i>Script</i> Proses <i>Stopword Removal</i>	46
Gambar 4.10 Hasil Data <i>Stopword Removal</i>	46
Gambar 4.11 <i>Script</i> Proses <i>Stemming</i>	46
Gambar 4.12 Hasil Data <i>Stemming</i>	47
Gambar 4.13 <i>Script</i> Proses TF	47
Gambar 4.14 Hasil Proses TF	47
Gambar 4.15 <i>Script</i> Proses IDF	47
Gambar 4.16 Hasil Proses IDF.....	48
Gambar 4.17 <i>Script</i> Proses TF-IDF	48
Gambar 4.18 Hasil Proses TF-IDF.....	48
Gambar 4.19 Grafik Perbandingan Langkah Uji Pada SVM.....	52
Gambar 4.20 Grafik Perbandingan Langkah Uji Pada <i>Borderline-SMOTE</i>	57
Gambar 4.21 Grafik Perbandingan Langkah Uji Pada SVM-SMOTE	61
Gambar 4.22 Grafik Perbandingan Langkah Uji Pada ADASYN	65
Gambar 4.23 <i>Confusion Matrix</i> Model SVM Dengan ADASYN	68

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	5
Tabel 2.2 <i>Confusion Matriks</i>	17
Tabel 3.1 Hasil <i>Scraping</i> Data	21
Tabel 3.2 Pelabelan	21
Tabel 3.3 <i>Cleaning</i>	23
Tabel 3.4 <i>Case Folding</i>	24
Tabel 3.5 <i>Tokenizing</i>	24
Tabel 3.6 <i>Stopword Removal</i>	25
Tabel 3.7 <i>Stemming</i>	26
Tabel 3.8 Nilai <i>Term Frequency</i>	26
Tabel 3.9 Nilai <i>Document Frequency</i>	27
Tabel 3.10 Hasil <i>Inverse Document Frequency</i>	29
Tabel 3.11 Nilai <i>Term Frequency-Invers Document Frequency</i>	30
Tabel 3.12 Hasil Interpolasi	32
Tabel 3.13 Hasil Akhir SVM-SMOTE.....	34
Tabel 3.14 Hasil Data ADASYN	37
Tabel 3.15 Sampel Data Vektorisasi.....	37
Tabel 3.16 Perhitungan Fungsi Kernel.....	38
Tabel 3.17 Hasil Perhitungan Kernel	38
Tabel 3.18 Hasil Perhitungan Matriks.....	39
Tabel 3.19 Hasil Perhitungan Nilai Error.....	39
Tabel 3.20 Hasil Perhitungan Delta Alpa	40
Tabel 3.21 Hasil Perhitungan Nilai Alpa Baru.....	40
Tabel 3.22 Hasil Perhitungan Sentimen	41
Tabel 4.1 Hasil Pelabelan	44
Tabel 4.2 Hasil Evaluasi Model SVM Tanpa <i>Balancing</i>	49
Tabel 4.3 Hasil Evaluasi Model SVM Dengan <i>Balancing Bordeline-SMOTE</i>	53
Tabel 4.4 Hasil Evaluasi Model SVM Dengan <i>Balancing SVM-SMOTE</i>	58
Tabel 4.5 Hasil Evaluasi Model SVM Dengan <i>Balancing ADASYN</i>	62
Tabel 4.6 Hasil Evaluasi Untuk Setiap Metode Per Label	66

DAFTAR RUMUS

Rumus 2.1 Menentukan IDF	9
Rumus 2.2 Menentukan TF-IDF	10
Rumus 2.3 Elemen Matriks Kernel	11
Rumus 2.4 Perhitungan <i>Error</i>	11
Rumus 2.5 Menghitung <i>Alpha</i>	11
Rumus 2.6 Nilai <i>Alpha</i> Baru	11
Rumus 2.7 Perhitungan Bias	11
Rumus 2.8 Fungsi Keputusan.....	11
Rumus 2.9 Skor Keputusan <i>Linear</i>	12
Rumus 2.10 Fungsi Keputusan dengan <i>Kernel</i>	12
Rumus 2.11 Tetangga Terdekat dari Sampel	13
Rumus 2.12 Cosine <i>Similarity</i>	14
Rumus 2.13 Deteksi <i>Borderline</i> Sampel	14
Rumus 2.14 Pembuatan Sampel Sintetik	14
Rumus 2.15 Fungsi Keputusan SVM.....	14
Rumus 2.16 Sampel Sintetik antara <i>Support Vector</i> dan Tetangga	14
Rumus 2.17 Rasio Ketidakseimbangan.....	15
Rumus 2.18 Bobot Sampel Sintetik	15
Rumus 2.19 Jumlah Total Sampel Sintetik	15
Rumus 2.20 Interpolasi Sintetik <i>ADASYN</i>	15
Rumus 2.21 Pembagagian data <i>K-fold cross validation</i>	16
Rumus 2.22 <i>Accuracy</i>	17
Rumus 2.23 <i>Precision</i>	18
Rumus 2.24 <i>Recall</i>	18
Rumus 2.25 <i>F1-score</i> per kelas	18
Rumus 2.26 <i>F1-score</i> rata-rata	18

LAMPIRAN

Lampiran 1 Data Validasi.....	75
Lampiran 2 Berita Acara Validasi Data.....	75
Lampiran 3 Dokumentasi.....	76

