

**SENTIMEN ANALISIS KETERBATASAN  
KESEMPATAN KERJA DALAM BIDANG YANG  
DIMINATI BAGI GEN Z MENGGUNAKAN  
ALGORITME NAIVE BAYES**

**Tugas Akhir**

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat Sarjana  
Komputer**



Dibuat oleh :

**SELA HADIYANA**

**160708834**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ATMA JAYA YOGYAKARTA  
2023**

## HALAMAN PENGESAHAN

Tugas Akhir Berjudul

SENTIMEN ANALISIS KETERBATASAN KESEMPATAN KERJA DALAM BIDANG YANG DIMINATI BAGI GEN  
Z MENGGUNAKAN ALGORITMA NAIVE BAYES

yang disusun oleh

Sela Hadiyana

160708834

dinyatakan telah memenuhi syarat pada tanggal 09 Agustus 2023

		Keterangan
Dosen Pembimbing 1	: Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D.	Telah Menyetujui
Dosen Pembimbing 2	: Paulus Mudjihartono, S.T.,M.T., Ph. D	Telah Menyetujui
Tim Penguji		
Penguji 1	: Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D.	Telah Menyetujui
Penguji 2	: Th. Adi Purnomo Sidhi, S.T., M.T.	Telah Menyetujui
Penguji 3	: Aloysius Gonzaga Pradnya Sidhawara, S.T., M.Eng.	Telah Menyetujui

Yogyakarta, 09 Agustus 2023

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Dr. A. Teguh Siswanto, M.Sc.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

# **PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH**

Saya yang bertanda tangan di bawah ini :

Nama Lengkap : Sela Hadiyana  
NPM : 160708834  
Program Studi : Informatika  
Fakultas : Teknologi Industri  
Judul Penelitian : Sentimen Analisis Keterbatasan Kesempatan Kerja Dalam Bidang  
Yang Diminati Bagi Gen Z Menggunakan Algoritme Naive Bayes

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama Penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 10 Juli 2023

Yang menyatakan,

Sela Hadiyana

160708834

## HALAMAN PERSEMBAHAN

*“Jadilah generasi yang cerdas secara intelektual dan moral, serta memiliki rasa empati yang tinggi. Selalu memberi kontribusi kebaikan dalam hal sekecil apa pun. Berfikirlah secara kritis dan terbuka. Bacalah buku sebanyak mungkin”*

## KATA PENGANTAR

Puji dan syukur Penulis haturkan kepada Tuhan yang maha Esa karena berkat rahmat dan karunia-Nya sehingga Penulis dapat menyelesaikan pembuatan tugas akhir “Sentimen Analisis Keterbatasan Kesempatan Kerja Dalam Bidang Yang Diminati Bagi Gen Z Menggunakan Algoritme Naive Bayes”.

Penulis menyadari bahwa penulisan laporan Tugas Akhir dapat Penulis selesaikan dengan baik tidak terlepas dari dukungan, bantuan, dan bimbingan dari beberapa pihak. Adapun, Penulis mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa, atas segala berkah yang diberikan kepada Penulis.
2. Orang Tua, yang selalu memberikan dukungan, kesabaran, serta doa yang terbaik bagi Penulis.
3. Bapak Dr. A. Teguh Siswantoro, M.Sc., sebagai Dekan Fakultas Teknologi Industri UAJY.
4. Bapak Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D., sebagai Dosen Pembimbing 1 dan Ketua Departemen Informatika UAJY.
5. Bapak Th. Adi Purnomo Sidhi, S.T., M.T., sebagai Kepala Prodi Informatika UAJY.
6. Bapak Paulus Mudjihartono, S.T., M.T., Ph.D., sebagai Ketua Kantor Sistem Informasi UAJY, Pembimbing Lapangan Kerja Praktik, serta Dosen Pembimbing 2.
7. Sofyan Saif Pahlevi yang sudah menemani Penulis.
8. Sahabat Muhammad Fitrah Pandjalu yang telah membantu Penulis saat Penulis mengalami kesulitan.
9. Teman kuliah dan sahabat, Dedi Wahyono, Theodorus C.L Kusetyapradana, Widya Laurensia Sitohang, dan Rosalia Elshe Lilis Permathasari, yang telah berteman dan mendukung Penulis selama masa kuliah.
10. Mas Deni Sutrisna yang telah memberi Penulis saran dan masukan.

11. Setiap orang baik yang pernah Penulis temui yang memperlancar urusan Penulis, yang tidak dapat Penulis sebutkan satu per satu.

Demikian laporan Tugas Akhir ini dibuat, semoga laporan ini dapat bermanfaat bagi pihak yang membaca dan membutuhkan. Penulis mengucapkan terimakasih kepada semua pihak.

Yogyakarta, 20 Juli 2023

Sela Hadiyana

160708834

# DAFTAR ISI

SENTIMEN ANALISIS KETERBATASAN KESEMPATAN KERJA DALAM BIDANG YANG DIMINATI BAGI GEN Z MENGGUNAKAN ALGORITME NAIVE BAYES.....	1
LEMBAR PENGESAHAN.....	<b>Error! Bookmark not defined.</b>
PERNYATAAN ORISINALITAS & PUBLIKASI.....	ii
HALAMAN PERSEMBAHAN .....	iii
KATA PENGANTAR .....	iv
DAFTAR ISI.....	vi
DAFTAR GAMBAR .....	ix
DAFTAR TABEL .....	xii
INTISARI.....	xiii
BAB I.....	1
A. Latar Belakang .....	1
B. Rumusan Masalah .....	4
C. Batasan Masalah.....	4
D. Tujuan Penelitian .....	5
E. Metode Pengembangan Model .....	5
1. Studi Literatur .....	5
2. Pengumpulan Data ( <i>Data Collection</i> ).....	6
3. Analisis Data Eksploratif ( <i>Exploratory Data Analysis</i> ).....	6
4. <i>Preprocessing Data</i> .....	6
5. Visualisasi Data.....	6
6. Pelatihan Model.....	7
7. Evaluasi Model.....	7
F. Sistematika Penulisan.....	8

BAB II.....	9
BAB III.....	14
A. <i>Artificial Intelligence (AI)</i> .....	14
B. <i>Machine Learning</i> .....	14
C. <i>Natural Language Processing (NLP)</i> .....	15
D. <i>TF-IDF (Term Frequency – Inverse Document Frequency)</i> .....	17
E. <i>Supervised Learning</i> .....	19
F. <i>Klasifikasi</i> .....	19
G. <i>Algoritme Naive Bayes</i> .....	19
H. <i>Analisis Data (Data Analysis)</i> .....	21
I. <i>Sentimen Analisis (Sentiment Analysis)</i> .....	22
J. <i>Aspect Based Sentiment Analysis</i> .....	23
K. <i>Evaluasi Model (Model Evaluation)</i> .....	24
L. <i>Confusion Matrix</i> .....	26
M. <i>Gen Z</i> .....	27
N. <i>Ekonomi Berkembang</i> .....	28
O. <i>Twitter</i> .....	31
P. <i>Media Sosial</i> .....	32
Q. <i>Blog</i> .....	32
R. <i>Data Pelatihan dan Data Pengujian (Data Training dan Data Testing)</i> .....	32
S. <i>Preprocessing Data</i> .....	34
T. <i>Analisis Data Eksploratif (Exploratory Data Analysis)</i> .....	36
U. <i>Visualisasi Data (Data Visualization)</i> .....	36
V. <i>Pelatihan Model (Training Model)</i> .....	37
W. <i>Python</i> .....	37
X. <i>Tools</i> .....	38
BAB IV .....	50
A. <i>Deskripsi</i> .....	50



B. <i>Dataset</i> .....	50
1. Pengumpulan Data.....	50
2. Analisis Data Eksploratif ( <i>Exploratory Data Analysis</i> ).....	55
3. <i>Preprocessing Data</i> .....	60
C. Pengembangan Model Sentimen Analisis .....	71
1. Pelatihan dan Evaluasi Model .....	71
BAB V.....	75
A. Alur Pengembangan Model .....	75
B. Hasil Evaluasi Model .....	77
C. Hasil Analisis Sentimen.....	79
BAB VI .....	80
A. Kesimpulan .....	80
B. Saran.....	81
DAFTAR PUSTAKA .....	83

## DAFTAR GAMBAR

Gambar 3.1 Confusion Matrix.....	26
Gambar 3.2 Twitter.....	31
Gambar 3.3 Blog.....	32
Gambar 3.4 Google Form.....	39
Gambar 3.5 Excel .....	39
Gambar 3.6 Tampilan Spreadsheet.....	39
Gambar 3.7 Jupyter.....	40
Gambar 3.8 Halaman Jupyter Notebook.....	40
Gambar 3.9 Numpy.....	42
Gambar 3.10 Numpy dari Python.....	42
Gambar 3.11 Pandas .....	43
Gambar 3.12 Matplotlib.....	43
Gambar 3.13 Matplotlib dari Python .....	44
Gambar 3.14 Seaborn.....	44
Gambar 3.15 Wujud Word Cloud.....	45
Gambar 3.16 Library word cloud dari Python .....	45
Gambar 3.17 Regex dari Python.....	46
Gambar 3.18 Library Sastrawi dari Python.....	47
Gambar 3.19 Scikit-learn .....	57
Gambar 4.1 Pertanyaan_1 survei.....	51
Gambar 4.2 Pertanyaan_2 survei.....	51
Gambar 4.3 Impor library .....	56
Gambar 4.4 Dataframe dataset .....	56
Gambar 4.5 Informasi dataset.....	57
Gambar 4.6 Informasi tipe data.....	58
Gambar 4.7 Informasi nilai unik dan non-unik .....	59

Gambar 4.8 Visualisasi dataset.....	59
Gambar 4.9 Dataset.....	60
Gambar 4.10 Lower casing .....	61
Gambar 4.11 Proses menghilangkan Punctuations.....	61
Gambar 4.12 Proses menghilangkan karakter angka.....	62
Gambar 4.13 Proses menghilangkan URL.....	63
Gambar 4.14 Daftar stop words Bahasa Indonesia.....	63
Gambar 4.15 Proses menghilangkan stop words.....	64
Gambar 4.16 Proses menghilangkan kata <3 karakter .....	64
Gambar 4.17 Proses visualisasi dataset menggunakan Word Cloud .....	65
Gambar 4.18 Visualisasi dataset dengan Word Cloud .....	65
Gambar 4.19 Proses visualisasi data sentimen negative dengan Word Cloud.....	66
Gambar 4.20 Visualisasi sentimen negative .....	66
Gambar 4.21 Proses visualisasi data sentimen positive dengan Word Cloud .....	67
Gambar 4.22 Visualisasi sentimen positive .....	67
Gambar 4.23 Proses stemming .....	68
Gambar 4.24 Proses splitting data .....	68
Gambar 4.25 Proses ekstraksi fitur.....	69
Gambar 4.26 Menampilkan jumlah fitur.....	70
Gambar 4.27 Menampilkan matriks fitur numerik .....	70
Gambar 4.28 Menampilkan fitur kata.....	71
Gambar 4.29 Proses pelatihan model.....	72
Gambar 4.30 Inisialisasi evaluasi model.....	72
Gambar 4.31 Proses evaluasi model dengan classification report.....	73
Gambar 4.32 Proses evaluasi model dengan confusion matrix.....	73
Gambar 5.1 Flowchart alur pengembangan model.....	75

Gambar 5.2 Hasil classification report.....77  
Gambar 5.3 Hasil confusion matrix.....78

## DAFTAR TABEL

Tabel 2.1 Perbandingan penelitian .....	12
Tabel 4.1 Sampel dataset.....	52

# INTISARI

## SENTIMEN ANALISIS KETERBATASAN KESEMPATAN KERJA DALAM BIDANG YANG DIMINATI BAGI GEN Z MENGGUNAKAN ALGORITME NAIVE BAYES

Sela Hadiyana

160708834

Di era digital saat ini, orang-orang dapat dengan mudah berbagi opini atau pendapat melalui media sosial. Dengan demikian media sosial dapat menjadi sumber data opini atau sentimen masyarakat, di mana data tersebut dapat digunakan untuk studi sosial. Studi sosial yang saat ini menjadi masalah krusial dalam negara berkembang, termasuk Indonesia adalah persoalan keterbatasan kesempatan kerja. Masalah sosial dapat dirasakan oleh siapa saja, tidak terkecuali *Gen Z* sebagai bagian dari penduduk yang akan dan telah memasuki usia kerja.

Terlepas dari permasalahan tentang keterbatasan kesempatan kerja, dapat dilakukan penelitian sentimen *Gen Z* dari media sosial. Dalam melakukan penelitian sentimen diperlukan algoritme klasifikasi dan teknik ekstraksi fitur untuk mengembangkan model Sentimen Analisis dengan nilai akurasi yang tinggi. Dalam penelitian ini menggunakan algoritme *Naive Bayes* dan teknik TF-IDF, di mana berdasarkan penelitian sebelumnya menghasilkan nilai akurasi yang cukup tinggi.

Dari hasil evaluasi dengan 131 data pengujian, algoritme *Naive Bayes* dengan TF-IDF menghasilkan nilai *Accuracy* 72%. Nilai *Precision* untuk sentimen *Negative* 69% dan sentimen *Positive* 100%. Nilai *Recall* sentimen *Negative* 100% dan sentimen *Positive* 26%. Kemudian, nilai *F1-Score* sentimen *Negative* 81% dan sentimen *Positive* 41%.

Kata Kunci : *Machine Learning*, Sentimen Analisis, TF-IDF, *Natural Language Processing*, *Gen Z*.