

**KLASIFIKASI TINGKAT KEPARAHAN KECELAKAAN
KERJA BERBASIS LONG SHORT TERM MEMORY (LSTM)**

SKRIPSI

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



disusun oleh

EVI ZULFIA RAHMAH

21.11.4064

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

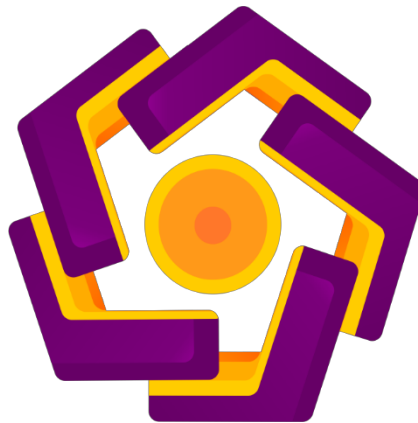
YOGYAKARTA

2025

KLASIFIKASI TINGKAT KEPARAHAN KECELAKAAN KERJA BERBASIS LONG SHORT TERM MEMORY (LSTM)

SKRIPSI

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



disusun oleh

EVI ZULFIA RAHMAH

21.11.4064

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

HALAMAN PERSETUJUAN**SKRIPSI****KLASIFIKASI TINGKAT KEPARAHAN KECELAKAAN KERJA
BERBASIS LONG SHORT TERM MEMORY (LSTM)**

yang disusun dan diajukan oleh

Evi Zulfia Rahmah

21.11.4064

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 Agustus 2025

Dosen Pembimbing,

Dr. Andi Sunyoto, S.Kom., M.Kom.
NIK. 190302052

HALAMAN PENGESAHAN**SKRIPSI****KLASIFIKASI TINGKAT KEPARAHAN KECELAKAAN KERJA
BERBASIS LONG SHORT TERM MEMORY (LSTM)**

yang disusun dan diajukan oleh

Evi Zulfia Rahmah

21.11.4064

Telah dipertahankan di depan Dewan Penguji
pada tanggal 11 Agustus 2025

Susunan Dewan Penguji**Nama Penguji****Tanda Tangan**

Bayu Setiaji, M.Kom.
NIK. 190302216

Yoga Pristyanto, S.Kom., M.Eng.
NIK. 190302412

Dr. Andi Sunyoto, S.Kom., M.Kom.
NIK. 190302052

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 11 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, S.Kom., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Evi Zulfia Rahmah

NIM : 21.11.4064

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Tingkat Keparahan Kecelakaan Kerja Berbasis Long Short Term Memory (LSTM)

Dosen Pembimbing : Bapak Dr. Andi Sunyoto, S.Kom., 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, 29 Juli 2025

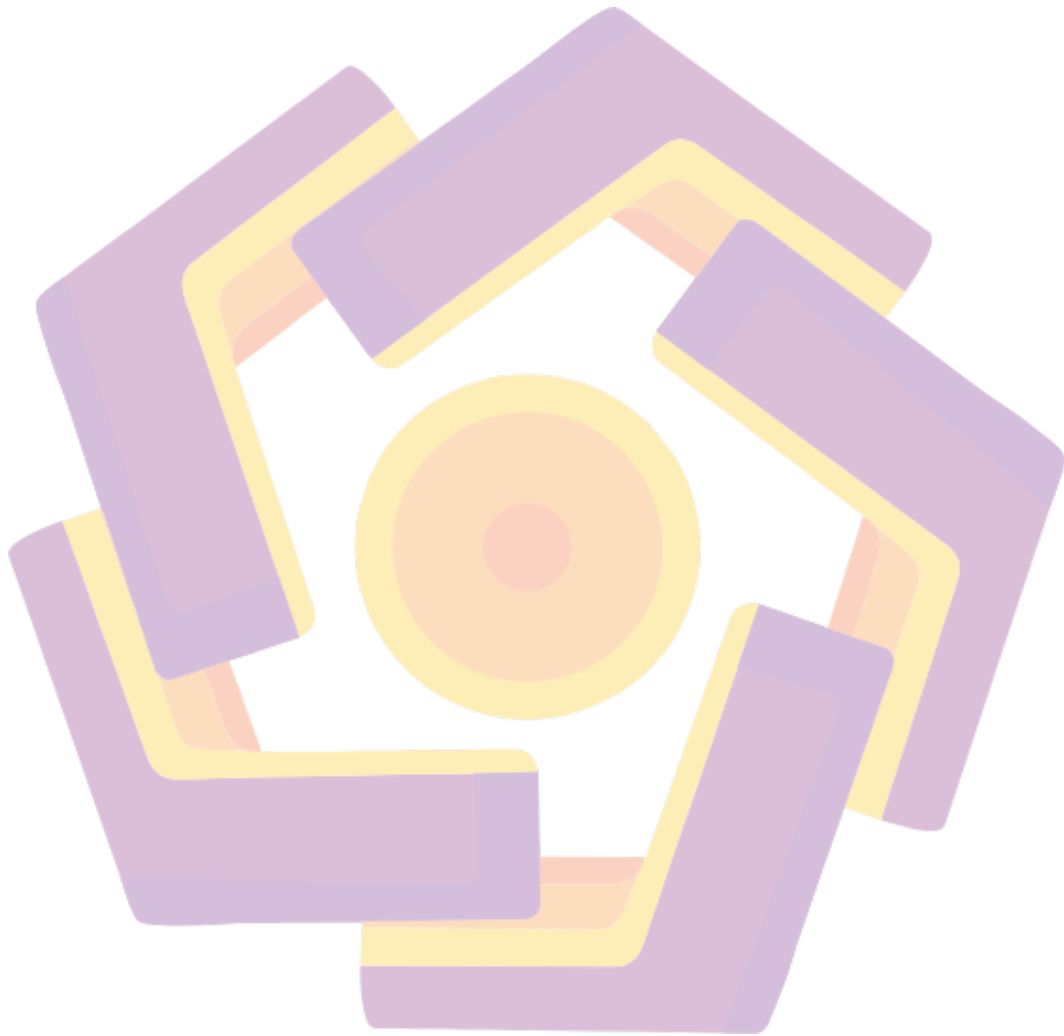
Yang Menyatakan,



Evi Zulfia Rahmah

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan kepada kedua orang tua saya dan semua pihak yang telah membantu dalam penyusunan skripsi ini.



KATA PENGANTAR

Puji syukur saya panjatkan kehadirat Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penyusunan skripsi dengan judul “Klasifikasi Tingkat Keparahen Kecelakaan Kerja Berbasis Long Short Term Memory (LSTM)” ini telah selesai dilaksanakan. Atas selesainya penyusunan penelitian ini, saya ucapkan terimakasih kepada:

1. Bapak Prof. Dr. Suyanto, M.M, selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusrini, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Bapak Dr. Andi Sunyoto, S.Kom., M.Kom., selaku Dosen pembimbing yang sudah memberikan arahan dan bimbingan sehingga penulisan skripsi ini dapat terselesaikan.
5. Semua pihak yang telah membantu dalam penyusunan skripsi, semoga skripsi ini dapat bermanfaat bagi perkembangan ilmu pada umumnya dan khususnya ilmu informatika.

Saya menyadari dalam penulisan laporan skripsi ini masih jauh dari sempurna, maka dari itu saya terbuka dalam kritik dan saran membangun demi perbaikan dikemudian hari

Yogyakarta, 29 Juli 2025

Penulis

DAFTAR ISI

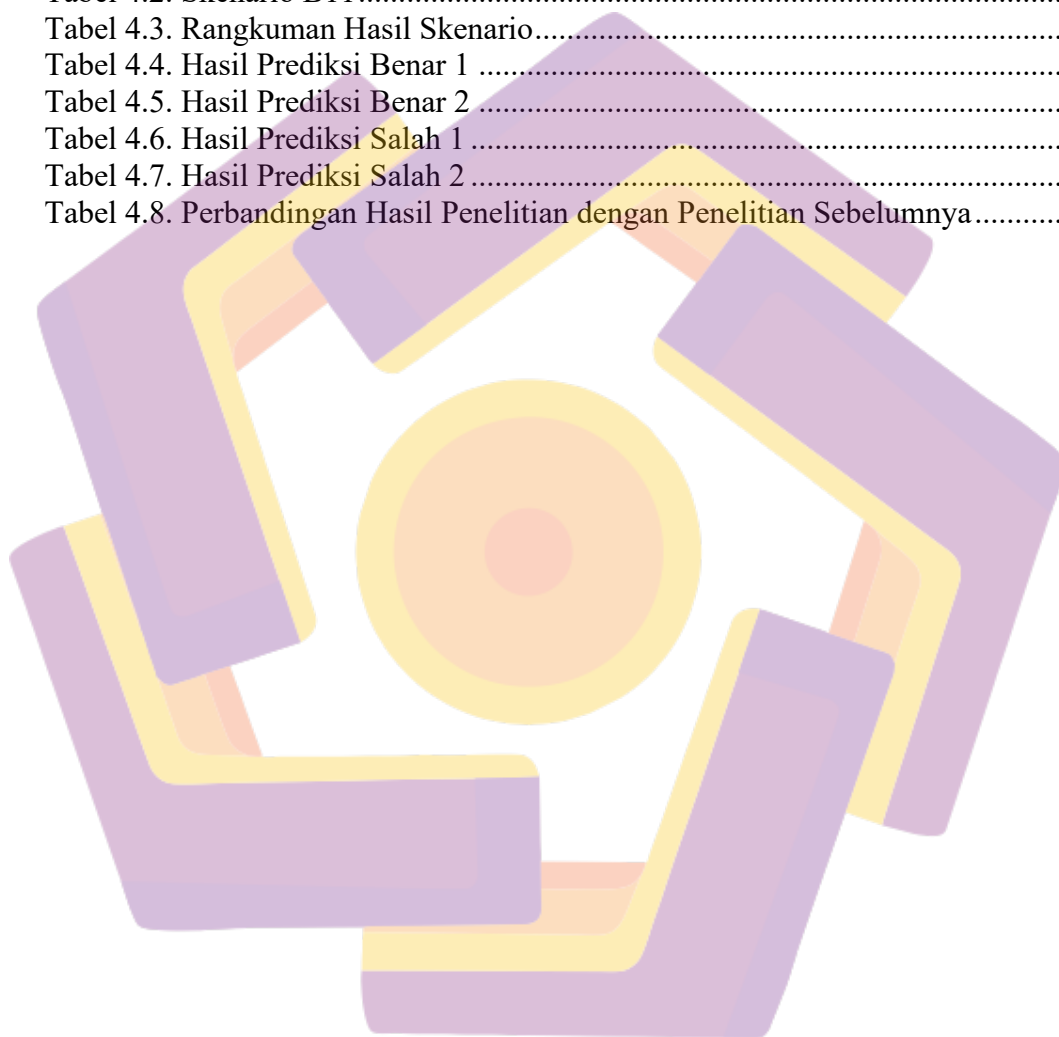
HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	11
2.2.1 Kecelakaan kerja.....	11
2.2.2 Klasifikasi Tingkat Keparahan Kecelakaan Kerja.....	11
2.2.3 Deep Learning	12
2.2.4 Natural Language Processing	13
2.2.5 GloVe	13
2.2.6 Recurrent Neural Network.....	15
2.2.7 Long Short Term Memory.....	16

BAB III METODE PENELITIAN.....	19
3.1 Objek Penelitian.....	19
3.2 Alur Penelitian	19
3.2.1 Pengumpulan Data.....	20
3.2.2 Pelabelan Data	20
3.2.3 Pre-Processing Data.....	21
3.2.4 Tokenisasi dan Padding	22
3.2.5 Pembagian Data.....	23
3.2.6 GloVe Embedding (Pembobotan Kata).....	23
3.2.7 Encoding Label dan One-Hot Encoding.....	24
3.2.8 Penyeimbangan Data	25
3.2.9 Pembangunan dan Pelatihan Model	26
3.2.10 Evaluasi Model.....	27
3.2.11 Representasi Hasil	28
3.3 Alat dan Bahan.....	28
3.3.1 Data Penelitian.....	28
3.3.2 Instrumen Penelitian	29
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Dataset dan Hasil Pelabelan.....	31
4.2 Analisis Panjang Teks.....	31
4.3 Arsitektur dan Pelatihan Model	32
4.3.1. Arsitektur Model	33
4.3.2 Proses Pelatihan Model	34
4.4. Hasil Evaluasi dan Analisis Model pada Setiap Skenario	35
4.4.1 Skenario A1	35
4.4.2 Skenario A2	39
4.4.3 Skenario A3	43
4.4.4 Skenario A4	46
4.4.5 Skenario A5	49
4.4.6 Skenario B6	52
4.4.7 Skenario B7	55
4.4.8 Skenario B8	58

4.4.9 Skenario B9	61
4.4.10 Skenario B10	64
4.4.11 Skenario B11	67
4.5 Pemilihan dan Pembahasan Skenario Terbaik	71
4.5.1. Pemilihan Skenario Terbaik	71
4.5.2 Pembahasan dan Alasan Pemilihan	73
4.5.3 Implikasi Pemilihan Skenario Terbaik	73
4.6 Analisis Prediksi Benar dan Salah pada Skenario Terbaik	74
4.6.1 Analisis Prediksi Benar	74
4.6.2 Analisis Prediksi Salah	76
4.7 Perbandingan Hasil Penelitian dengan Penelitian Sebelumnya	78
BAB V PENUTUP	80
5.1 Kesimpulan	80
5.2 Saran	80
REFERENSI	82

DAFTAR TABEL

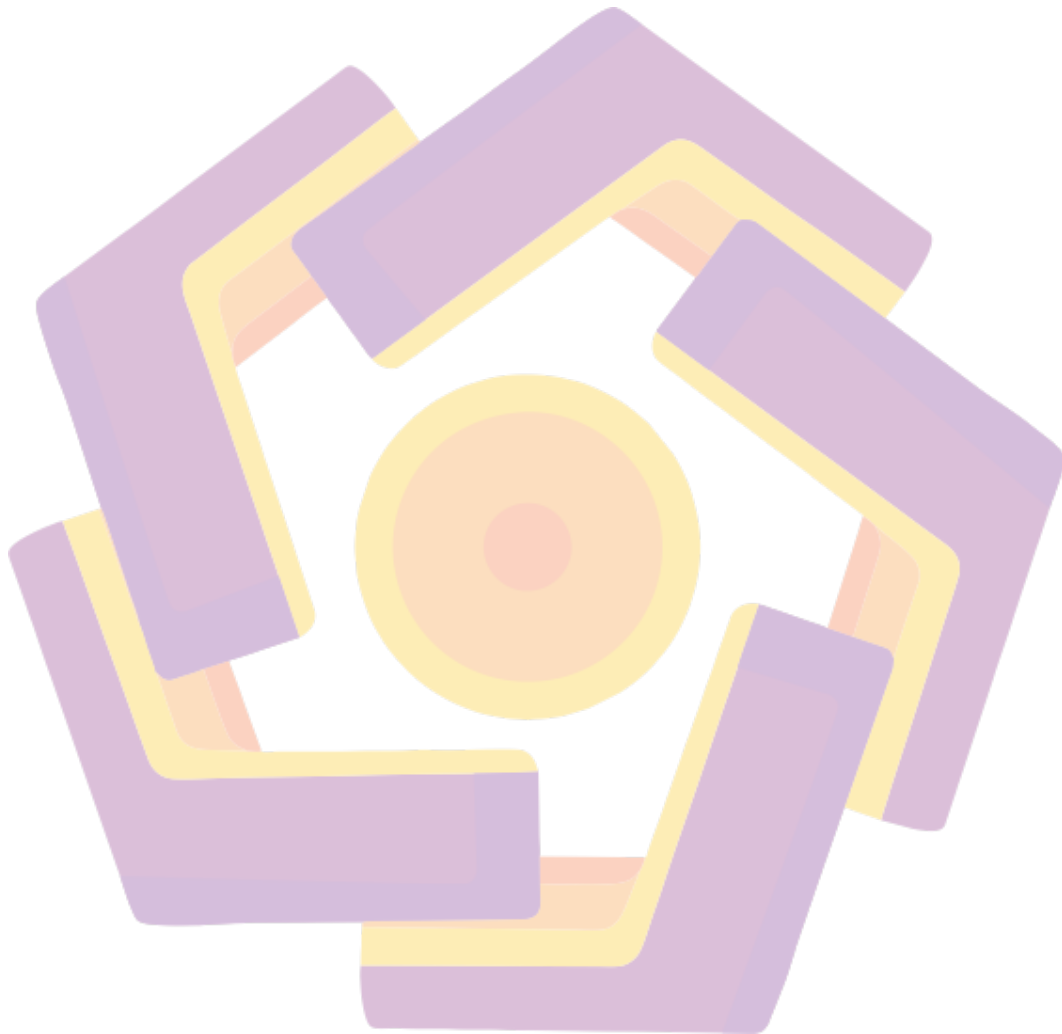
Tabel 2.1 Tabel Keaslian Peneliti.....	8
Tabel 3.1 Konfigurasi Model pada Setiap Skenario Eksperimen.....	27
Tabel 3.1 Spesifikasi Komputer	29
Tabel 4.1. Distribusi Label Kelas	31
Tabel 4.2. Skenario B11	67
Tabel 4.3. Rangkuman Hasil Skenario.....	71
Tabel 4.4. Hasil Prediksi Benar 1	74
Tabel 4.5. Hasil Prediksi Benar 2	75
Tabel 4.6. Hasil Prediksi Salah 1	76
Tabel 4.7. Hasil Prediksi Salah 2	77
Tabel 4.8. Perbandingan Hasil Penelitian dengan Penelitian Sebelumnya.....	78



DAFTAR GAMBAR

Gambar 2.1. Ilustrasi Forget Gate	16
Gambar 2.2. Ilustrasi Input Gate	17
Gambar 2.3. Ilustrasi Cell State Update	17
Gambar 2.4. Ilustrasi Output Gate	18
Gambar 3.1. Gambar Alur Penelitian	19
Gambar 4.1. Distribusi Panjang Teks Sebelum Padding	32
Gambar 4.2. Grafik Akurasi Model per Epoch Skenario A1	35
Gambar 4.3. Grafik Loss Model per Epoch Skenario A1	36
Gambar 4.4. Confusion Matrix dan Classification Report Skenario A1	37
Gambar 4.5. Grafik Receiver Operating Characteristic (ROC) Curve A1	38
Gambar 4.6. Grafik Akurasi Model per Epoch Skenario A2	39
Gambar 4.7. Grafik Loss Model per Epoch Skenario A2	40
Gambar 4.8. Confusion Matrix dan Classification Report Skenario A2	41
Gambar 4.9. Grafik Receiver Operating Characteristic (ROC) Curve A2	42
Gambar 4.10. Grafik Akurasi Model per Epoch Skenario A3	43
Gambar 4.11. Grafik Loss Model per Epoch Skenario A3	43
Gambar 4.12. Confusion Matrix dan Classification Report Skenario A3	44
Gambar 4.13. Grafik Receiver Operating Characteristic (ROC) Curve A3	45
Gambar 4.14. Grafik Akurasi Model per Epoch Skenario A4	46
Gambar 4.15. Grafik Loss Model per Epoch Skenario A4	46
Gambar 4.16. Confusion Matrix dan Classification Report Skenario A4	47
Gambar 4.17. Grafik Receiver Operating Characteristic (ROC) Curve A4	48
Gambar 4.18. Grafik Akurasi Model per Epoch Skenario A5	49
Gambar 4.19. Grafik Loss Model per Epoch Skenario A5	50
Gambar 4.20. Confusion Matrix dan Classification Report Skenario A5	50
Gambar 4.21. Grafik Receiver Operating Characteristic (ROC) Curve A5	51
Gambar 4.22. Grafik Akurasi Model per Epoch Skenario B6	52
Gambar 4.23. Grafik Loss Model per Epoch Skenario B6	53
Gambar 4.24. Confusion Matrix dan Classification Report Skenario B6	53
Gambar 4.25. Grafik Receiver Operating Characteristic (ROC) Curve	54
Gambar 4.26. Grafik Akurasi Model per Epoch Skenario B7	55
Gambar 4.27. Grafik Loss Model per Epoch Skenario B7	56
Gambar 4.28. Confusion Matrix dan Classification Report Skenario B7	56
Gambar 4.29. Grafik Receiver Operating Characteristic (ROC) Curve B7	57
Gambar 4.30. Grafik Akurasi Model per Epoch Skenario B8	58
Gambar 4.31. Grafik Loss Model per Epoch Skenario B8	58
Gambar 4.32. Confusion Matrix dan Classification Report Skenario B8	59
Gambar 4.33. Grafik Receiver Operating Characteristic (ROC) Curve B8	60
Gambar 4.34. Grafik Akurasi Model per Epoch Skenario B9	61
Gambar 4.35. Grafik Loss Model per Epoch Skenario B9	61
Gambar 4.36. Confusion Matrix dan Classification Report Skenario B9	62
Gambar 4.37. Grafik Receiver Operating Characteristic (ROC) Curve B9	63
Gambar 4.38. Grafik Akurasi Model per Epoch Skenario B10	64

Gambar 4.39. Grafik Loss Model per Epoch Skenario B10	64
Gambar 4.40. Confusion Matrix dan Classification Report Skenario B10.....	65
Gambar 4.41. Grafik Receiver Operating Characteristic (ROC) Curve B10.....	66
Gambar 4.42. Grafik Akurasi Model per Epoch Skenario B11	68
Gambar 4.43. Grafik Loss Model per Epoch Skenario B11	68
Gambar 4.44. Confusion Matrix dan Classification Report Skenario B11.....	69
Gambar 4.45. Grafik Receiver Operating Characteristic (ROC) Curve B11.....	70
Gambar 4.46. Grafik Akurasi Per Skenario	72

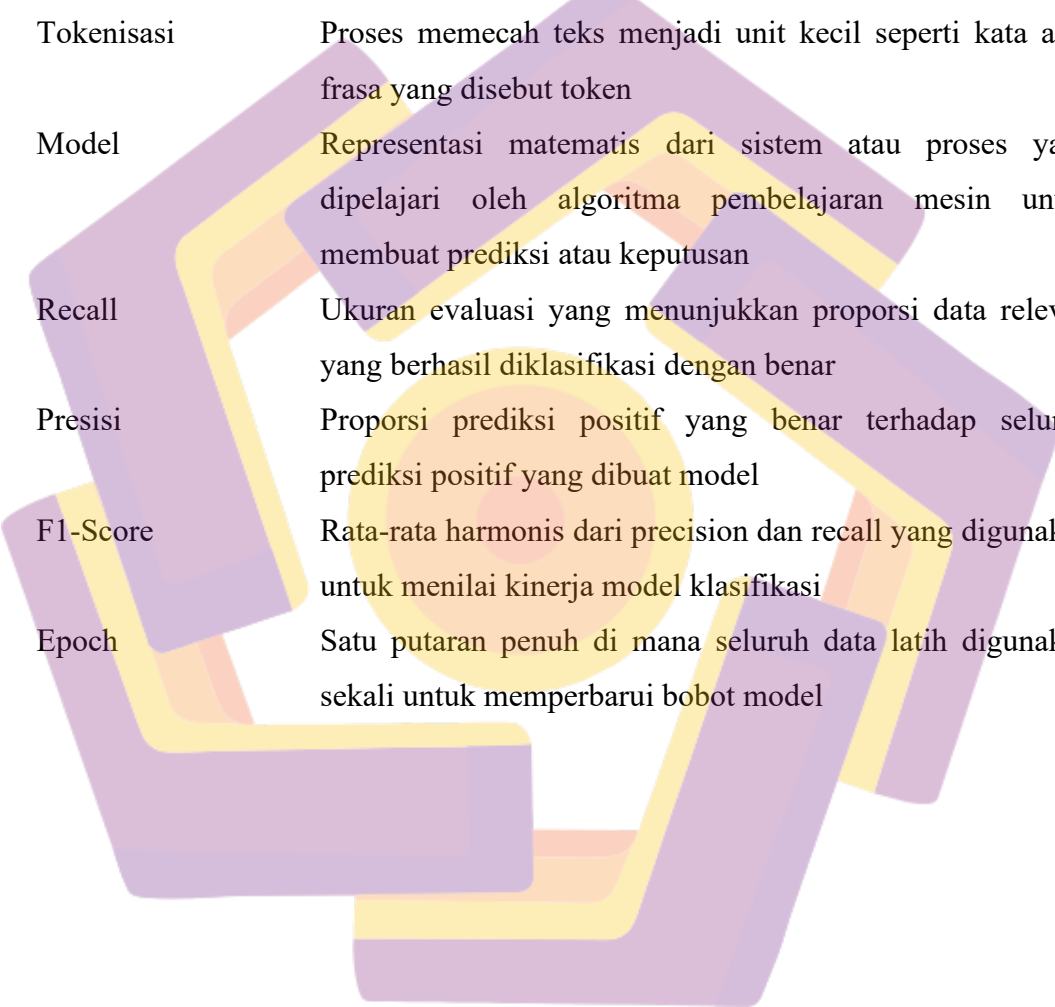


DAFTAR LAMBANG DAN SINGKATAN



LSTM	Long Short-Term Memory
OSHA	Occupational Safety and Health Administration
ILO	International Labour Organization
SMOTE	Synthetic Minority Oversampling Technique
GloVe	Global Vectors for Word Representation
HIRA	Hazard Identification and Risk Assessment
RNN	Recurrent Neural Network
ITE	Informasi dan Transaksi Elektronik
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
C-LSTM	Convolutional Long Short-Term Memory
CNN	Convolutional Neural Network
NER	Named Entity Recognition
FOM	Field Operations Manual
NLP	Natural Language Processing
NLU	Natural Language Understanding
NLG	Natural Language Generation
LSA	Latent Semantic Analysis
RNN	Recurrent Neural Networks
GPU	Graphics Processing Unit

DAFTAR ISTILAH



Memory Cell	unit penyimpanan internal dalam LSTM untuk menyimpan dan mempertahankan informasi jangka panjang selama pemrosesan.
Sigmoid	fungsi aktivasi yang memetakan input dengan nilai 0-1
Tokenisasi	Proses memecah teks menjadi unit kecil seperti kata atau frasa yang disebut token
Model	Representasi matematis dari sistem atau proses yang dipelajari oleh algoritma pembelajaran mesin untuk membuat prediksi atau keputusan
Recall	Ukuran evaluasi yang menunjukkan proporsi data relevan yang berhasil diklasifikasi dengan benar
Presisi	Proporsi prediksi positif yang benar terhadap seluruh prediksi positif yang dibuat model
F1-Score	Rata-rata harmonis dari precision dan recall yang digunakan untuk menilai kinerja model klasifikasi
Epoch	Satu putaran penuh di mana seluruh data latih digunakan sekali untuk memperbarui bobot model

INTISARI

Kecelakaan kerja merupakan salah satu permasalahan utama dalam dunia industri karena berdampak serius terhadap keselamatan tenaga kerja dan efisiensi operasional perusahaan. Dalam banyak kasus, informasi terkait tingkat keparahan kecelakaan kerja tidak tersedia secara eksplisit dalam laporan insiden, khususnya yang berbentuk deskripsi naratif. Hal ini menyulitkan pihak manajemen dan ahli keselamatan dalam menentukan prioritas penanganan insiden. Penelitian ini bertujuan untuk membangun model klasifikasi otomatis untuk mengidentifikasi tingkat keparahan kecelakaan kerja berdasarkan deskripsi naratif menggunakan pendekatan Long Short-Term Memory (LSTM). Dataset yang digunakan diperoleh dari OSHA (*Occupational Safety and Health Administration*) dan terdiri atas 91147 entri laporan kecelakaan kerja dalam bahasa Inggris. Proses pelabelan dilakukan secara bertahap dengan kombinasi metode *zero-shot classification*, serta verifikasi manual untuk memastikan kualitas label. Selanjutnya, dilakukan pra pemrosesan teks, tokenisasi, serta pembobotan kata menggunakan GloVe pre-trained embedding. Data dibagi menjadi data latih dan data uji dengan perbandingan 80:20 dan disesuaikan secara stratifikasi. Model LSTM dibangun dengan arsitektur berlapis dan dilatih selama 35 epoch menggunakan class weight untuk mengatasi ketidakseimbangan data. Hasil evaluasi menunjukkan bahwa model mampu mencapai akurasi sebesar 97% dengan nilai precision, recall, dan F1-score yang seimbang pada tiap kelas.

Kata kunci: kecelakaan kerja, LSTM, klasifikasi, teks naratif, keselamatan kerja

ABSTRACT

Workplace accidents remain a critical issue in industrial sectors due to their significant impact on employee safety and operational efficiency. In many incident reports, especially those in narrative form, explicit information regarding the severity level of each case is often unavailable, making it difficult for safety officers and management to prioritize appropriate responses. This study aims to develop an automated classification model to predict the severity level of workplace accidents based on narrative descriptions using the Long Short-Term Memory (LSTM) method. The dataset used was obtained from the Occupational Safety and Health Administration (OSHA), consisting of 91147 English-language incident reports. The data labeling process was carried out in stages using zero-shot classification with a pre-trained language model, and manual verification to ensure label accuracy. Text preprocessing included normalization, stopword removal, lemmatization, tokenization, and word representation using GloVe embeddings. The dataset was divided into training and testing sets with an 80:20 ratio, stratified based on class distribution. A multi-layered LSTM architecture was trained for 35 epochs using class weighting to address data imbalance. Evaluation results show that the model achieved an accuracy of 97%, with balanced precision, recall, and F1-score across the three classes.

Keyword: workplace accidents, LSTM, classification, narrative text, occupational safety