

**STUDI PERBANDINGAN ALGORITMA MACHINE
LEARNING UNTUK KLASIFIKASI SENTIMEN DALAM
KONTEKS DUNIA KERJA GEN Z**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

MUHAMMAD FATHIR RUMADAUL
21.11.4235

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

**STUDI PERBANDINGAN ALGORITMA MACHINE
LEARNING UNTUK KLASIFIKASI SENTIMEN DALAM
KONTEKS DUNIA KERJA GEN Z**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

MUHAMMAD FATHIR RUMADAUL

21.11.4235

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

STUDI PERBANDINGAN ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI SENTIMEN DALAM KONTEKS DUNIA KERJA GEN Z

yang disusun dan diajukan oleh

MUHAMMAD FATHIR RUMADAUL

21.11.4235

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 30 Juli 2025

Dosen Pembimbing,


Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248

HALAMAN PENGESAHAN
SKRIPSI
STUDI PERBANDINGAN ALGORITMA MACHINE LEARNING UNTUK
KLASIFIKASI SENTIMEN DALAM KONTEKS DUNIA KERJA GEN Z

yang disusun dan diajukan oleh

MUHAMMAD FATHIR RUMADAUL

21.11.4235

Telah dipertahankan di depan Dewan Penguji
pada tanggal 30 Juli 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dwi Nurani, S.Kom., M.Kom.
NIK. 190302236



Mujiyanto, M.Kom
NIK. 190302492



Mulia Sulistiyono, S.Kom., M.Kom
NIK. 190302248



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : MUHAMMAD FATHIR RUMADAUL
NIM : 21.11.4235

Menyatakan bahwa Skripsi dengan judul berikut:

Studi Perbandingan Algoritma *Machine Learning* Untuk Klasifikasi Sentimen Dalam Konteks Dunia Kerja Gen Z

Dosen Pembimbing : Mulia Sulistiyono, 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, 30 Juli 2025

Yang Menyatakan,



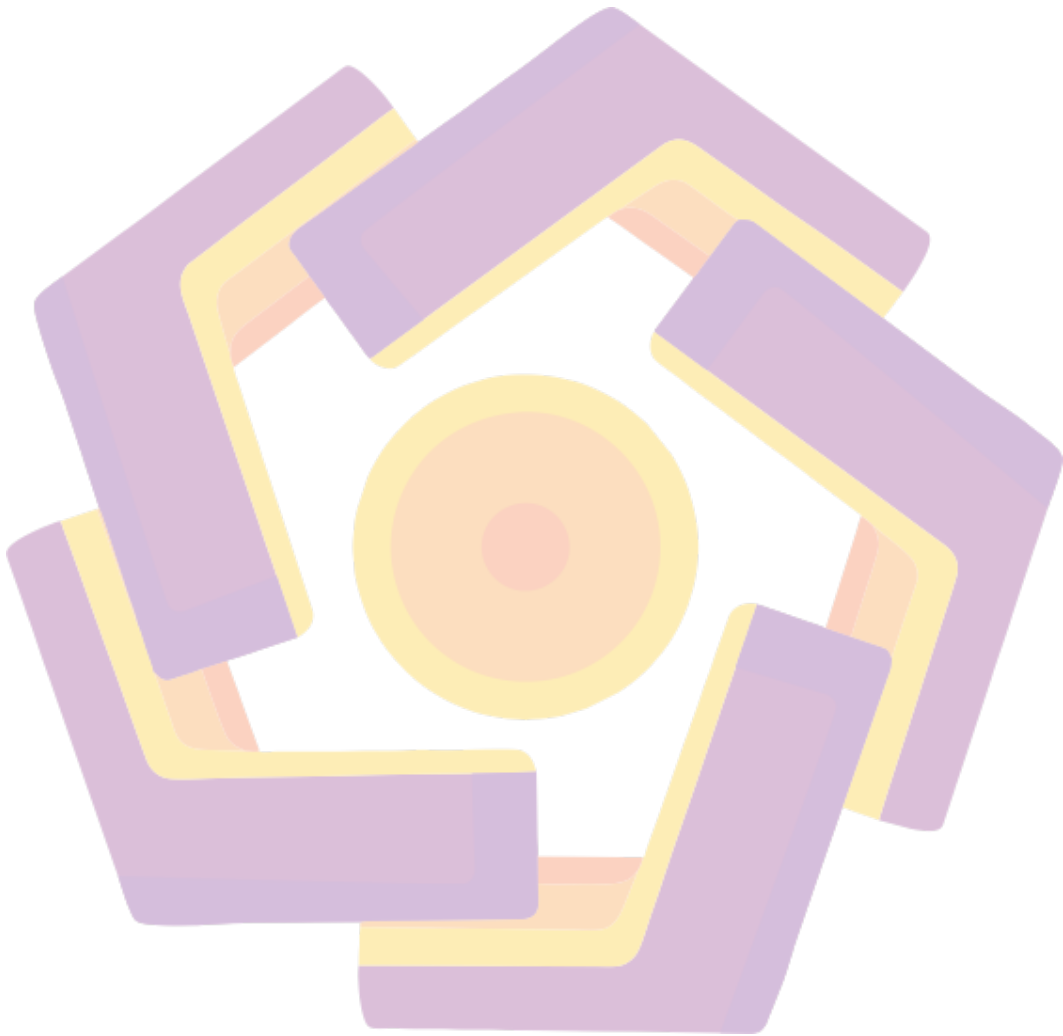
(Muhammad Fathir Rumadaul)

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam atas kehadiran Allah SWT, yang telah memberikan hidayah-Nya atas kekuatan dan kesehatan penulis, sehingga skripsi ini dapat terselesaikan dengan baik dan tanpa terkendala, demikian skripsi ini juga penulis persembahkan kepada:

1. Yuyun Yuniarti Layn dan Rustam Hidayat Rumadaul, selaku orang tua terkasih dan tercinta, terima kasih atas doa, dukungan dan bimbingan yang tiada henti, serta kasih sayang yang tulus. Segala pengorbanan terlihat maupun tak terlihat yang kalian berikan dalam setiap kisah hidup ini, semoga kalian senantiasa sehat, panjang umur, dilancarkan segala rezeki dan selalu berbahagia.
2. Keluarga besar tersayang, yang senantiasa memberikan semangat, dukungan dan juga doa restu dalam setiap langkah perjalanan ini.
3. Keluarga pengurus Himpunan Mahasiswa Informatika, terima kasih atas segala kebersamaan dan kenangan yang sangat berarti.
4. Teman-teman kelas, terima kasih atas segala kebersamaan dan kenangan yang telah dijalani bersama.
5. Fisfrydania Tiara Rahmawati, Terima kasih atas setiap dorongan semangat dan dukungan yang diberikan agar pengerjaan skripsi ini dapat terus berjalan sampai selesai.
6. Terima kasih kepada diri saya sendiri yang mampu selalu bertahan dalam kondisi dan situasi apapun, sehingga dapat menyelesaikan skripsi ini dengan baik.
7. Almamater tercinta, Universitas AMIKOM Yogyakarta, tempat dimana cerita perjuangan ini bermula, terima kasih atas segala ilmu yang bermanfaat dalam menimba ilmu, pembentukan karakter, serta pengembangan potensi diri.

Skripsi ini adalah bukti perjuangan melalui ketekunan, doa dan semangat pantang menyerah dalam awal perjalanan menuju kesuksesan. Semoga skripsi ini dapat memberikan manfaat, baik bagi penulis maupun bagi para pembaca di masa yang akan datang.



KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan skripsi yang berjudul **“STUDI PERBANDINGAN ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI SENTIMEN DALAM KONTEKS DUNIA KERJA GEN Z”**. Skripsi ini disusun sebagai salah satu syarat untuk meraih gelar sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, mengingat keterbatasan kemampuan, pengetahuan, dan pengalaman penulis. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun agar karya tulis berikutnya dapat lebih baik lagi.

Proses penyusunan skripsi ini tidak lepas dari berbagai tantangan, namun berkat bantuan, bimbingan, dan kerja sama dari banyak pihak, penulis akhirnya dapat menyelesaikannya dengan baik. Oleh karena itu penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom, selaku Dekan Fakultas Ilmu Komputer di Universitas AMIKOM Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom, selaku Ketua Program Studi Informatika di Universitas AMIKOM Yogyakarta
4. Bapak Mulia Sulistiyono, M.Kom, selaku Dosen Pembimbing yang dengan sabar dan ikhlas meluangkan waktu, tenaga, dan pikiran, serta memberikan arahan, motivasi dan saran-s
5. aran yang sangat berharga kepada penulis selama proses penyusunan skripsi ini.
6. Bapak Nuri Cahyono, M.Kom, selaku Dosen Wali yang telah memberikan bimbingan dan bantuan selama penulis menempuh masa studi di Universitas AMIKOM Yogyakarta.

7. Seluruh Staff Dosen Universitas AMIKOM Yogyakarta, khususnya di Fakultas Ilmu Komputer, Program Studi Informatika yang telah memberikan ilmu dan bimbingan yang tidak ternilai harganya, serta masukan, pemikiran dan juga tenaga selama penulis menempuh pendidikan sarjana.
8. Semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan dukungan kepada penulis.

Semoga Allah SWT memberikan balasan yang berlipat ganda kepada kalian semua yang senantiasa membantu dan menemani penulis dalam pembuatan skripsi ini. Penulis berharap skripsi ini dapat memberikan manfaat khususnya bagi penulis dan juga para pembaca.

Yogyakarta, 30 Juli 2025

Muhammad Fathir Rumadaul

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
DAFTAR LAMBANG DAN SINGKATAN.....	xv
DAFTAR ISTILAH.....	xvi
INTISARI.....	xvii
ABSTRACT.....	xviii
BAB I	
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II	
TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur.....	6
2.2 Dasar Teori.....	12
2.2.1 Media Sosial.....	12
2.2.2 Web Scraping.....	12
2.2.3 Analisis Sentimen.....	12
2.2.4 Data Mining.....	13
2.2.5 Text Mining.....	13
2.2.6 Preprocessing Data.....	13
2.2.7 Term Frequency-Inverse Document Frequency (TF-IDF).....	14
2.2.8 Bert Tokenization and Encoding.....	15
2.2.9 Random Oversampling (ROS).....	16
2.2.10 Support Vector Machine.....	16
2.2.11 Random Forest.....	16
2.2.12 Bidirectional Encoder Representations from Transformers (BERT).....	16

17	
2.2.13 Google Collaboratory.....	18
2.2.14 Python.....	18

BAB III

METODE PENELITIAN..... 19

3.1 Objek Penelitian.....	19
3.2 Alur Penelitian.....	19
3.2.1 Crawling Data.....	20
3.2.2 Preprocessing Data.....	20
3.2.2 Pelabelan Data.....	21
3.2.4 Splitting Data.....	21
3.2.5 TF-IDF (Term Frequency-Inverse Document Frequency).....	22
3.2.6 Bert Tokenization and Encoding.....	22
3.2.7 Random Oversampling (ROS).....	22
3.2.8 Implementasi Model.....	23
3.2.9 Evaluasi.....	23
3.3 Alat dan Bahan.....	23
3.3.1 Alat Penelitian.....	23
3.3.2 Bahan Penelitian.....	24

BAB IV

HASIL DAN PEMBAHASAN..... 25

4.1 Pengumpulan Data.....	25
4.1.1 Crawling Data.....	25
4.2 Preprocessing Data.....	26
4.2.1 Case Folding & Cleaning Data.....	26
4.2.2 Normalisasi.....	28
4.2.3 Tokenisasi & Stopword Removal.....	29
4.2.4 Stemming.....	30
4.3 Pelabelan Data.....	31
4.4 TF-IDF (Term Frequency-Inverse Document Frequency).....	32
4.5 Splitting Data.....	33
4.6 Bert Tokenization and Encoding.....	34
4.7 Random OverSampling.....	34
4.8 Implementasi Model.....	35
4.8.1 Support Vector Machine.....	35
4.8.2 Random Forest.....	36
4.8.3 Bidirectional Encoder Representations from Transformers.....	37
4.9 Evaluasi.....	39

4.9.1 Support Vector Machine & Random Forest Confusion Matrix.....	39
4.9.1.1 Model Support Vector Machine (SVM).....	40
4.9.1.2 Model Random Forest (RF).....	40
4.9.2 BERT Confusion Matrix.....	41
4.9.3 Evaluasi Hasil Klasifikasi Confusion Matrix.....	42
4.9.3.1 Support Vector Machine (SVM).....	42
4.9.3.2 Random Forest (RF).....	42
4.9.3.3 Bidirectional Encoder Representation from Transformers (BERT).....	43
4.9.4 Analisa Risiko Overfitting.....	43
BAB V	
PENUTUP.....	44
5.1 Kesimpulan.....	44
5.2 Saran.....	45
REFERENSI.....	46
LAMPIRAN.....	48

DAFTAR TABEL

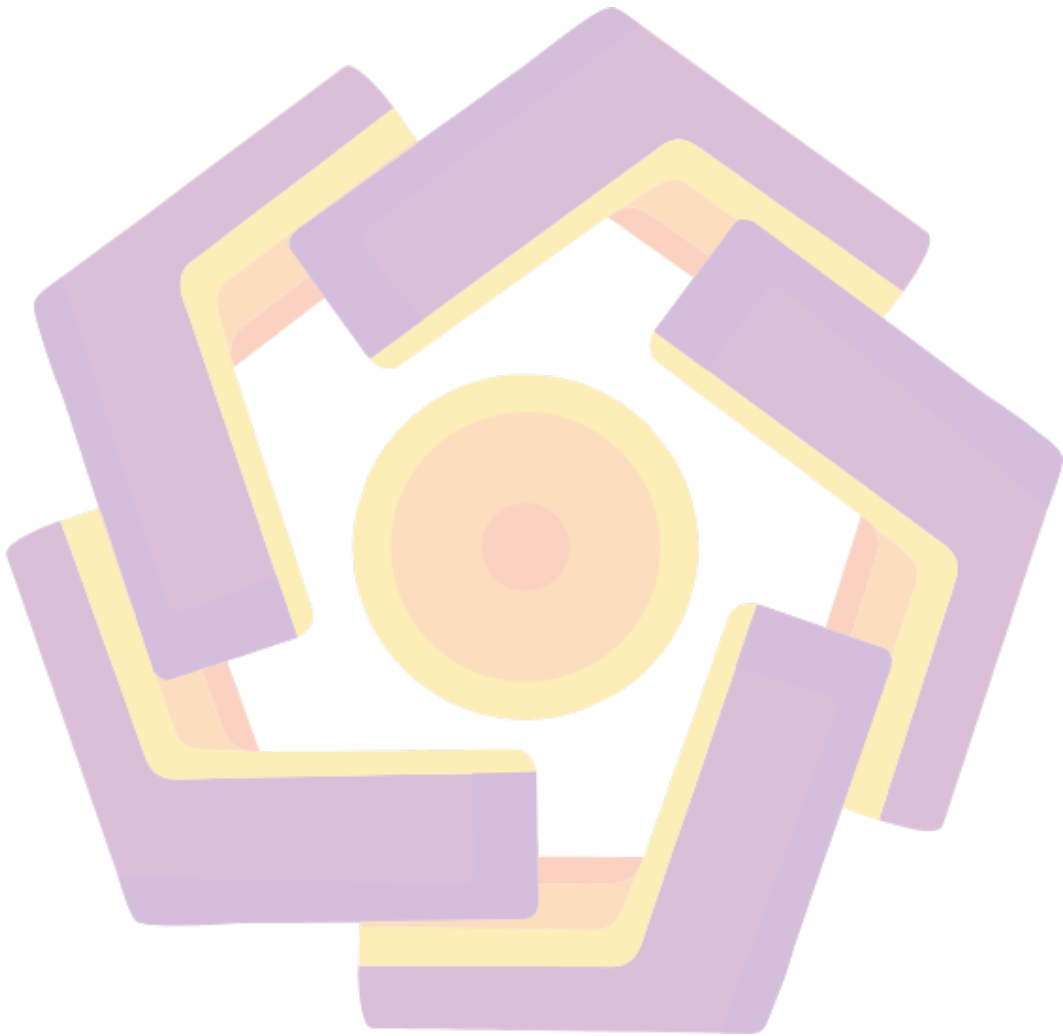
Tabel 2.1 Keaslian Penelitian	8
Tabel 4.1 Hasil Pengambilan Data	25
Tabel 4.2 Hasil setelah Case Folding & Cleaning Data	27
Tabel 4.3 Hasil sesudah Normalisasi	28
Tabel 4.4 Hasil sesudah Tokenisasi & Stopword Removal	29
Tabel 4.5 Hasil sesudah Stemming	30
Tabel 4.6 Jumlah Distribusi Data pada Masing-Masing Label	32
Tabel 4.7 Hasil Ulasan Sesudah Labeling	32
Tabel 4.8 Skor TF-IDF Tertinggi	33
Tabel 4.9 Hasil Tokenisasi dan Encoding BERT	34
Tabel 4.10 Hasil Klasifikasi SVM	36
Tabel 4.11 Hasil Klasifikasi Random Forest	37
Tabel 4.12 Hasil pada setiap Epoch	38
Tabel 4.13 Hasil Klasifikasi BERT	39

DAFTAR GAMBAR

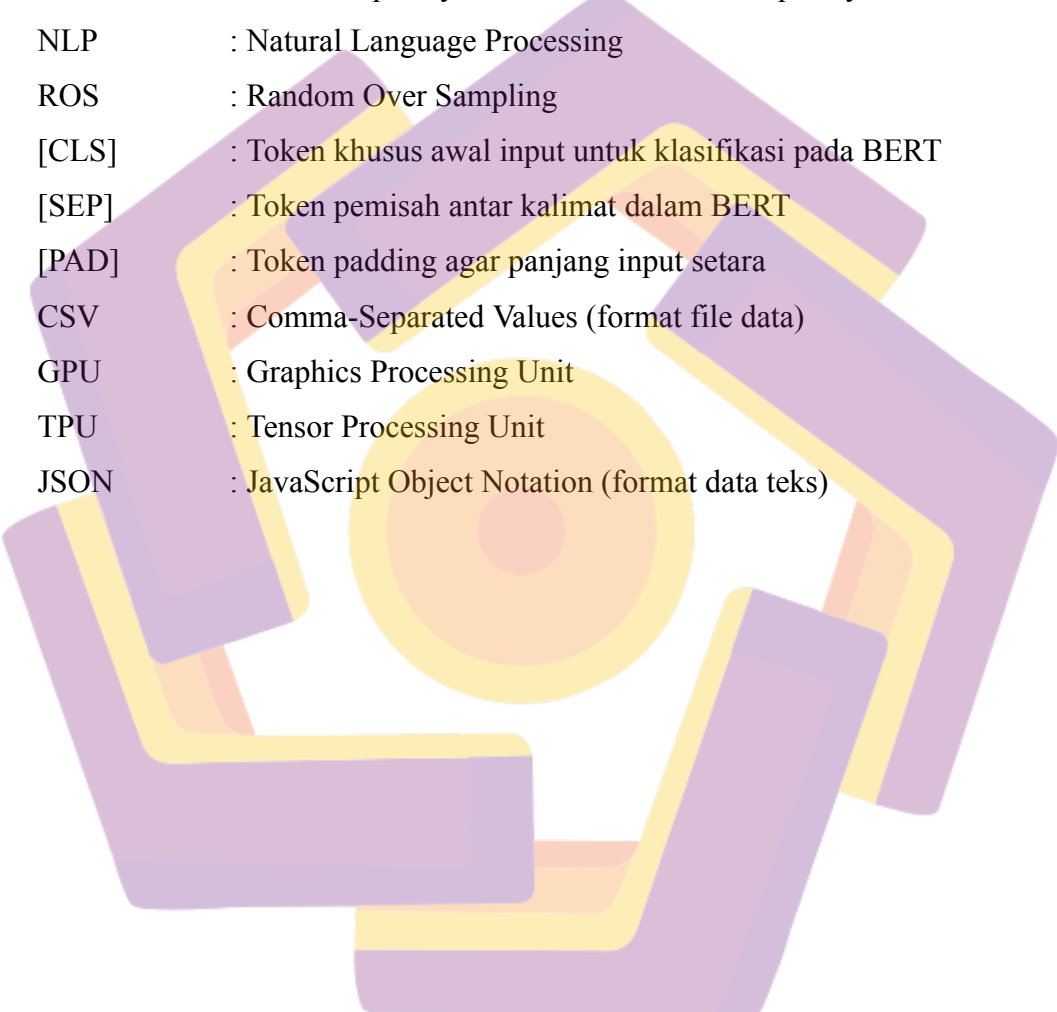
Gambar 2.1 Ilustrasi pre-training dan fine-tuning BERT	17
Gambar 3.1 Alur Penelitian	19
Gambar 4.1 Source Code Crawling Data	25
Gambar 4.2 Source Code Case Folding dan Cleaning Data	26
Gambar 4.3 Source Code Normalisasi	28
Gambar 4.4 Source Code Tokenisasi & Stopword Removal	29
Gambar 4.5 Source Code Stemming	30
Gambar 4.6 Source Code untuk Labeling Data	31
Gambar 4.7 Source Code untuk TF-IDF	33
Gambar 4.8 Hasil sebelum dan sesudah ROS	35
Gambar 4.9 Source Code untuk model SVM	36
Gambar 4.10 Source Code untuk model Random Forest	37
Gambar 4.11 Source Code untuk Epoch	38
Gambar 4.12 Source Code untuk model BERT	39
Gambar 4.13 Confusion Matrix pada SVM dan Random Forest	40
Gambar 4.14 Confusion Matrix Model BERT	41

DAFTAR LAMPIRAN

Lampiran 1. Link Dataset	47
Lampiran 2. Link Google Colab	47



DAFTAR LAMBANG DAN SINGKATAN



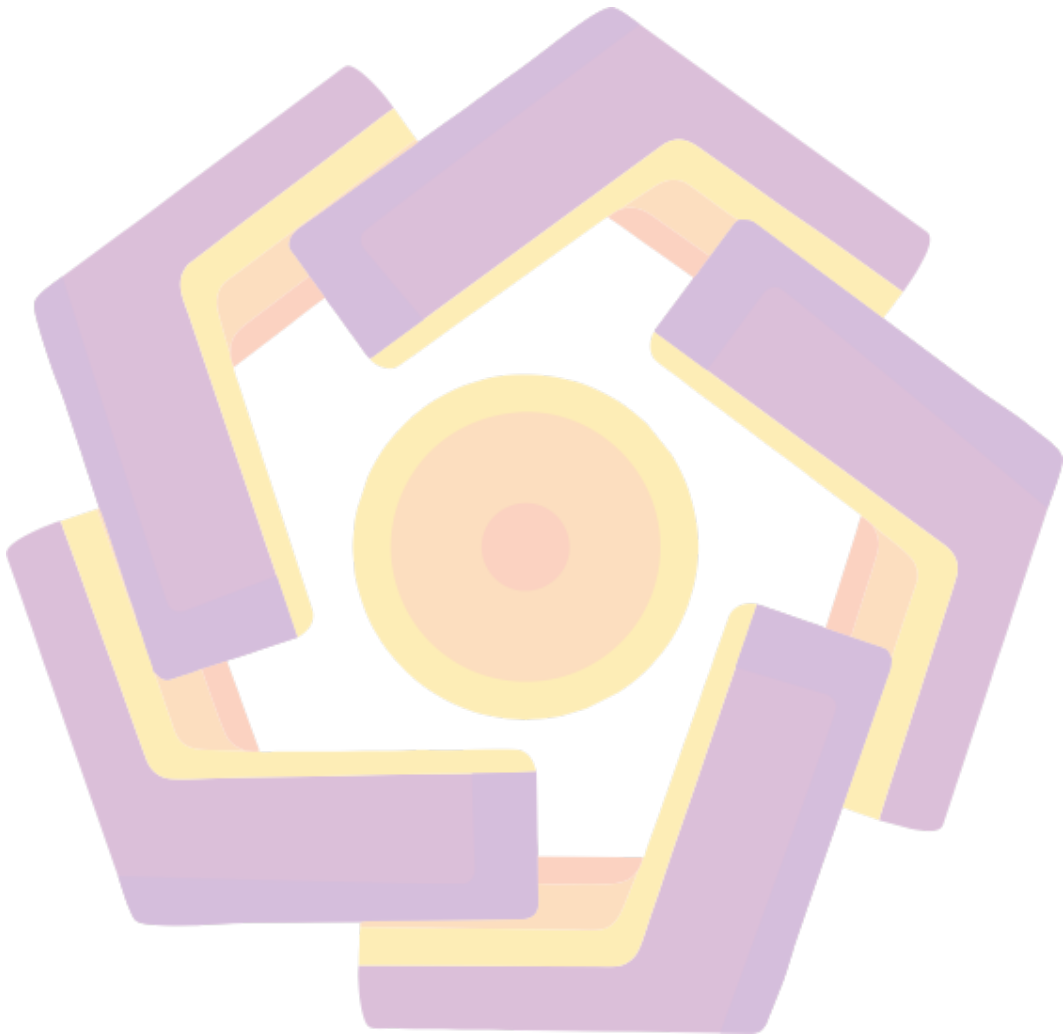
SVM	: Support Vector Machine
RF	: Random Forest
BERT	: Bidirectional Encoder Representations from Transformers
TF-IDF	: Term Frequency - Inverse Document Frequency
NLP	: Natural Language Processing
ROS	: Random Over Sampling
[CLS]	: Token khusus awal input untuk klasifikasi pada BERT
[SEP]	: Token pemisah antar kalimat dalam BERT
[PAD]	: Token padding agar panjang input setara
CSV	: Comma-Separated Values (format file data)
GPU	: Graphics Processing Unit
TPU	: Tensor Processing Unit
JSON	: JavaScript Object Notation (format data teks)

DAFTAR ISTILAH

Analisis Sentimen : Proses mengidentifikasi opini atau ulasan dalam teks

Ulasan : Pendapat atau tanggapan pengguna

Epoch : Jumlah siklus penuh pelatihan model



INTISARI

Penelitian ini bertujuan untuk mengklasifikasikan sentimen opini Generasi Z terhadap dunia kerja menggunakan tiga pendekatan model machine learning, yaitu Random Forest, Support Vector Machine (SVM), dan BERT. Dataset yang digunakan berisi opini-opini dalam bahasa Indonesia yang telah melalui proses pra-pemrosesan dan pelabelan sentimen (negatif, netral, positif). Hasil penelitian menunjukkan bahwa model SVM memberikan performa terbaik dengan akurasi mencapai 90,22% dan distribusi kinerja yang seimbang di semua kelas. Model BERT menyusul dengan akurasi 88,38%, memperlihatkan kemampuan yang kuat dalam mengenali opini positif dan netral, meskipun sedikit lemah dalam mendeteksi opini negatif. Sementara itu, model Random Forest memiliki performa paling rendah dengan akurasi 80,82%, terutama kurang efektif dalam menangkap sentimen negatif. Kesimpulan dari penelitian ini adalah bahwa model SVM paling optimal untuk klasifikasi sentimen opini Gen Z terhadap dunia kerja, baik dari segi akurasi maupun keseimbangan antar kelas. Penelitian ini diharapkan dapat menjadi dasar dalam pengembangan sistem analisis opini publik yang lebih efektif, khususnya yang menyasar generasi muda di dunia kerja.

Kata kunci: *Generasi Z, Dunia Kerja, Support Vector Machine (SVM), Random Forest (RF), Bidirectional Encoder Representations from Transformers (BERT), Machine Learning.*

ABSTRACT

This study aims to classify the sentiment of Generation Z's opinions regarding the world of work using three machine learning models: Random Forest, Support Vector Machine (SVM), and BERT. The dataset consists of Indonesian-language opinions that have undergone preprocessing and sentiment labeling into negative, neutral, and positive classes. The results show that the SVM model achieved the best performance with an accuracy of 90.22% and a balanced distribution across all sentiment classes. The BERT model followed with an accuracy of 88.38%, demonstrating strong capabilities in recognizing positive and neutral opinions, although slightly less effective in detecting negative sentiments. Meanwhile, the Random Forest model recorded the lowest performance with an accuracy of 80.82%, particularly underperforming in identifying negative opinions. In conclusion, the SVM model is the most optimal for classifying Generation Z's opinions on the world of work, offering both high accuracy and class balance. This research is expected to serve as a foundation for developing more effective public opinion analysis systems, especially those targeting young generations in the workforce.

Keyword: Generation Z, Workplace, Support Vector Machine (SVM), Random Forest (RF), Bidirectional Encoder Representations from Transformers (BERT), Machine Learning