

**STUDI PERBANDINGAN TEKNIK PENANGANAN DATA
TIDAK SEIMBANG DI DEEP LEARNING**

SKRIPSI

(Jalur Scientist)

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HANIF AL IRSYAD

20.11.3740

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**STUDI PERBANDINGAN TEKNIK PENANGANAN DATA
TIDAK SEIMBANG DI DEEP LEARNING**

SKRIPSI

(Jalur Scientist)

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HANIF AL IRSYAD

20.11.3740

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PERSETUJUAN

SKRIPSI

**STUDI PERBANDINGAN TEKNIK PENANGANAN DATA TIDAK
SEIMBANG DI DEEP LEARNING**

yang disusun dan diajukan oleh

Hanif Al Irsyad

20.11.3740

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Desember 2023

Dosen Pembimbing,



Arif Dwi Laksito, M.Kom

NIK. 190302150

HALAMAN PENGESAHAN
SKRIPSI

STUDI PERBANDINGAN TEKNIK PENANGANAN DATA
TIDAK SEIMBANG DI DEEP LEARNING

yang disusun dan diajukan oleh

Hanif Al Irsyad

20.11.3740

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Desember 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anna Baita, M.Kom

NIK. 190302290

Arifivanto Hadincgoro, S.Kom, MT

NIK. 190302289

Windha Mega Pradnya D, M.Kom

NIK. 190302185



Skrripsi ini telah diterima sebagai salah satu persyaratan
Untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Desember 2023

DEKAN FALKUTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom, M.Kom

NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hanif Al Irsyad
NIM : 20.11.3740

Menyatakan bahwa Skripsi dengan judul berikut:

STUDI PERBANDINGAN TEKNIK PENANGANAN DATA TIDAK SEIMBANG DI DEEP LEARNING

Dosen Pembimbing : Arif Dwi Laksito, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Desember 2023

Yang Menyatakan,

Hanif Al Irsyad

HALAMAN PERSEMBAHAN

Terima kasih kepada Allah SWT atas rahmat dan hidayah-Nya yang senantiasa melimpahkan keberkahan, kesabaran, serta petunjuk sehingga penulis dapat menyelesaikan penelitian ini. Penelitian ini saya persembahkan kepada:

1. Allah SWT yang telah memberikan anugerah, rahmat, hidayah dan inayah sehingga penulis dapat menyelesaikan skripsi ini.
2. Orang tua saya, Bapak Sholikhin dan Ibu Sumarni yang selalu memberikan doa, kasih sayang dan dukungan yang begitu tulus.
3. Saudara saya, Maulana Luthfi Karim dan Faiq Zahir Akmal yang selalu memberikan pengertian dan dukungan.
4. Bapak Arif Dwi Laksito, selaku dosen pembimbing yang telah memberikan banyak saran bimbingan dan masukan sehingga dapat menyelesaikan tugas akhir ini.
5. Teman-teman seperjuangan kelas 20-SIIF-08, terima kasih banyak.
6. Dan segenap keluarga besar Universitas Amikom Yogyakarta terima kasih banyak.
7. Bapak/Ibu dosen pengajar, yang telah memberikan ilmu selama hampir 4 (empat) tahun ini.
8. Serta tanpa mengurangi sedikit hormat, seluruh pihak yang telah banyak membantu yang tidak dapat disebutkan satu persatu, saya ucapkan terima kasih banyak.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas berkat, rahmat dan inayah- Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pemodelan Prediktif untuk Deteksi Kejadian Bencana Menggunakan Support Vector Machine”

Penulis menyampaikan ucapan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Arif Dwi Laksito, M.Kom selaku dosen pembimbing.
3. Keluarga, teman-teman dan berbagai pihak yang telah mendukung hingga terselesaikanya skripsi ini.

Akhir kata, semoga pembuatan skripsi dan penelitian ini dapat bermanfaat bagi pembaca dalam menambah wawasan dan pengetahuan, khususnya dalam bidang kecerdasan buatan.

Yogyakarta, 20 Desember 2023

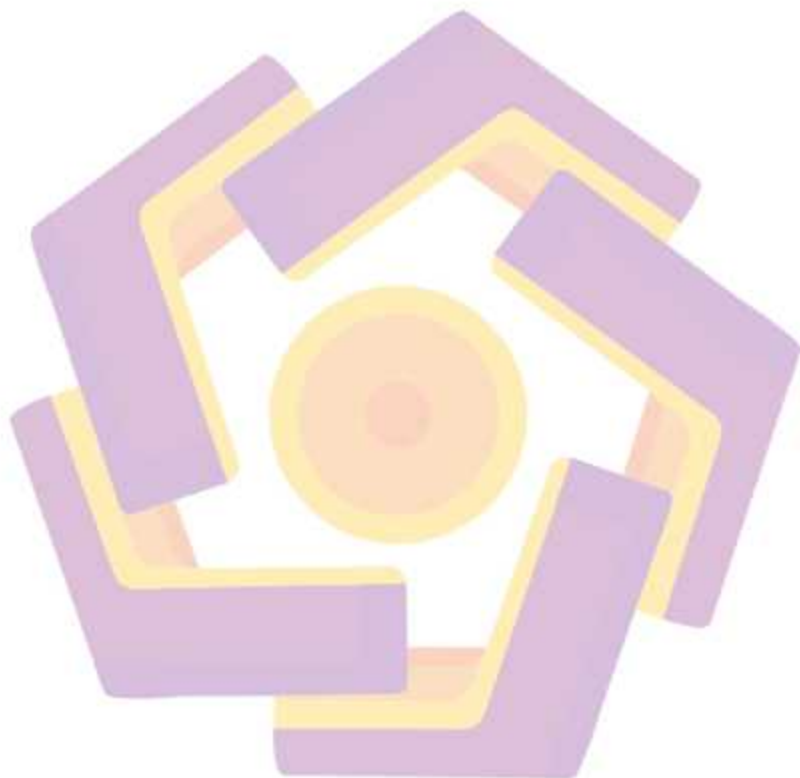
Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN.....	x
INTISARI.....	xi
<i>ABSTRACT</i>	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
BAB II TINJAUAN PUSTAKA	4
2.1 Landasan Teori.....	4
BAB III METODE PENELITIAN	8
3.1 Pra-Pemrosesan Data.....	8
3.2 Pemisahan Dataset.....	9
3.3 Konversi Teks ke Urutan.....	9
3.4 Sequence Padding.....	9
3.5 Pembuatan Matriks Embedding.....	10
3.6 Pemodelan dengan Long Short Term Memory (LSTM).....	10
3.7 Evaluasi Model.....	12
BAB IV HASIL DAN PEMBAHASAN	15
BAB V PENUTUP	18
5.1 Kesimpulan.....	18
REFERENSI	19
LAMPIRAN	22

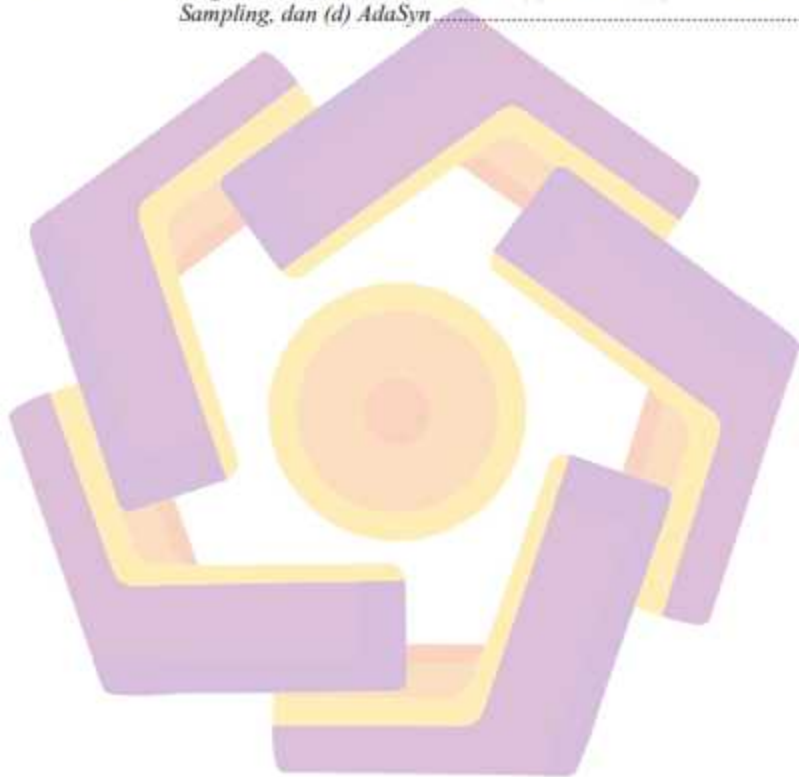
DAFTAR TABEL

<i>Tabel 2.1 Pratinjau Dataset</i>	6
<i>Tabel 3.2 Arsitektur LSTM Model</i>	11
<i>Tabel 4.3 Hasil Presisi, Recall, F1-Score, Akurasi Untuk Semua Metode</i>	15



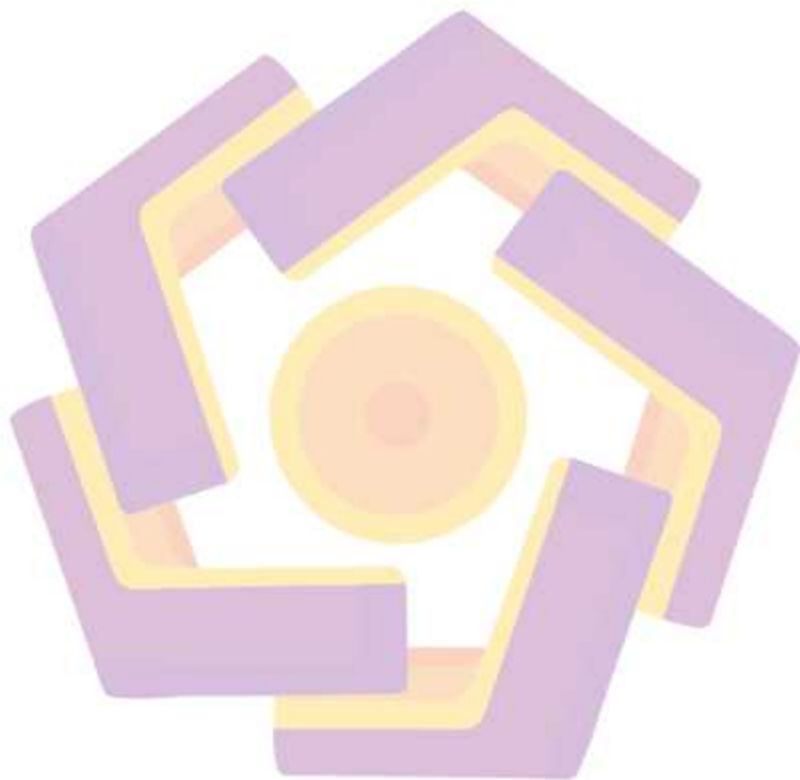
DAFTAR GAMBAR

<i>Gambar 2.1 Distribusi Kelas Dataset</i>	7
<i>Gambar 2.2 Distribusi Kata Dataset</i>	7
<i>Gambar 4.3 Akurasi Pelatihan (a) Dataset Asli, (b) SMOTE, (c) Random Over-Sampling, dan (d) AdaSyn</i>	16
<i>Gambar 4.4 Kerugian Pelatihan (a) Dataset Asli, (b) SMOTE, (c) Random Over-Sampling, dan (d) AdaSyn</i>	17



DAFTAR LAMPIRAN

<i>Lampiran 1.1 Letter of Acceptance</i>	22
<i>Lampiran 1.2 Lembar Review</i>	23
<i>Lampiran 1.3 Sertifikat Penyaji</i>	24
<i>Lampiran 1.4 Sertifikat Author</i>	24



INTISARI

Penelitian ini mengeksplorasi dampak dari tiga metode penyeimbangan data (SMOTE, Random Over-Sampling, dan AdaSyn) terhadap akurasi model LSTM dalam klasifikasi teks tweet bencana.

Penelitian ini bertujuan untuk mengidentifikasi teknik penyeimbangan data yang paling efektif untuk meningkatkan kinerja algoritma LSTM. Penelitian ini membandingkan akurasi model LSTM dengan menggunakan berbagai metode penyeimbangan data dan menentukan pendekatan yang optimal untuk klasifikasi tweet bencana yang tepat.

Kombinasi dari pendekatan-pendekatan ini memungkinkan pemrosesan data berurutan yang efektif, meningkatkan analisis keseluruhan tweet dalam skenario bencana. Algoritma LSTM, tanpa penyeimbangan, mencapai akurasi rata-rata 81,6% di sepuluh pengujian pada data tweet bencana. Ketika dikombinasikan dengan metode penyeimbangan, akurasi berkisar antara 81% hingga 81,1%. Temuan ini menunjukkan kinerja algoritma LSTM yang kuat dalam mengklasifikasikan tweet bencana, dengan sedikit penurunan akurasi ketika menggunakan teknik balancing. Penelitian ini menyimpulkan bahwa menggunakan algoritma LSTM untuk klasifikasi teks tweet bencana tanpa menyeimbangkan dataset akan menghasilkan akurasi yang lebih tinggi. Hal ini menunjukkan bahwa LSTM efektif untuk memproses data teks dalam analisis bencana dan berpotensi untuk digunakan dalam sistem peringatan dini dan manajemen krisis. Namun, penelitian lebih lanjut diperlukan untuk peningkatan algoritma dan peningkatan kinerja.

Kata Kunci: Metode Penyeimbangan, LSTM, Klasifikasi Tweet Bencana, Ketidakseimbangan Data, Pemrosesan Data Sekuensial

ABSTRACT

This research explores the impact of three balancing methods (SMOTE, Random Over-Sampling, and AdaSyn) on LSTM model accuracy in disaster tweet text classification.

It aims to identify the most effective data balancing technique to enhance the LSTM algorithm's performance. The study compares LSTM model accuracy using different balancing methods and determines the optimal approach for precise disaster tweet classification.

The combination of these approaches enables effective processing of sequential data, enhancing the overall analysis of tweets in disaster scenarios. The LSTM algorithm, without balancing, achieved 81.6% average accuracy across ten tests on disaster tweet data. When combined with balancing methods, accuracy ranged from 81% to 81.1%. These findings demonstrate the LSTM algorithm's strong performance in classifying disaster tweets, with a slight accuracy reduction when using balancing techniques. The research concludes that using the LSTM algorithm for text classification of disaster tweets without balancing the dataset achieves higher accuracy. However, further research is required for algorithm enhancement and performance improvement.

Keywords: *Balancing Methods, LSTM, Disaster Tweet Classification, Data Imbalance, Sequential Data Processing*