

**PEMANFAATAN NATURAL LANGUAGE PROCESSING
(NLP) UNTUK ANALISIS DAN KLASIFIKASI TOPIK
KERESAHAN KARIR BERDASARKAN KOMENTAR
TIKTOK BERBASIS WEB MENGGUNAKAN
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

DISUSUN OLEH

NATASYA MAHARANI

NPM. 2209010079



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

**PEMANFAATAN NATURAL LANGUAGE PROCESSING
(NLP) UNTUK ANALISIS DAN KLASIFIKASI TOPIK
KERESAHAN KARIR BERDASARKAN KOMENTAR
TIKTOK BERBASIS WEB MENGGUNAKAN
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

NATASYA MAHARANI

NPM. 2209010079

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN**

2026

LEMBAR PENGESAHAN

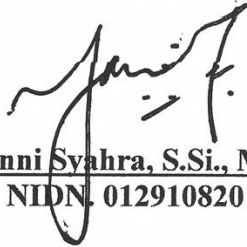
Judul Skripsi : Pemanfaatan Natural Language Processing (NLP) untuk Analisis dan Klasifikasi Topik Keresahan Karir Berdasarkan Komentar TikTok Berbasis Web Menggunakan Support Vector Machine (SVM)

Nama Mahasiswa : NATASYA MAHARANI

NPM : 2209010079

Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Yohanni Syahra, S.Si., M.Kom)
NIDN. 0129108201

Ketua Program Studi



(Mahardika Abdi Prawira Tanjung,
S.Kom, M.Kom)
NIDN. 0117088902

Dekan



Prof. Dr. Al-Khowarizmi,
M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

PEMANFAATAN NATURAL LANGUAGE PROCESSING (NLP) UNTUK ANALISIS DAN KLASIFIKASI TOPIK KERESAHAN KARIR BERDASARKAN KOMENTAR TIKTOK BERBASIS WEB MENGUNAKAN SUPPORT VECTOR MACHINE (SVM)

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, September 2026

Yang membuat pernyataan



Natasya Maharani

NPM. 2209010079

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Natasya Maharani
NPM : 2209010079
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**Pemanfaatan Natural Language Processing (NLP) Untuk Analisis Dan
Klasifikasi Topik Keresahan Karir Berdasarkan Komentar Tiktok Berbasis
Web Menggunakan Support Vector Machine (SVM)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya

Medan, September 2026

Yang membuat pernyataan



Natasya Maharani

NPM. 2209010079

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Natasya Maharani
Tempat dan Tanggal Lahir	: Medan, 28 September 2004
Alamat Rumah	: JL. RAWE VII LINGKUNGAN X
Telepon/Faks/HP	: 082269006994
E-mail	: natasyamaharani47@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: SDN 064005	TAMAT: 2016
SMP	: SMPS Islam Annur Prima Medan	TAMAT: 2019
SMA	: MAPN 4 Medan	TAMAT: 2022

KATA PENGANTAR



Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Pemanfaatan Natural Language Processing (NLP) untuk Analisis dan Klasifikasi Topik Keresahan Karir Berdasarkan Komentar TikTok Berbasis Web Menggunakan Support Vector Machine (SVM)". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.

Selama proses penyusunan skripsi ini, penulis memperoleh banyak bantuan dan dukungan dari berbagai pihak, baik berupa bimbingan, arahan, doa, motivasi, maupun dukungan lainnya. Oleh karena itu, dengan penuh rasa hormat, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Akrim, M.Pd., selaku Rektor Universitas Muhammadiyah Sumatera Utara.
2. Bapak Assoc. Prof. Dr. Al-Khowarizmi, M.Kom., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
3. Ibu Dr. Firaahmi Rizky M.Kom, selaku Wakil Dekan 1 Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
4. Bapak Mhd Basri, S.Si., M.Kom, selaku Wakil Dekan 3 Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
5. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.

6. Bapak Mulkan Azhari, S.Kom., M.Kom., selaku Sekretaris Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
7. Ibu Yohanni Syahra, S.Si., M.Kom., selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, motivasi, serta masukan yang sangat berarti selama proses penyusunan skripsi ini.
8. Seluruh dosen Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara yang telah memberikan ilmu pengetahuan, pengalaman, serta bimbingan kepada penulis selama menempuh pendidikan.
9. Kedua orang tua tercinta dan seluruh keluarga yang selalu menjadi sumber kekuatan melalui doa, kasih sayang, dan dukungan yang tak pernah berhenti. Terima kasih telah menjadi alasan penulis untuk terus berjuang hingga sampai pada titik ini. Semoga pencapaian ini dapat menjadi salah satu bentuk kecil dari rasa terima kasih dan kebahagiaan yang dapat penulis persembahkan kepada keluarga tercinta.
10. Pemilik NPM 220901097, yang telah menemani dan memberikan dukungan, semangat, doa, serta motivasi kepada penulis selama menjalani perkuliahan hingga proses penyusunan skripsi ini.
11. Teman-teman seperjuangan Program Studi Sistem Informasi serta semua pihak yang telah memberikan bantuan, dukungan, dan semangat kepada penulis, baik secara langsung maupun tidak langsung, sehingga skripsi ini dapat diselesaikan.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, penulis terbuka terhadap kritik dan saran yang membangun sebagai bahan perbaikan untuk penyempurnaan skripsi ini.

Wassalamu'alaikum

Warahmatullahi

Wabarakatuh.

PEMANFAATAN NATURAL LANGUAGE PROCESSING (NLP) UNTUK ANALISIS DAN KLASIFIKASI TOPIK KERESAHAN KARIR BERDASARKAN KOMENTAR TIKTOK BERBASIS WEB MENGUNAKAN SUPPORT VECTOR MACHINE (SVM)

ABSTRAK

Keresahan karir menjadi salah satu fenomena yang banyak diutarakan masyarakat melalui media sosial, khususnya TikTok, dalam bentuk komentar yang berisi pengalaman, pendapat, dan keluhan terkait dunia kerja. Komentar tersebut mengandung informasi yang berpotensi dimanfaatkan untuk mengidentifikasi topik keresahan karir, namun bentuknya masih berupa data teks tidak terstruktur sehingga menyulitkan proses analisis secara manual. Studi ini ditujukan untuk membangun sistem klasifikasi topik keresahan karir berdasarkan komentar TikTok memanfaatkan pendekatan *Natural Language Processing* (NLP) dan algoritma *Support Vector Machine* (SVM), beserta penerapan model tersebut pada aplikasi berbasis web.

Penelitian ini mengolah 305 komentar TikTok berbahasa Indonesia yang dikelompokkan ke dalam tiga kategori, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Penetapan label dilakukan secara manual, sebelum data diproses melalui tahapan *pre-processing* yaitu *cleansing*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya, dilakukan pembobotan fitur menerapkan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20. Model klasifikasi dirancang menggunakan algoritma SVM dengan kernel linear, lalu diukur kinerjanya menggunakan *confusion matrix* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

The results of the study show that the SVM-based classification model achieved an accuracy of 96.72%, a macro precision of 96.67%, a macro recall of 96.67%, and a macro F1-score of 96.67%. Based on the confusion matrix analysis, 59 out of 61 test data points were correctly predicted, while 2 data points were misclassified between the Workload and Work Environment categories. The developed web-based system successfully integrated the processes of dataset management, preprocessing, model training, model evaluation, and automatic classification of new comments. Thus, the synergy of NLP, TF-IDF, and SVM proved reliable in classifying career-related concerns based on TikTok comments and was successfully implemented in the web-based system.

Kata Kunci: Keresahan Karir; TikTok; *Natural Language Processing* (NLP); TF-IDF; *Support Vector Machine* (SVM).

UTILIZATION OF NATURAL LANGUAGE PROCESSING (NLP) FOR THE ANALYSIS AND CLASSIFICATION OF CAREER CONCERN TOPICS BASED ON TIKTOK COMMENTS USING SUPPORT VECTOR MACHINE (SVM) WITH A WEB-BASED SYSTEM

ABSTRACT

Career concerns have become a common topic discussed by the public on social media, particularly TikTok, in the form of comments sharing experiences, opinions, and complaints about the workplace. These comments contain information that could be used to identify topics of career-related concerns; however, they are in the form of unstructured text data, making manual analysis difficult. This study aims to develop a system for classifying career anxiety topics based on TikTok comments using a Natural Language Processing (NLP) approach and the Support Vector Machine (SVM) algorithm, along with the implementation of this model in a web-based application.

This study used 305 Indonesian-language TikTok comments classified into three categories: Workload, Work Environment, and Work Compensation. The comments were manually labeled before the data underwent preprocessing steps, including data cleansing, case folding, normalization, tokenization, stopword removal, and stemming. Next, feature weighting was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The dataset was split into training and test sets at an 80:20 ratio. A classification model was designed using the SVM algorithm with a linear kernel, and its performance was evaluated using a confusion matrix based on the metrics of accuracy, precision, recall, and F1-score.

The results showed that the SVM-based classification model achieved an accuracy of 96.72%, a macro precision of 96.67%, a macro recall of 96.67%, and a macro F1-score of 96.67%. Based on the confusion matrix, 59 out of 61 testing data were correctly classified, while only 2 data were misclassified between the Workload and Work Environment categories. The developed web-based system successfully integrated dataset management, pre-processing, model training, model evaluation, and automatic classification of new comments. Therefore, the combination of NLP, TF-IDF, and SVM proved to be effective in classifying career concern topics based on TikTok comments and was successfully implemented in a web-based system.

Keywords: Career Concerns, TikTok, Natural Language Processing (NLP), TF-IDF, Support Vector Machine (SVM).

DAFTAR ISI

LEMBAR PENGESAHAN	i
PENYATAAN ORISINALITAS	ii
PENYATAAN PERSETUJUAN PUBLIKASI.....	iii
RIWAYAT HIDUP	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
BAB I. PENDAHULUAN.....	1
1.1. LATAR BELAKANG MASALAH.....	1
1.2. RUMUSAN MASALAH	5
1.3. BATASAN MASALAH	6
1.4. TUJUAN PENELITIAN	6
1.5. MANFAAT PENELITIAN	7
BAB II. LANDASAN TEORI	9
2.1. KERESAHAN KARIR	9
2.2. MEDIA SOSIAL TIKTOK	13
2.3. KLASIFIKASI	14
2.4. <i>NATURAL LANGUAGE PROCESSING</i> (NLP).....	15
2.5. EKSTRAKSI FITUR TF-IDF	16
2.6. <i>SUPPORT VECTOR MACHINE</i> (SVM)	18
2.7. EVALUASI MODEL KLASIFIKASI	20
2.8. SISTEM BERBASIS WEB.....	23
2.9. PENELITIAN TERDAHULU	23
BAB III. ANALISA DAN PERANCANGAN SISTEM.....	26
3.1. JENIS DAN PENDEKATAN PENELITIAN.....	26
3.2. OBJEK DAN SUBJEK PENELITIAN	26
3.3. TAHAPAN PENELITIAN.....	27
3.4. TEKNIK PENGUMPULAN DATA.....	31
3.5. TEKNIK PELABELAN DATA.....	33
3.6. PRE-PROCESSING DATA TEKS.....	35
3.7. EKSTRAKSI FITUR TF-IDF	36
3.8. PEMBANGUNAN MODEL KLASIFIKASI	37
3.9. EVALUASI MODEL	39
3.10. PERANCANGAN SISTEM BERBASIS WEB.....	40
BAB IV. HASIL DAN PEMBAHASAN	58
4.1. HASIL PENGUMPULAN DATA.....	58
4.2. ATRIBUT DATA.....	59
4.3. HASIL PELABELAN DATA.....	60
4.4. HASIL PRE-PROCESSING DATA	62
4.5. HASIL PEMBOBOTAN TF-IDF	65
4.6. HASIL PEMBAGIAN DATA DAN TRAINING MODEL SVM	68
4.7. HASIL EVALUASI MODEL	70
4.8. HASIL IMPLEMENTASI SISTEM WEB	73

4.9. HASIL PENGUJIAN SISTEM	84
BAB V. PENUTUP.....	86
5.1. KESIMPULAN	86
5.2. SARAN	87
DAFTAR PUSTAKA.....	88
LAMPIRAN.....	95

DAFTAR TABEL

		HALAMAN
TABEL 2.1.	INDIKATOR PELABELAN TOPIK KERESAHAN KARIR	12
TABEL 2.2.	PENELITIAN TERDAHULU	24
TABEL 3.1.	CONTOH DATA MENTAH	32
TABEL 3.2.	CONTOH PELABELAN DATA	34
TABEL 3.3.	TAHAP PRE-PROCESSING	36
TABEL 4.1.	ATRIBUT DATA	59
TABEL 4.2.	DISTRIBUSI LABEL DATASET	61
TABEL 4.3.	HASIL PRE-PROCESSING DATA	64
TABEL 4.4.	PERHITUNGAN TERM FREQUENCY (TF)	66
TABEL 4.5.	CONTOH PERHITUNGAN TF-IDF	67
TABEL 4.6.	PEMBAGIAN DATA TRAINING DAN TESTING	68
TABEL 4.7.	PARAMETER TRAINING MODEL SVM	69
TABEL 4.8.	HASIL EVALUASI MODEL SVM	70
TABEL 4.9.	CONFUSION MATRIX MODEL SVM	71
TABEL 4.10.	METRIKS EVALUASI PERLABEL	72
TABEL 4.11.	HASIL PENGUJIAN SISTEM	84

DAFTAR GAMBAR

	HALAMAN
GAMBAR 3.1	TAHAPAN PENELITIAN 30
GAMBAR 3.2	USE CASE DIAGRAM 42
GAMBAR 3.3	FLOWCHART SISTEM 44
GAMBAR 3.4	ENTITY RELATIONSHIP DIAGRAM (ERD) 46
GAMBAR 3.5	WIREFRAME LOGIN 49
GAMBAR 3.6	WIREFRAME DASHBOARD 50
GAMBAR 3.7	WIREFRAME DATASET 51
GAMBAR 3.8	WIREFRAME PRE-PROCESSING 52
GAMBAR 3.9	WIREFRAME PEMBAGIAN DATA 53
GAMBAR	WIREFRAME TRAINING MODEL 54
3.10	WIREFRAME EVALUASI MODEL 55
GAMBAR	WIREFRAME KLASIFIKASI KOMENTAR 56
3.11	WIREFRAME RIWAYAT KOMENTAR 57
GAMBAR 3.12	HASIL CLASSIFICATION 71
GAMBAR 3.13	REPORT MENGGUNAKAN 72
GAMBAR 4.1	PYTHON 72
	HASIL CONFUSION MATRIX MENGGUNAKAN
GAMBAR 4.2	PYTHON 74
	HALAMAN LOGIN 75
GAMBAR 4.3	HALAMAN DASHBOARD 76
GAMBAR 4.4	HALAMAN DATASET 77
GAMBAR 4.5	HALAMAN PRE-PROCESSING 78
GAMBAR 4.6	HALAMAN PEMBAGIAN DATA 79
GAMBAR 4.7	HALAMAN TRAINING MODEL 80
GAMBAR 4.8	HALAMAN EVALUASI MODEL 81
GAMBAR 4.9	HALAMAN MANAJEMEN MODEL 82
GAMBAR 4.10	HALAMAN KLASIFIKASI KOMENTAR 83
GAMBAR 4.11	HALAMAN RIWAYAT KOMENTAR
GAMBAR 4.12	

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi digital telah memberikan pengaruh besar terhadap perubahan dunia kerja. Kemajuan teknologi, otomatisasi, dan digitalisasi proses bisnis menyebabkan kebutuhan terhadap keterampilan tenaga kerja ikut berubah. Pekerja tidak hanya dituntut untuk mampu menyelesaikan tugas sesuai bidangnya, tetapi juga perlu beradaptasi dengan perkembangan teknologi, pola kerja baru, serta tuntutan organisasi yang semakin dinamis. Laporan World Economic Forum menunjukkan bahwa perkembangan teknologi dan otomatisasi turut mendorong perubahan jenis pekerjaan serta kompetensi yang dibutuhkan di masa depan (Bahram, 2023).

Perubahan tersebut tidak hanya berdampak pada kebutuhan keterampilan, tetapi juga memengaruhi stabilitas pekerjaan dan kepastian karir. Kondisi pasar kerja yang semakin kompetitif dapat menimbulkan tekanan bagi individu, khususnya bagi pencari kerja, pekerja muda, maupun fresh graduate yang sedang memasuki dunia kerja. Ketidaksesuaian antara harapan dan kenyataan di dunia kerja, seperti sulitnya memperoleh pekerjaan, beban kerja yang tinggi, serta kompensasi yang belum sesuai, dapat memunculkan keresahan terhadap masa depan karir. Dalam beberapa kondisi, tekanan tersebut juga dapat berkaitan dengan kejenuhan kerja, stres, dan rasa tidak aman terhadap keberlanjutan pekerjaan (Luwei & Huimin, 2024).

Keresahan karir dapat dipahami sebagai bentuk kekhawatiran individu terhadap kondisi pekerjaan, peluang kerja, beban kerja, maupun kesejahteraan

yang berkaitan dengan karirnya. Keresahan ini tidak hanya muncul dalam percakapan langsung, tetapi juga banyak diekspresikan melalui media sosial. Saat ini, media sosial tidak lagi hanya digunakan sebagai sarana hiburan, melainkan juga menjadi ruang bagi pengguna untuk menyampaikan pendapat, pengalaman, dan keluhan pribadi secara terbuka (Kurniawan et al., 2021). Fenomena berbagi cerita atau “curhat” di media sosial menunjukkan bahwa pengguna memanfaatkan platform digital sebagai tempat untuk memperoleh dukungan sosial, validasi, maupun respons dari pengguna lain (Wicaksana & Rudy, 2021).

Salah satu platform media sosial yang banyak digunakan adalah TikTok. TikTok memiliki jumlah pengguna yang besar dan tingkat interaksi yang tinggi, termasuk di Indonesia. Indonesia bahkan tercatat sebagai salah satu negara dengan jumlah pengguna TikTok terbesar, yaitu mencapai 99,1 juta pengguna (Indriyani et al., 2023). Selain digunakan untuk hiburan, TikTok juga menjadi ruang diskusi mengenai berbagai isu sosial, termasuk permasalahan dunia kerja. Melalui kolom komentar, pengguna dapat memberikan tanggapan secara spontan terhadap suatu konten, baik berupa pendapat, pengalaman pribadi, maupun keluhan yang berkaitan dengan pekerjaan dan karir (Mardianto, 2023).

Komentar TikTok memiliki potensi untuk dianalisis karena memuat berbagai bentuk ekspresi pengguna terhadap isu yang sedang terjadi. Dalam konteks keresahan karir, komentar-komentar tersebut dapat menggambarkan berbagai persoalan, seperti ketidaksesuaian gaji, beban kerja berlebih, lingkungan kerja yang kurang mendukung, maupun kekhawatiran terhadap kondisi pekerjaan. Namun, data komentar media sosial termasuk ke dalam data teks tidak terstruktur. Komentar yang ditulis pengguna sering kali menggunakan bahasa tidak baku,

singkatan, ragam bahasa, simbol, emotikon, serta gaya penulisan informal. Kondisi ini membuat proses analisis secara manual menjadi kurang efektif, terutama apabila jumlah data yang dianalisis cukup banyak (Abdurrazik & Wirawana, 2025).

Untuk mengolah data teks tidak terstruktur, diperlukan pendekatan *Natural Language Processing* (NLP) agar komputer dapat memproses dan menganalisis bahasa manusia dalam bentuk teks. Dalam penelitian ini, NLP dimanfaatkan dalam fase *pre-processing* komentar, mencakup *case folding*, tokenisasi, pembuangan *stopword*, normalisasi, dan *stemming*, sehingga data lebih bersih dan siap untuk tahap analisis. Tahapan *pre-processing* tersebut umum digunakan dalam *text mining* untuk mempersiapkan data teks sebelum dilakukan pengolahan lebih lanjut (Rifaldi et al., 2023). Selanjutnya, data teks dikonversi ke dalam bentuk numerik memakai teknik *Term Frequency-Inverse Document Frequency* (TF-IDF), yakni pendekatan pembobotan kata yang berfungsi menentukan tingkat kepentingan tiap kata pada data teks (Wibowo et al., 2024).

Setelah data komentar diubah menjadi bentuk numerik, proses klasifikasi dapat dilakukan menggunakan algoritma *machine learning*. Salah satu algoritma yang sering diandalkan untuk pengelompokan teks yaitu *Support Vector Machine* (SVM). Prinsip kerja SVM ialah menemukan garis atau bidang pemisah optimal yang dikenal sebagai *hyperplane* guna memisahkan data ke kelas spesifik (Mutawalli et al., 2019). Pemilihan algoritma ini didasari oleh ketangguhannya dalam mengolah data berdimensi tinggi, seperti halnya data teks yang dipetakan lewat TF-IDF. Selain itu, SVM juga terbukti luas digunakan dalam penelitian

klasifikasi teks dan terbukti mampu menghasilkan performa yang baik pada data media sosial (Fide et al., 2021).

Beberapa algoritma lain, seperti Naïve Bayes dan K-Nearest Neighbor (KNN), juga dapat digunakan dalam klasifikasi teks. Namun, Naïve Bayes memiliki asumsi bahwa setiap fitur saling independen, sedangkan pada data teks keterkaitan antar kata sering kali tetap terjadi (Agustian & Bisri, 2019). Sedangkan, K-Nearest Neighbor (KNN) sensitif terhadap data pencilan (*outlier*) dan cukup bergantung pada pemilihan nilai parameter k , yang dapat memengaruhi performa klasifikasi (Indrayuni et al., 2021). Oleh karena itu, SVM digunakan dalam penelitian ini karena dinilai sesuai untuk mengklasifikasikan komentar TikTok ke dalam beberapa topik keresahan karir.

Berdasarkan tinjauan terhadap berbagai penelitian terdahulu, analisis teks pada media sosial umumnya masih berfokus pada analisis sentimen yang mengelompokkan opini ke dalam kategori positif, negatif, dan netral. Sebagai contoh, penelitian Hilmi & Irwiensyah (2024), menerapkan analisis sentimen pada data media sosial untuk mengklasifikasikan opini pengguna. Pendekatan tersebut memang dapat menunjukkan kecenderungan emosi pengguna, tetapi belum menjelaskan topik atau bentuk permasalahan yang sedang dibicarakan secara lebih spesifik.

Selain itu, sebagian penelitian sebelumnya masih banyak menggunakan Twitter, ulasan aplikasi, atau survei sebagai sumber data penelitian. Sebagai contoh, penelitian oleh, Herlinda et al. (2024) menggunakan Twitter sebagai sumber data penelitian. Padahal, TikTok juga memiliki potensi sebagai sumber

data karena interaksi pengguna melalui kolom komentar cukup tinggi dan sering memuat ungkapan pengalaman pribadi secara spontan.

Berdasarkan uraian tersebut, terdapat kesenjangan penelitian dalam pemanfaatan komentar TikTok untuk mengklasifikasikan topik keresahan karir secara terstruktur. Studi terdahulu mayoritas masih berfokus pada polaritas sentimen, seperti positif, negatif, atau netral, sehingga belum mengenali topik spesifik yang menggambarkan bentuk keresahan karir secara lebih mendalam. Penelitian ini ditujukan untuk mengembangkan model klasifikasi yang mampu mengidentifikasi dan mengelompokkan topik keresahan karir dari komentar TikTok, serta menganalisis hasil pengelompokan topik berdasarkan data komentar yang digunakan dalam penelitian. Model yang dibangun kemudian diterapkan ke dalam sistem berbasis web sehingga hasil klasifikasi dapat disajikan secara terstruktur dan mudah dipahami. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran berbasis empiris mengenai pola keresahan karir yang berkembang di masyarakat digital.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana membangun model klasifikasi topik keresahan karir berdasarkan komentar TikTok menggunakan pendekatan *Natural Language Processing* (NLP) dan algoritma *Support Vector Machine* (SVM)?
2. Bagaimana tingkat kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan topik keresahan karir pada data komentar TikTok.

3. Bagaimana mengimplementasikan model klasifikasi tersebut ke dalam sistem berbasis web?

1.3. Batasan Masalah

Agar penelitian ini tetap fokus dan terarah, maka ruang lingkup penelitian dibatasi. Batasan masalah ditetapkan sebagai berikut:

1. Penelitian ini menggunakan 305 data komentar TikTok berbahasa Indonesia yang relevan dengan isu keresahan karir.
2. Data komentar yang digunakan merupakan komentar yang tersedia secara publik dan tidak mencantumkan identitas pribadi pengguna TikTok.
3. Proses pelabelan dilakukan secara manual dengan pendekatan *single-label classification*, di mana setiap komentar hanya diberikan satu label utama berdasarkan topik yang paling dominan.
4. Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi dan tidak membandingkannya dengan metode klasifikasi lainnya.
5. Penelitian ini berfokus pada komentar yang membahas pengalaman, keluhan, atau persepsi pengguna terhadap kondisi dunia kerja, sehingga tidak membahas keresahan akademik, pendidikan, maupun masalah pribadi di luar konteks pekerjaan.
6. Penelitian ini tidak melakukan analisis sentimen maupun analisis psikologis secara klinis, melainkan berfokus pada pengelompokan topik keresahan karir berdasarkan isi teks komentar.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan sebelumnya, penelitian ini diarahkan untuk mencapai tujuan sebagai berikut:

1. Untuk membangun model klasifikasi topik keresahan karir berdasarkan komentar TikTok menggunakan pendekatan *Natural Language Processing* (NLP) dan algoritma *Support Vector Machine* (SVM).
2. Untuk mengetahui tingkat kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan topik keresahan karir pada data komentar TikTok berdasarkan metrik evaluasi yang digunakan.
3. Untuk merancang dan mengimplementasikan sistem berbasis web yang mampu menyajikan hasil klasifikasi topik keresahan karir secara informatif dan mudah dipahami.

1.5. Manfaat Penelitian

Hasil penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Perguruan Tinggi

Penelitian ini diharapkan dapat menjadi bahan referensi bagi penelitian selanjutnya dalam bidang *text mining* dan *Natural Language Processing* (NLP), khususnya terkait penerapan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan teks dari media sosial. Selain itu, penelitian ini dapat menjadi tambahan informasi bagi pengembangan penelitian mengenai analisis komentar dan klasifikasi teks berbahasa Indonesia.

2. Bagi Penulis

Penelitian ini menambah pengalaman dan pemahaman penulis dalam menerapkan *Natural Language Processing* (NLP) dan algoritma *Support*

Vector Machine (SVM) untuk mengolah serta mengklasifikasikan data komentar dari media sosial. Selain itu, penelitian ini juga memberikan pengalaman dalam mengembangkan sistem berbasis web yang dapat mendukung proses pengolahan dan klasifikasi data.

3. Bagi Masyarakat

Penelitian ini diharapkan dapat memberikan gambaran mengenai hasil analisis topik keresahan karir berdasarkan komentar TikTok yang digunakan dalam penelitian, sehingga dapat menjadi informasi pendukung dalam memahami berbagai isu karir yang dibahas oleh pengguna media sosial.

BAB II

LANDASAN TEORI

2.1. Keresahan Karir

Karir dipahami sebagai pola pengalaman yang berkaitan dengan pekerjaan, mencakup posisi, tugas, tanggung jawab, serta perkembangan individu dalam dunia kerja (Kusumawati, 2017). Selain itu, karir juga dipandang sebagai bagian penting dalam perjalanan hidup karena merepresentasikan peran sosial, pencapaian, dan identitas individu di masyarakat (Juliana & Gunawan, 2021). Dalam proses menjalani karir, individu tidak selalu berada dalam kondisi yang stabil secara emosional.

Dalam konteks dunia kerja, keresahan berkaitan dengan kekhawatiran yang dirasakan individu ketika menjalani atau menghadapi lingkungan pekerjaan. Kondisi ini dapat ditandai dengan rasa tidak aman, kurang percaya diri, tekanan mental, serta perasaan tidak siap dalam menghadapi tuntutan kerja (Anggini & Syarqawi, 2023).

Berdasarkan berbagai pandangan tersebut, keresahan karir dalam penelitian ini didefinisikan sebagai kondisi emosional berupa perasaan gelisah, tidak nyaman, atau ketidaktenangan yang muncul akibat tekanan, ketidakpastian, maupun ketidaksesuaian antara harapan karir dengan realitas kerja. Penelitian ini berfokus pada individu yang telah bekerja, sehingga analisis didasarkan pada pengalaman empiris yang tercermin dalam komentar media sosial.

2.1.1. Faktor-Faktor Pemicu Keresahan Karir

Kondisi lingkungan kerja modern yang dinamis memiliki pengaruh terhadap stabilitas psikologis pekerja. (Indriyani & Dihan, 2025) menjelaskan

bahwa faktor yang dapat memicu stres kerja dan ketidakpuasan kerja antara lain beban kerja, kondisi lingkungan kerja, serta sistem kompensasi. Faktor-faktor tersebut dapat menjadi sumber keresahan apabila tidak sesuai dengan harapan, kebutuhan, atau kemampuan individu dalam menjalani pekerjaan.

1. Beban Kerja (*Workload*)

Beban kerja mencakup tuntutan tugas yang tinggi, volume pekerjaan yang berlebihan, tekanan waktu, serta target kerja yang tidak realistis. Menurut (Bakker & Demerouti, 2017) dalam teori *Job Demands-Resources*, tuntutan kerja yang berlangsung terus-menerus tanpa dukungan sumber daya dan pemulihan yang memadai dapat memicu kelelahan kronis atau *burnout*. Dalam konteks pekerjaan, keresahan akibat beban kerja dapat muncul dalam bentuk keluhan mengenai lembur, tekanan mental, kelelahan fisik, serta tuntutan pekerjaan yang berlebihan.

2. Lingkungan Kerja (*Work Environment*)

Lingkungan kerja meliputi berbagai aspek, baik secara fisik maupun nonfisik, yang dapat memengaruhi tingkat kenyamanan seseorang dalam menjalankan pekerjaannya. Dalam konteks nonfisik atau psikososial, lingkungan kerja berkaitan dengan hubungan antara pekerja dengan atasan, rekan kerja, budaya organisasi, serta pola komunikasi di tempat kerja. Penelitian oleh (Sull et al., 2022) mengungkapkan bahwa budaya kerja yang bersifat toksik atau *toxic culture* dapat menjadi salah satu faktor yang menimbulkan ketidaknyamanan bagi pekerja serta meningkatkan keinginan untuk meninggalkan pekerjaannya. Kondisi tersebut dapat meliputi perilaku tidak sopan, kurangnya dukungan, konflik interpersonal, serta budaya kerja yang tidak sehat.

3. Kompensasi Kerja (*Work Compensation*)

Kompensasi berkaitan dengan imbalan finansial maupun nonfinansial yang diterima pekerja. Imbalan finansial dapat berupa gaji, upah, bonus, atau tunjangan, sedangkan imbalan nonfinansial dapat berupa penghargaan, kesempatan pengembangan karir, dan jaminan kesejahteraan. (Schmitt et al., 2024) menyatakan bahwa kekhawatiran finansial merupakan bagian dari keresahan karir karena individu dapat merasa bahwa karirnya belum mampu memberikan jaminan kesejahteraan dan keamanan ekonomi di masa depan. Keresahan dapat muncul ketika terdapat ketidakseimbangan antara beban kerja, kontribusi, dan imbalan yang diterima.

2.1.2. Operasionalisasi Konsep ke dalam Label Klasifikasi

Konsep keresahan karir dapat dioperasionalkan ke dalam beberapa kategori agar dapat digunakan dalam proses klasifikasi teks. Operasionalisasi konsep bertujuan untuk mengubah konsep teoretis menjadi kategori yang lebih konkret dan dapat diamati berdasarkan isi komentar. Pendekatan ini dilakukan dengan mengacu pada teori mengenai faktor pemicu keresahan karir serta pengamatan terhadap data komentar media sosial.

Penyusunan kategori klasifikasi mengacu pada prinsip *content analysis* sebagaimana dijelaskan oleh (Krippendorff, 2018), yaitu kategori harus disusun secara sistematis, konsisten, dan mampu merepresentasikan makna yang terkandung dalam data berdasarkan kerangka teori yang jelas.

Berdasarkan teori mengenai faktor pemicu keresahan karir yang telah diuraikan sebelumnya, peneliti menyusun indikator pelabelan sebagai acuan dalam proses klasifikasi komentar. Rincian indikator tersebut dapat dilihat pada Tabel 2.1.

Tabel 2.1 Indikator Pelabelan Topik Keresahan Karir

Label	Indikator
Beban Kerja	Lembur, target kerja, tekanan kerja, beban kerja berlebih, kelelahan fisik, kelelahan mental, jam kerja panjang, tuntutan pekerjaan.
Lingkungan Kerja	Atasan, rekan kerja, budaya organisasi, toxic workplace, konflik, komunikasi, diskriminasi, suasana kerja, politik kantor.
Kompensasi Kerja	Gaji, upah, bonus, tunjangan, insentif, keterlambatan pembayaran, pendapatan tidak sesuai, kesejahteraan, ketidakseimbangan antara beban kerja dan imbalan.

Berdasarkan indikator tersebut, kategori klasifikasi dijelaskan sebagai berikut:

- a. Label keresahan tentang Beban Kerja yaitu kategori yang digunakan untuk mengklasifikasikan komentar yang berfokus pada kuantitas pekerjaan, tekanan waktu, tuntutan lembur, target kerja, serta dampak kelelahan fisik maupun mental.
- b. Label keresahan tentang Lingkungan Kerja yaitu kategori yang digunakan untuk mengklasifikasikan komentar yang berkaitan dengan hubungan sosial di tempat kerja, perilaku atasan, perilaku rekan kerja, budaya organisasi, politik kantor, serta ketidaknyamanan terhadap suasana kerja.
- c. Label keresahan tentang Kompensasi Kerja yaitu kategori yang digunakan untuk mengklasifikasikan komentar yang berfokus pada isu finansial, seperti gaji, upah, tunjangan, keterlambatan pembayaran, ketidakcukupan pendapatan, serta ketidakseimbangan antara beban kerja yang diberikan dengan imbalan yang diterima.

Pembagian kategori tersebut digunakan agar komentar yang berkaitan dengan keresahan karir dapat dikelompokkan secara lebih terstruktur. Dengan adanya indikator yang jelas, proses klasifikasi dapat dilakukan secara konsisten sesuai dengan makna utama yang terkandung dalam komentar.

2.2. Media Sosial TikTok

Media sosial merupakan sarana komunikasi digital yang memungkinkan pengguna untuk berinteraksi, berbagi informasi, menyampaikan pendapat, serta membahas beragam konten melalui platform daring. Aktivitas tersebut dapat dilakukan melalui berbagai bentuk media, termasuk teks, gambar, video, dan audio, dengan dukungan koneksi internet (Ilpa Hasanah, 2026).

Salah satu media sosial yang mengalami perkembangan pesat adalah TikTok. TikTok telah bertransformasi menjadi media informasi baru yang populer di kalangan masyarakat. Pengguna TikTok tidak hanya memanfaatkan platform ini sebagai media hiburan, tetapi juga memanfaatkannya untuk berbagi pengetahuan, pengalaman, wawasan, dan informasi dalam berbagai bidang (Bur et al., 2023).

Selain sebagai media berbagi informasi, TikTok juga menyediakan fitur kolom komentar yang memungkinkan terjadinya interaksi antarpengguna (Indriyani et al., 2023). Melalui kolom komentar tersebut, pengguna tidak hanya menyampaikan tanggapan terhadap konten, tetapi juga mengekspresikan opini publik terkait berbagai isu sosial, termasuk isu ketenagakerjaan dan karir.

Pola penyampaian opini melalui komentar TikTok memiliki kesamaan dengan platform digital lain, seperti YouTube, di mana kolom komentar digunakan sebagai sarana ekspresi pandangan masyarakat secara terbuka (Mulyono & Saprudin, 2025). Oleh karena itu, komentar pengguna TikTok memiliki potensi untuk dimanfaatkan sebagai sumber data penelitian karena mencerminkan pandangan, pengalaman, dan keresahan masyarakat secara aktual dalam konteks dunia kerja.

2.3. Klasifikasi

Klasifikasi merupakan suatu proses untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik serta pola yang terdapat di dalamnya. Dalam pembelajaran mesin, Klasifikasi merupakan bagian dari metode *supervised learning* yang memanfaatkan data berlabel dalam proses pembelajaran untuk membentuk model yang dapat memprediksi kelas dari data baru (Nurhalizah et al., 2024).

Menurut (Yudana et al., 2023) klasifikasi digunakan untuk mengidentifikasi kesamaan karakteristik dalam suatu kelompok data dan mengelompokkannya ke dalam kelas tertentu dengan tujuan memprediksi label kelas yang telah ditentukan sebelumnya. Proses klasifikasi dilakukan dengan mengidentifikasi pola pada data latih, kemudian pola tersebut digunakan sebagai dasar untuk menentukan kelas pada data uji maupun data baru.

Salah satu penerapan klasifikasi adalah klasifikasi teks. Klasifikasi teks, yang juga dikenal sebagai *text tagging* atau *text labeling*, merupakan proses pengorganisasian teks ke dalam kategori atau label tertentu berdasarkan isi atau kontennya (Abdurrazik & Wirawana, 2025). Secara umum, klasifikasi teks ditujukan guna mengategorikan suatu teks ke dalam dua atau lebih kelompok yang telah ditetapkan sebelumnya (Rahma & Suadaa, 2023). Dalam konteks klasifikasi topik, teks dikelompokkan berdasarkan tema atau permasalahan utama yang terkandung di dalamnya. Dengan demikian, klasifikasi teks dapat dimanfaatkan untuk mengelompokkan komentar pada media sosial ke dalam kategori tertentu berdasarkan isi yang terkandung dalam komentar tersebut.

2.4. *Natural Language Processing (NLP)*

Natural Language Processing (NLP) ialah cabang disiplin yang mengombinasikan ilmu komputer, kecerdasan buatan, maupun linguistik guna memfasilitasi komputer dalam memahami, memproses, hingga menganalisis bahasa manusia. NLP dimanfaatkan untuk memproses data berbentuk bahasa alami, baik berupa teks maupun suara, agar dapat dianalisis oleh sistem komputer. NLP memungkinkan komputer untuk memahami serta menganalisis bahasa manusia melalui tahapan pengolahan bahasa alami (Rivaldi et al., 2024).

Dalam klasifikasi teks, NLP berfungsi mengolah data teks tanpa struktur menjadi data yang lebih teratur, bersih, serta siap diolah lebih lanjut oleh sistem. Data teks dari media sosial umumnya memiliki karakteristik tidak baku, seperti penggunaan singkatan, emotikon, tanda baca berlebihan, simbol, serta variasi penulisan kata. Oleh sebab itu, data perlu melalui tahap *pre-processing* teks sebelum masuk ke tahapan ekstraksi fitur dan klasifikasi. Tahapan *pre-processing* dilakukan untuk menyeleksi dan mengolah data teks agar makin rapi dan terstruktur sebelum masuk ke tahapan analisis (Astuti, 2024). Selain itu, tahapan seperti tokenisasi, penghapusan *stopword*, konversi huruf kecil, dan stemming umum digunakan dalam proses klasifikasi teks (HaCohen-Kerner et al., 2020).

Tahap *pre-processing* teks dalam NLP meliputi beberapa proses, yaitu:

1. *Cleansing*, yakni proses membersihkan teks dari karakter atau elemen yang tidak relevan, seperti tanda baca, angka, maupun simbol, tautan, username, dan emotikon.

2. *Case Folding*, yakni proses penyetaraan bentuk teks dengan mengubah seluruh huruf besar menjadi huruf kecil.
3. Normalisasi, yakni proses untuk mengonversi kata-kata yang tidak baku, singkatan, maupun bahasa tidak formal menjadi bentuk yang lebih baku.
4. *Tokenizing*, yakni proses pemecahan teks menjadi unit-unit yang lebih kecil, seperti kata, frasa, atau simbol.
5. *Stopword Removal*, yakni proses pengeliminasian kata-kata umum yang kurang memiliki makna signifikan, contohnya "yang", "di", "ke", "dari", dan "adalah".
6. *Stemming*, yaitu tahapan mengubah kata berimbuhan menjadi bentuk kata dasar untuk menyederhanakan representasi teks.

2.5. Ekstraksi Fitur TF-IDF

Setelah tahap *pre-processing* selesai, teks perlu dikonversi ke dalam bentuk numerik agar dapat diproses oleh algoritma *machine learning*. Salah satu metode yang umum digunakan untuk merepresentasikan teks secara numerik adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF yakni metode pembobotan yang digunakan untuk mengukur nilai suatu kata berdasarkan tingkat kepentingannya dalam sebuah dokumen dengan memperhitungkan keseluruhan dokumen dalam *corpus* (Abdurrazik & Wirawana, 2025).

TF-IDF merupakan metode ekstraksi fitur sering diimplementasikan dalam *text mining* dan klasifikasi teks. Metode ini menghasilkan nilai statistik untuk menggambarkan tingkat kepentingan suatu kata dalam sebuah dokumen. TF-IDF mencakup dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Nilai TF menggambarkan frekuensi kemunculan

suatu istilah dalam dokumen, sedangkan nilai IDF menunjukkan tingkat kepentingan istilah tersebut dalam kumpulan dokumen. Pembobotan menggunakan TF-IDF dilakukan untuk mengubah kata menjadi nilai numerik yang selanjutnya dapat digunakan sebagai masukan dalam proses klasifikasi teks (Herlinda et al., 2024). Untuk memahami proses perhitungan TF-IDF, berikut disajikan rumus dari masing-masing komponen pembentuknya.

1. *Term Frequency* (TF).

Komponen pertama dalam pembobotan TF-IDF yakni *Term Frequency* (TF). TF berfungsi menghitung intensitas kemunculan suatu kata dalam suatu dokumen. Rumus perhitungan TF dapat dinyatakan sebagai berikut:

$$TF(t, d) = \frac{f(t, d)}{\sum_{t' \in d} f(t', d)} \quad (1)$$

Keterangan:

- a. t = term atau kata yang dihitung
- b. d = dokumen tempat term berada
- c. $f(t, d)$ = frekuensi kemunculan term t pada dokumen d
- d. $\sum_{t' \in d} f(t', d)$ = total seluruh kata dalam dokumen d

2. *Inverse Document Frequency* (IDF)

Inverse Document Frequency (IDF) berperan menentukan derajat kepentingan suatu *term* atau kata pada populasi dokumen. Kata yang tersebar di berbagai dokumen cenderung mempunyai bobot IDF lebih kecil, sedangkan kata yang hanya ditemukan pada sebagian kecil dokumen akan

mendapatkan skor IDF yang lebih tinggi. Rumus IDF dapat dituliskan sebagai berikut:

$$IDF(t) = \log\left(\frac{N}{df_t}\right) \quad (2)$$

Keterangan:

- a. t = term atau kata yang dihitung
- b. N = jumlah semua dokumen dalam dataset
- c. df_t = jumlah dokumen yang memiliki term t
- d. $\log$ = fungsi logaritma yang digunakan untuk menstabilkan nilai IDF

3. TF-IDF

Nilai dari TF-IDF didapatkan dari perkalian antara nilai TF dan IDF.

Rumus perhitungan TF-IDF dapat dituliskan sebagai berikut:

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

Keterangan:

- a. $TF(t, d)$ = nilai *term frequency* dari kata t pada dokumen d
- b. $IDF(t)$ = nilai *inverse document frequency* dari term t

Nilai TF-IDF menggambarkan tingkat kepentingan suatu kata pada sebuah dokumen. Semakin tinggi nilai TF-IDF, semakin besar kontribusi kata tersebut guna memisahkan suatu dokumen dengan dokumen lainnya. Oleh sebab itu, TF-IDF sering diimplementasikan pada klasifikasi teks karena mampu mentransformasikan data teks menjadi format numerik yang dapat diolah oleh algoritma pembelajaran mesin (*machine learning*).

2.6. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode *supervised learning* yang bisa dimanfaatkan untuk proses klasifikasi maupun regresi. Pada

proses klasifikasi, SVM mencari *hyperplane* optimal yang berperan sebagai batas pemisah antara data dari kelas yang berbeda. *Hyperplane* yang optimal yakni batas pemisah yang mempunyai jarak atau *margin* terbesar terhadap data terdekat dari setiap kelas. Data yang terletak paling dekat dengan *hyperplane* tersebut dikenal sebagai *support vector* (Fide et al., 2021).

Kelebihan utama SVM berada pada kemampuannya mengolah data berdimensi tinggi serta membentuk model klasifikasi yang optimal dan stabil. Kelebihan ini membuat SVM sering diandalkan dalam berbagai penelitian terkait klasifikasi teks. Selain itu, SVM dikenal sebagai salah satu algoritma *machine learning* yang mampu membentuk model klasifikasi dengan kinerja yang optimal (Aulia et al., 2021).

Secara matematis, fungsi klasifikasi pada SVM dapat dituliskan sebagai berikut:

$$f(x) = w \cdot x + b \quad (4)$$

Keterangan:

- a. w = vektor bobot (weight)
- b. x = vektor fitur (data input)
- c. b = bias (intersep)
- d. $f(x)$ = fungsi keputusan untuk menentukan posisi data terhadap *hyperplane*

Fungsi tersebut digunakan untuk menentukan posisi suatu data terhadap *hyperplane*. Jika nilai fungsi berada pada satu sisi *hyperplane*, maka data akan diklasifikasikan ke dalam kelas tertentu. Sebaliknya, jika berada pada sisi lain, maka data akan diklasifikasikan ke kelas yang berbeda.

Pada klasifikasi teks, hasil ekstraksi fitur berbasis TF-IDF dipetakan menjadi vektor numerik. Selanjutnya, vektor tersebut dijadikan masukan untuk algoritma SVM guna mengenali pola dari data yang sudah dilabeli dan memetakan kelas pada data baru.

2.7. Evaluasi Model Klasifikasi

Evaluasi model merupakan tahapan yang dilakukan untuk mengukur kualitas, keandalan, dan kinerja model dalam menghasilkan prediksi yang akurat berdasarkan data yang telah dipelajari (Merdiansah et al., 2024). Tahapan ini dilakukan setelah model terbentuk dengan tujuan untuk mengetahui sejauh mana model mampu melakukan klasifikasi secara tepat terhadap data uji (Budiman, 2021). Selain itu, evaluasi juga digunakan untuk mengukur efektivitas model klasifikasi *machine learning* yang digunakan (Alhaqq et al., 2022).

Pada klasifikasi, khususnya klasifikasi teks, kinerja model umumnya diukur memakai *confusion matrix* yang mencakup beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* (Hendriyanto & Sari, 2022). *Confusion matrix* yakni metode evaluasi yang difungsikan untuk membandingkan hasil prediksi model dengan kelas sebenarnya, baik dalam klasifikasi dua kelas maupun klasifikasi dengan lebih dari dua kelas.

Dalam penelitian ini, klasifikasi dilakukan ke dalam tiga kategori, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Oleh karena itu, perhitungan komponen evaluasi dilakukan dengan melihat hasil prediksi pada

masing-masing kelas. Menurut (Putra et al., 2022), komponen utama dalam *confusion matrix* yaitu:

1. True Positive (TP), yakni total data yang terprediksi sebagai suatu kelas dan hasil prediksinya sesuai dengan kelas sebenarnya.
2. False Positive (FP), yakni total data yang teridentifikasi sebagai bagian dari suatu kelas, namun sebenarnya masuk ke dalam kelas yang berbeda.
3. True Negative (TN), yakni total data yang terprediksi bukan termasuk suatu kelas dan memang benar bukan termasuk bagian dalam kelas tersebut.
4. False Negative (FN), yakni total data yang sebenarnya masuk dalam suatu kelas, namun luaran prediksinya dimasukkan ke dalam kelas lain.

Berdasarkan nilai TP, FP, TN, dan FN, beberapa metrik evaluasi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Accuracy

Accuracy ialah indikator yang difungsikan guna mengukur derajat ketepatan model secara keseluruhan. Besaran *accuracy* dikalkulasikan melalui komparasi antara total prediksi yang benar dengan keseluruhan data yang diuji. Persamaan *accuracy* dapat dinyatakan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Metrik ini mudah digunakan untuk melihat performa model secara umum. Namun, *accuracy* kurang cukup apabila digunakan pada dataset yang memiliki ketidakseimbangan distribusi kelas.

2. Precision

Precision ialah indikator yang dimanfaatkan guna menilai presisi model dalam mengategorikan suatu kelas. Besaran *precision* dikalkulasikan dari rasio antara total data yang terprediksi secara tepat pada suatu kelas dengan keseluruhan data yang dimasukkan ke dalam kelas tersebut. Persamaan *precision* dapat diformulasikan sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Semakin tinggi nilai *precision*, semakin rendah total data yang sebenarnya milik kelas lain, namun terprediksi sebagai bagian dari kelas tersebut.

3. Recall

Recall ialah indikator yang dimanfaatkan guna menilai kapasitas model dalam mendeteksi data yang sejatinya masuk ke dalam suatu kelas. Nilai *recall* dihitung berdasarkan perbandingan antara data yang berhasil diprediksi sesuai dengan kelas sebenarnya dan keseluruhan data aktual pada kelas tersebut. Rumus *recall* dapat dituliskan sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Semakin tinggi nilai *recall* yang diperoleh, semakin baik kinerja model dalam mengidentifikasi data yang termasuk ke dalam kelas tertentu.

4. F1-Score

F1-score ialah indikator evaluasi yang difungsikan guna mengombinasikan besaran *precision* dan *recall*. Metrik ini menghasilkan performa yang lebih proporsional bagi capaian model, khususnya saat muncul selisih antara nilai *precision* dan *recall*. Rumus *F1-score* dapat dituliskan sebagai berikut:

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

Pada penelitian ini, indikator *accuracy*, *precision*, *recall*, dan *F1-score* diterapkan guna menilai ketangguhan algoritma *Support Vector Machine* (SVM) dalam mengategorikan komentar TikTok ke dalam kelompok topik keresahan karir.

2.8. Sistem Berbasis Web

Sistem berbasis web adalah suatu aplikasi atau sistem informasi yang dapat dijalankan menggunakan jaringan internet dan diakses melalui peramban web (*web browser*) (Sintaro, 2022). Web berfungsi sebagai media penyedia informasi yang tersusun dalam berbagai halaman digital dan memuat beragam konten, seperti teks, gambar, dan video, sehingga memungkinkan pengguna untuk mengakses informasi secara fleksibel tanpa batasan ruang dan waktu selama semuanya terhubung dengan internet (Rahmi et al., 2023). Selain itu, menurut (Hadi & Rokhman, 2020), fungsi web secara umum adalah sebagai berikut:

1. Fungsi komunikasi, yaitu web bersifat dinamis dan memungkinkan terjadinya interaksi antara pengguna dan pengelola.
2. Fungsi informasi, yaitu web menyajikan konten yang berfokus pada penyampaian informasi secara jelas dan efektif.
3. Fungsi hiburan (*entertainment*), yaitu web menyediakan konten visual atau interaktif untuk tujuan hiburan.
4. Fungsi transaksi, yaitu web digunakan sebagai sarana transaksi bisnis, baik barang maupun jasa secara daring.

2.9. Penelitian Terdahulu

Penelitian terdahulu dimanfaatkan sebagai referensi untuk memperkuat landasan teori, mendukung pemilihan metode, serta menunjukkan posisi penelitian yang dilakukan. Beberapa penelitian yang berkaitan dengan klasifikasi teks, penerapan algoritma SVM, dan analisis data media sosial akan diuraikan pada Tabel 2.2.

Tabel 2.2. Penelitian Terdahulu

NO	Peneliti & Tahun	Judul Penelitian	Objek penelitian	Hasil Penelitian
1	Abdurrazik, I Made Widhi Wirawana (2025)	Analisis Klasifikasi Tweet Berdasarkan Topik Sosial Menggunakan SVM	Tweet Twitter	Penggunaan SVM dalam pengelompokan tweet berdasarkan topik sosial menghasilkan akurasi sebesar 84%. Hasil tersebut menunjukkan bahwa SVM dapat diterapkan pada data teks media sosial yang memiliki karakteristik singkat dan tidak terstruktur.
2	Sola Fide, Suparti, Sudarno (2021)	Analisis Sentimen Ulasan Aplikasi TikTok di Google Play Menggunakan Metode SVM	Ulasan TikTok Google Play	Penerapan SVM dengan fitur berbobot TF-IDF menghasilkan akurasi tertinggi sebesar 90,62% dalam melakukan klasifikasi sentimen.
3	Enda Esyudha Pratama, Bambang Riyanto Trilaksono (2015)	Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode <i>Support Vector Machine</i> (SVM)	Tweet Twitter	Setelah dilakukan penggabungan fitur, model SVM mampu mengelompokkan topik keluhan pelanggan dengan akurasi sebesar 89,17%.

4	Lutfi Budi Ilmawan, Muhammad Aliyazid Mude (2020)	Perbandingan Metode Klasifikasi <i>Support Vector Machine</i> dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store	Ulasan Google Play Store	SVM memperoleh akurasi 81,46%, lebih tinggi dibandingkan Naïve Bayes yang memperoleh 75,41%. Hasil perbandingan tersebut menunjukkan kinerja SVM yang lebih baik pada data ulasan yang digunakan dalam penelitian
5	Anne Mudya Yolanda, Ridho Tri Mulya (2024)	Implementasi Metode <i>Support Vector Machine</i> untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store	Ulasan aplikasi Sayurbox pada Google Play Store	Klasifikasi menggunakan SVM menghasilkan akurasi sebesar 89,29% pada data ulasan pengguna aplikasi Sayurbox. Hasil tersebut menunjukkan bahwa SVM dapat digunakan untuk mengklasifikasikan data berbentuk ulasan pengguna.

Berdasarkan Tabel 2.2, terlihat bahwa algoritma SVM telah diimplementasikan pada berbagai studi terkait klasifikasi teks dan mencapai performa yang baik. Riset yang dilakukan oleh (Abdurrazik & Wirawana, 2025), (Fide et al., 2021), (Pratama & Trilaksono, 2015), (Ilmawan & Mude, 2025), serta (Yolanda & Mulya, 2024) menunjukkan bahwa SVM mampu memperoleh tingkat akurasi yang tinggi pada berbagai jenis data teks, baik berupa tweet media sosial maupun ulasan pengguna aplikasi. Di samping itu, beberapa studi juga menggunakan pembobotan TF-IDF untuk membentuk representasi fitur teks yang lebih kaya informasi agar menunjang tahapan klasifikasi secara lebih optimal.

Meskipun demikian, mayoritas studi terdahulu masih berfokus pada analisis sentimen maupun klasifikasi topik secara umum, dengan basis data yang

berasal dari Twitter serta ulasan aplikasi di Google Play Store. Penelitian yang memanfaatkan komentar TikTok sebagai sumber data untuk mengklasifikasikan topik keresahan karir masih relatif terbatas.

Berdasarkan uraian tersebut, penelitian ini berfokus pada klasifikasi topik keresahan karir berdasarkan komentar publik di TikTok dengan menerapkan *Natural Language Processing* (NLP), TF-IDF, serta *Support Vector Machine* (SVM). Topik yang diangkat meliputi Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Di samping itu, model klasifikasi yang dirancang bakal diintegrasikan ke dalam sistem berbasis web agar keluaran klasifikasi dapat disajikan secara lebih terstruktur dan mudah dipahami.

BAB III

ANALISA DAN PERANCANGAN SISTEM

3.1. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif yang didukung oleh penerapan *text mining* dan algoritma *machine learning*. Pemilihan pendekatan tersebut disesuaikan dengan tujuan penelitian, yaitu mengolah data berupa teks secara sistematis serta mengukur kemampuan model dalam melakukan klasifikasi berdasarkan nilai metrik evaluasi yang diperoleh.

Teknik *text mining* digunakan untuk mengolah dan memperoleh informasi dari komentar pengguna TikTok yang berkaitan dengan keresahan karir. Selanjutnya, metode machine learning berupa *Support Vector Machine* (SVM) digunakan untuk membangun model klasifikasi yang bertugas mengelompokkan komentar ke dalam kategori topik keresahan karir yang telah ditetapkan.

Dalam penelitian ini, topik keresahan karir dikelompokkan menjadi tiga kategori, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Hasil pengelompokan tersebut selanjutnya dimanfaatkan sebagai acuan dalam proses perancangan dan penerapan sistem berbasis web.

3.2. Objek dan Subjek Penelitian

Objek pada penelitian ini adalah topik keresahan karir yang diekspresikan oleh pengguna TikTok melalui kolom komentar. Keresahan karir yang dianalisis mencakup tiga kategori utama, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja.

Subjek atau sumber data pada penelitian ini adalah komentar pengguna TikTok berbahasa Indonesia yang relevan dengan isu karir dan dunia kerja.

Komentar yang dianalisis berasal dari video TikTok dengan tema pekerjaan, pengalaman kerja, beban kerja, lingkungan kerja, gaji, dan kompensasi. Data komentar tersebut kemudian diseleksi dan diberi label sesuai dengan kategori yang telah ditetapkan.

Dataset penelitian terdiri dari 305 komentar TikTok berbahasa Indonesia. Sebelum digunakan, komentar terlebih dahulu melalui proses seleksi dan pelabelan berdasarkan kategori yang telah ditetapkan. Data yang telah selesai dipersiapkan selanjutnya digunakan untuk melatih dan menguji model klasifikasi dengan algoritma *Support Vector Machine* (SVM).

TikTok dipilih sebagai sumber data karena memiliki tingkat interaksi yang tinggi dan memungkinkan pengguna menyampaikan pengalaman, pendapat, serta keresahan secara spontan. Karakteristik tersebut menjadikan komentar TikTok relevan sebagai sumber data untuk mengamati bentuk-bentuk ekspresi keresahan karir yang muncul secara alami di ruang digital.

3.3. Tahapan Penelitian

Tahapan penelitian disusun untuk menggambarkan rangkaian pelaksanaan penelitian, mulai dari identifikasi masalah hingga tahap perancangan sistem berbasis web. Penelitian ini dilakukan secara sistematis agar setiap proses, mulai dari pengumpulan data, pengolahan teks, pembangunan model klasifikasi, hingga perancangan sistem berbasis web dapat berjalan secara terarah.

Tahapan penelitian yang dilakukan meliputi beberapa proses sebagai berikut:

1. Identifikasi Permasalahan

Tahaapan ini dilakukan untuk mengamati fenomena keresahan karir yang banyak diungkapkan oleh pengguna TikTok melalui kolom komentar. Permasalahan utama yang diangkat adalah bagaimana komentar TikTok dapat dianalisis dan diklasifikasikan ke dalam topik keresahan karir secara otomatis dengan menerapkan pendekatan *Natural Language Processing* (NLP) serta algoritma *Support Vector Machine* (SVM).

2. Studi Literatur

Studi literatur dilakukan dengan menelaah berbagai sumber yang berkaitan dengan permasalahan dan metode yang digunakan dalam penelitian. Literatur yang dikaji mencakup konsep keresahan karir, media sosial TikTok, klasifikasi teks, NLP, TF-IDF, SVM, evaluasi model klasifikasi, serta sistem berbasis web. Hasil studi literatur digunakan sebagai dasar dalam menentukan kategori klasifikasi dan metode yang digunakan.

3. Pengumpulan Data

Pengumpulan data dilakukan dengan memperoleh komentar dari beberapa video TikTok yang membahas topik dunia kerja dan karir. Video dipilih berdasarkan kata kunci yang relevan, seperti pengalaman kerja, beban kerja, lingkungan kerja, gaji, kompensasi, dan keresahan karir. Komentar yang diperoleh selanjutnya diseleksi untuk memastikan kesesuaiannya dengan konteks penelitian.

4. Pelabelan Data

Tahapan pelabelan dilakukan dengan memberikan kategori pada setiap komentar berdasarkan topik keresahan karir yang telah ditetapkan. Kategori tersebut meliputi Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja.

Pelabelan dilakukan secara manual dengan memperhatikan makna utama yang terkandung dalam setiap komentar.

5. Pre-processing Data

Data komentar yang telah dikumpulkan selanjutnya dipersiapkan melalui tahap *pre-processing*. Tahap ini bertujuan untuk membersihkan teks dari unsur yang tidak diperlukan sekaligus menyeragamkan bentuk data sebelum masuk ke proses ekstraksi fitur. Proses *pre-processing* yang diterapkan meliputi *cleansing*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Setelah seluruh tahapan selesai, data teks yang diperoleh digunakan sebagai masukan pada tahap ekstraksi fitur.

6. Ekstraksi Fitur

Tahap ekstraksi fitur dilakukan dengan menerapkan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini digunakan untuk mengubah data teks hasil *pre-processing* menjadi bentuk numerik dengan memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam kumpulan dokumen. Hasil pembobotan tersebut selanjutnya menjadi representasi data yang digunakan oleh model klasifikasi.

7. Pembangunan Model Klasifikasi

Setelah data melalui proses pelabelan dan ekstraksi fitur, tahap berikutnya adalah membangun model klasifikasi menggunakan algoritma Support Vector Machine (SVM). Data yang telah direpresentasikan dalam bentuk bobot TF-IDF dibagi menjadi data latih dan data uji. Data latih digunakan untuk membentuk model, sedangkan data uji digunakan untuk mengetahui

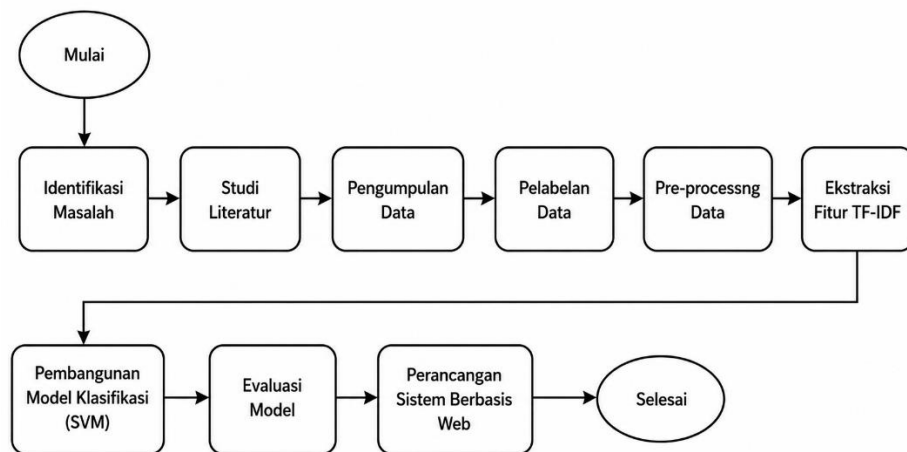
kemampuan model dalam mengklasifikasikan data yang belum digunakan pada proses pelatihan.

8. Evaluasi Model

Pengujian dilakukan untuk mengetahui tingkat kinerja model SVM dalam mengelompokkan komentar TikTok ke dalam kategori keresahan karir yang telah ditentukan. Penilaian terhadap hasil klasifikasi dilakukan menggunakan *confusion matrix* dan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai dari setiap metrik digunakan untuk melihat kemampuan model dalam menghasilkan klasifikasi yang tepat.

9. Perancangan Sistem Berbasis Web

Pada tahap perancangan, sistem berbasis web disusun sebagai sarana untuk menerapkan model yang digunakan dalam proses klasifikasi topik keresahan karir. Kegiatan yang dilakukan mencakup identifikasi kebutuhan sistem, penyusunan *use case diagram* dan *flowchart*, pembuatan *Entity Relationship Diagram (ERD)*, serta rancangan tampilan dalam bentuk *wireframe*. Tahap ini bertujuan untuk menentukan gambaran proses sistem, hubungan antar data dalam basis data, dan rancangan tampilan yang akan digunakan sebagai acuan pada proses pengembangan sistem.



Gambar 3.1 Tahapan Penelitian

3.4. Teknik Pengumpulan Data

Data penelitian menggunakan data primer yang bersumber dari komentar pengguna TikTok yang dikumpulkan secara langsung oleh peneliti. Pemilihan data dilakukan berdasarkan keterkaitannya dengan topik keresahan karir yang menjadi fokus penelitian. Komentar tersebut kemudian diseleksi dan diberi label secara manual sesuai dengan kategori penelitian.

Pengumpulan data dilakukan dengan mengambil beberapa komentar dari video TikTok yang membahas isu dunia kerja dan karir. Video dipilih berdasarkan kesesuaian topik dengan fokus penelitian, seperti pengalaman kerja, beban kerja, lingkungan kerja, gaji, kompensasi, lembur, serta ketidaknyamanan dalam pekerjaan. Penentuan data penelitian dilakukan dengan metode *purposive sampling*. Melalui metode ini, data dipilih secara selektif dengan mengacu pada sejumlah kriteria yang telah ditetapkan sesuai dengan tujuan dan kebutuhan penelitian.

Proses pencarian video dilakukan menggunakan kata kunci yang relevan, seperti “kerja capek”, “beban kerja”, “lingkungan kerja toxic”, “gaji kecil”,

“lembur tidak dibayar”, dan “keresahan kerja”. Kata kunci tersebut digunakan sebagai panduan awal untuk menemukan video yang berkaitan dengan isu keresahan karir. Setelah video ditemukan, komentar pada video tersebut dibaca dan diseleksi secara manual untuk memastikan kesesuaiannya dengan konteks penelitian.

Komentar yang digunakan pada penelitian ini adalah komentar berbahasa Indonesia yang mengandung ekspresi keresahan terkait karir dan dunia kerja. Komentar yang tidak relevan, bersifat spam, hanya berisi emotikon, promosi, atau tidak memiliki konteks yang jelas tidak disertakan dalam dataset. Selain itu, data yang digunakan hanya berupa isi komentar tanpa mencantumkan identitas pribadi pengguna TikTok.

Jumlah data yang digunakan dalam penelitian ini adalah 305 komentar TikTok berbahasa Indonesia yang relevan dengan isu keresahan karir. Pengumpulan data dilakukan secara langsung oleh peneliti pada periode April hingga Mei 2026 melalui komentar pada video TikTok yang relevan dengan topik penelitian. Jumlah tersebut dipilih karena dinilai cukup untuk menggambarkan variasi komentar pada tiga kategori klasifikasi, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Selain itu, jumlah tersebut masih memungkinkan proses seleksi, pelabelan manual, *pre-processing*, serta pengujian model dilakukan secara terkontrol sesuai dengan ruang lingkup penelitian.

Setelah seluruh komentar terkumpul, setiap data diklasifikasikan secara manual ke dalam salah satu dari tiga kategori yang telah ditentukan, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Dataset yang telah melalui proses pelabelan tersebut kemudian dipisahkan menjadi dua bagian menggunakan

perbandingan 80:20. Bagian pertama, sebesar 80%, digunakan sebagai data latih untuk membentuk model klasifikasi, sedangkan 20% lainnya digunakan sebagai data uji untuk mengetahui kemampuan model dalam mengklasifikasikan data yang tidak digunakan selama proses pelatihan. Contoh data komentar sebelum melalui proses pengolahan ditampilkan pada bagian berikut:

Tabel 3.1 Contoh Data Mentah

No	Komentar
1	atasan nya baik tapi partner kerja nya yg toxic sama ajaa bikin kita ga betah 😞
2	paling sedih gaji sebulan cuma 400k ya allah udh mah cairnya bisa 3 bulan sekali
3	Ngerasain bgt jd honorer dgn gaji 300k sebulan
4	Kerja tiap hari lemburan capek bgt 😞
5	percayalah. udh ditahap mentok pgn bgt risegn krna lingkungannya udah g sehat bnyk nangisnya.

Komentar pada tabel ditampilkan sesuai bentuk asli dari sumber data, termasuk penggunaan singkatan, bahasa tidak baku, dan emotikon. Data yang telah dikumpulkan dan diseleksi kemudian digunakan pada tahap pelabelan data.

3.5. Teknik Pelabelan Data

Pelabelan data dilakukan dengan mengelompokkan setiap komentar sesuai dengan topik keresahan karir yang diteliti. Pelabelan bertujuan membentuk dataset yang telah memiliki kategori sehingga dapat digunakan pada tahap pembentukan serta pengujian model klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

Setiap komentar dianalisis secara manual dengan memperhatikan isi dan konteks yang disampaikan oleh pengguna. Pemberian kategori dilakukan dengan

melihat makna yang terkandung dalam komentar secara keseluruhan. Hal ini dilakukan agar label yang diberikan benar-benar sesuai dengan topik keresahan yang paling dominan dalam komentar tersebut.

Dalam penelitian ini, komentar yang telah melalui proses pelabelan dibedakan ke dalam tiga kategori utama. Ketiga kategori tersebut ditetapkan berdasarkan topik keresahan karir yang menjadi fokus penelitian, yaitu:

1. Beban Kerja, yaitu komentar yang berkaitan dengan tuntutan pekerjaan, tekanan kerja, beban tugas yang berlebihan, lembur, target kerja, serta kelelahan fisik atau mental akibat pekerjaan.
2. Lingkungan Kerja, yaitu komentar yang berkaitan dengan kondisi tempat kerja, hubungan dengan atasan atau rekan kerja, budaya organisasi, konflik kerja, serta suasana kerja yang tidak nyaman.
3. Kompensasi Kerja, yaitu komentar yang berkaitan dengan gaji, upah, tunjangan, keterlambatan pembayaran, atau ketidaksesuaian antara imbalan dan beban pekerjaan.

Penelitian ini menerapkan pendekatan *single-label classification*, sehingga setiap komentar hanya memperoleh satu label berdasarkan topik yang paling dominan. Apabila suatu komentar mengandung lebih dari satu aspek keresahan, maka label ditentukan berdasarkan inti masalah yang paling menonjol dalam komentar tersebut.

Komentar yang ambigu, tidak memiliki konteks yang jelas, tidak relevan dengan isu karir, bersifat spam, atau tidak sesuai dengan kategori yang telah ditetapkan tidak disertakan dalam dataset. Langkah tersebut bertujuan untuk

memastikan kualitas dataset yang digunakan pada tahap pelatihan dan pengujian model.

Untuk memberikan gambaran proses pelabelan yang dilakukan, berikut disajikan contoh hasil pelabelan data :

Tabel 3.2 Contoh Pelabelan Data

NO	Komentar	Label
1	atasan nya baik tapi patner kerja nya yg toxic sama ajaa bikin kita ga betah 🤔	Lingkungan Kerja
2	paling sedih gaji sebulan cuma 400k ya allah udh mah cairnya bisa 3 bulan sekali	Kompensasi Kerja
3	Ngerasain bgt jd honorer dg gaji 300k sebulan	Kompensasi Kerja
4	Kerja tiap hari lemburan capek bgt 🤔	Beban Kerja
5	percayalah. udh ditahap mentok pgn bgt risegn krna lingkungannya udah g sehat bnyk nangisnya	Lingkungan kerja

Komentar pada tabel ditampilkan sesuai bentuk asli dari sumber data, termasuk penggunaan singkatan, bahasa tidak baku, dan emotikon. Data yang telah diberi label kemudian diproses pada tahap *pre-processing* data sebelum dilakukan ekstraksi fitur dan pembangunan model klasifikasi.

3.6. Pre-processing Data Teks

Tahap *pre-processing* dilakukan untuk menyiapkan data teks sebelum proses ekstraksi fitur dan klasifikasi. Tahap ini mencakup pembersihan teks, penyeragaman bentuk penulisan, serta pengurangan *noise* agar data lebih siap digunakan dalam proses klasifikasi.

Data komentar TikTok umumnya memiliki karakteristik bahasa yang tidak baku, seperti penggunaan singkatan, emotikon, simbol, tanda baca berlebihan, serta variasi penulisan kata. Oleh karena itu, data yang telah melalui tahap

pelabelan perlu diproses terlebih dahulu sehingga menghasilkan data yang terstruktur dan dapat diproses dalam tahap klasifikasi.

Tahap *pre-processing* dalam penelitian ini terdiri atas beberapa proses berikut:

1. Cleansing, dilakukan dengan membuang karakter yang tidak diperlukan, seperti angka, tanda baca, simbol, dan emotikon.
2. Case Folding, dilakukan dengan menyeragamkan seluruh huruf menjadi huruf kecil.
3. Normalisasi, dilakukan dengan mengubah kata tidak baku, singkatan, atau istilah tertentu menjadi bentuk yang sesuai.
4. Tokenizing, dilakukan dengan memisahkan teks menjadi bagian-bagian berupa kata atau token.
5. Stopword Removal, dilakukan dengan menghilangkan kata umum yang tidak memberikan informasi penting dalam proses klasifikasi.
6. Stemming, dilakukan dengan mengubah kata berimbuhan menjadi bentuk kata dasarnya.

Untuk memberikan gambaran proses pre-processing data teks, berikut disajikan contoh tahapan pengolahan komentar:

Tabel 3.3 Tahap Pre-processing

Tahapan	Hasil
Data Awal	Kerja tiap hari lemburan capek bgt 🤖
Cleansing	Kerja tiap hari lemburan capek bgt
Case Folding	kerja tiap hari lemburan capek bgt
Normalisasi	kerja tiap hari lemburan capek banget
Tokenizing	[kerja, tiap, hari, lemburan, capek, banget]

Stopword Removal	[kerja, lemburan, capek, banget]
Stemming	[kerja, lembur, capek, banget]

Berdasarkan Tabel 3.3, komentar yang awalnya masih mengandung emotikon dan kata tidak baku diproses melalui beberapa tahapan hingga menghasilkan bentuk teks yang lebih bersih. Hasil dari tahap pre-processing selanjutnya digunakan pada proses ekstraksi fitur dengan metode TF-IDF.

3.7. Ekstraksi Fitur TF – IDF

Tahap ekstraksi fitur dilakukan terhadap data yang telah selesai melalui tahap *pre-processing*. Tujuannya adalah mengubah teks menjadi representasi numerik sehingga dapat diproses oleh algoritma *Support Vector Machine* (SVM).

Penelitian ini menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk memberikan bobot pada setiap kata. Pembobotan mempertimbangkan frekuensi kata dalam suatu komentar serta kemunculannya pada keseluruhan dataset. Kata yang sering muncul pada satu komentar tetapi relatif jarang ditemukan pada komentar lain akan memiliki bobot yang lebih tinggi.

Hasil pembobotan TF-IDF direpresentasikan sebagai vektor numerik. Setiap komentar menjadi satu baris, sedangkan kata-kata yang terdapat dalam dataset menjadi fitur pada kolom. Matriks yang dihasilkan kemudian digunakan sebagai masukan dalam proses pelatihan model SVM.

Implementasi TF-IDF dilakukan menggunakan Python dengan memanfaatkan *library scikit-learn*. Tahapan ekstraksi fitur yang diterapkan meliputi:

1. memasukkan data teks hasil pre-processing

2. membentuk daftar kosakata (*vocabulary*) dari seluruh komentar
3. menghitung bobot TF-IDF untuk setiap kata
4. mengubah teks menjadi matriks numerik (*document-term matrix*)

Matriks TF-IDF yang dihasilkan kemudian digunakan sebagai masukan untuk proses pelatihan dan pengujian model klasifikasi dengan algoritma SVM..

3.8. Pembangunan Model Klasifikasi

Setelah fitur teks diperoleh melalui pembobotan TF-IDF, tahap berikutnya adalah membangun model klasifikasi. Data yang telah berbentuk matriks numerik digunakan sebagai masukan dalam proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

SVM digunakan untuk mempelajari hubungan antara fitur teks dan label yang telah diberikan pada setiap komentar. Penelitian ini menggunakan tiga kategori, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Dengan adanya tiga kategori tersebut, permasalahan yang dihadapi termasuk ke dalam klasifikasi multi-kelas.

Dataset terlebih dahulu dipisahkan menjadi data latih dan data uji dengan rasio 80:20. Sebagian besar data, yaitu 80%, digunakan untuk membentuk model, sedangkan 20% lainnya disisihkan untuk menguji hasil klasifikasi. Pembagian ini dilakukan agar model memperoleh data yang cukup untuk mempelajari karakteristik setiap kategori, sekaligus menyediakan data yang berbeda untuk mengetahui kemampuan model dalam melakukan prediksi.

Pada proses klasifikasi, SVM diterapkan dengan menggunakan *kernel* linear. Penggunaan *kernel* tersebut disesuaikan dengan karakteristik data teks yang telah diubah menjadi fitur TF-IDF dan memiliki jumlah fitur yang relatif

banyak. Selain itu, *kernel* linear memiliki proses komputasi yang lebih sederhana sehingga sesuai digunakan dalam klasifikasi teks. Parameter regularisasi yang diterapkan adalah $C = 1.0$, yang berfungsi mengatur keseimbangan antara margin pemisah dan kesalahan klasifikasi pada data latih.

Proses pembelajaran dilakukan menggunakan data latih yang telah direpresentasikan melalui TF-IDF. Model kemudian mempelajari pola fitur yang terdapat pada masing-masing kategori. Setelah model terbentuk, data uji diberikan sebagai masukan untuk menghasilkan prediksi kelas.

Prediksi yang diperoleh selanjutnya dibandingkan dengan label aktual pada data uji. Pengukuran kinerja model dilakukan menggunakan *confusion matrix* dan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengukuran tersebut menjadi dasar untuk mengetahui kemampuan SVM dalam mengklasifikasikan topik keresahan karir pada komentar TikTok.

3.9. Evaluasi Model

Tahap evaluasi dilakukan untuk menilai kinerja SVM setelah model menghasilkan prediksi pada data uji. Evaluasi ini digunakan untuk melihat sejauh mana model mampu mengklasifikasikan komentar TikTok sesuai dengan kategori keresahan karir yang telah ditentukan.

Pada penelitian ini, *confusion matrix* digunakan untuk mengevaluasi hasil klasifikasi dengan membandingkan label prediksi model terhadap label aktual pada data uji. Evaluasi ini menunjukkan jumlah prediksi yang benar maupun kesalahan klasifikasi pada setiap kategori. Selain memberikan gambaran kinerja model secara keseluruhan, *confusion matrix* juga memperlihatkan ketepatan klasifikasi pada tiga kategori yang digunakan, yaitu Beban Kerja, Lingkungan

Kerja, dan Kompensasi Kerja. Dari hasil tersebut dapat diketahui data yang berhasil ditempatkan pada kategori yang sesuai maupun data yang justru diprediksi ke kategori lainnya.

Kinerja model dianalisis berdasarkan hasil *confusion matrix* dengan menggunakan empat ukuran evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk melihat persentase prediksi yang benar dari seluruh data uji. *Precision* menggambarkan ketepatan model dalam memberikan suatu label terhadap data yang diprediksi pada kategori tertentu. Selanjutnya, *recall* menunjukkan kemampuan model dalam mengenali data yang sebenarnya termasuk dalam suatu kategori. Sementara itu, *F1-score* merupakan ukuran gabungan yang digunakan untuk melihat keseimbangan antara nilai *precision* dan *recall*.

Karena model yang digunakan termasuk klasifikasi multi-kelas, nilai *precision*, *recall*, dan *F1-score* diperoleh untuk masing-masing kategori. Nilai dari setiap kategori kemudian dirata-ratakan menggunakan metode *macro average*, sehingga setiap kelas memiliki pengaruh yang sama dalam penilaian. Dengan cara ini, hasil evaluasi tidak hanya dipengaruhi oleh jumlah data pada kelas tertentu. Nilai evaluasi tersebut digunakan untuk menilai kemampuan SVM dalam mengklasifikasikan komentar TikTok ke dalam kategori keresahan karir.

3.10. Perancangan Sistem Berbasis Web

Tahap perancangan sistem berbasis web dilakukan sebagai bentuk penerapan dari model klasifikasi yang telah dibangun dengan memanfaatkan algoritma *Support Vector Machine* (SVM). Sistem ini dirancang untuk

memudahkan admin dalam melakukan klasifikasi komentar TikTok ke dalam kategori topik keresahan karir secara otomatis dan terstruktur.

Sistem berbasis web yang dirancang memiliki fungsi utama untuk mengelola dataset, melakukan pelatihan model, mengevaluasi model, menyimpan dan mengaktifkan model, melakukan input komentar, menjalankan proses klasifikasi, serta menampilkan hasil klasifikasi. Model klasifikasi yang telah selesai dilatih kemudian diintegrasikan ke dalam sistem sehingga dapat digunakan untuk memprediksi kategori pada komentar baru yang dimasukkan oleh admin.

Secara umum, sistem bekerja dengan menerima input berupa komentar teks. Komentar tersebut kemudian diproses melalui tahapan *pre-processing* data, ekstraksi fitur dengan menggunakan TF-IDF, kemudian diklasifikasikan menggunakan SVM. Hasil dari proses tersebut selanjutnya ditampilkan oleh sistem dalam bentuk kategori topik keresahan karir, yaitu Beban Kerja, Lingkungan Kerja, atau Kompensasi Kerja.

Perancangan sistem tidak hanya berfokus pada tampilan antarmuka, tetapi juga pada perancangan alur proses serta pengelolaan data yang mendukung proses klasifikasi. Setiap komponen sistem dirancang agar saling terintegrasi sehingga proses mulai dari pengelolaan dataset, pelatihan model, evaluasi, hingga klasifikasi komentar dapat berjalan secara sistematis. Melalui perancangan yang dilakukan, proses implementasi sistem diharapkan dapat dilakukan secara lebih sistematis serta mampu memenuhi kebutuhan pengguna sesuai dengan tujuan penelitian

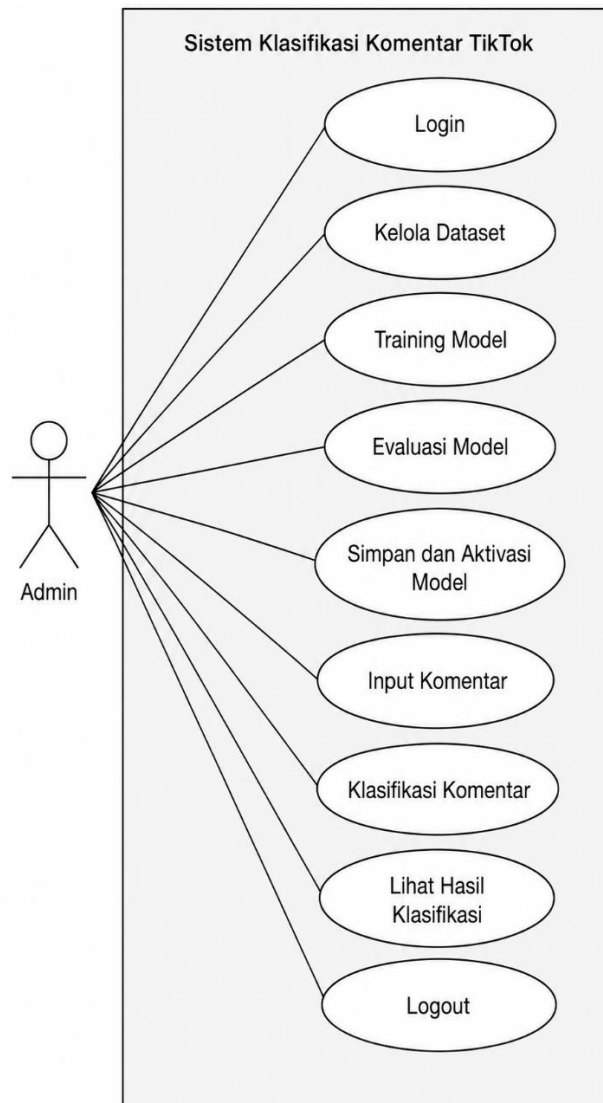
Perancangan sistem bertujuan memberikan gambaran mengenai kebutuhan sistem, struktur basis data, alur proses, serta hubungan interaksi antara admin dan

sistem. Dalam penelitian ini, perancangan sistem menggunakan *use case diagram*, *flowchart*, dan *Entity Relationship Diagram* (ERD). *Use case diagram* berfungsi untuk menggambarkan fitur utama yang dapat diakses oleh admin, *flowchart* digunakan untuk menggambarkan tahapan proses dalam sistem, sedangkan ERD berfungsi untuk menggambarkan struktur database beserta hubungan antarentitas dalam sistem.

3.10.1. Perancangan Sistem

Perancangan sistem bertujuan menjelaskan alur kerja serta hubungan admin dengan berbagai fitur yang terdapat dalam sistem. Sistem yang dirancang memiliki satu aktor utama, yaitu admin. Admin memiliki hak akses untuk mengelola dataset, melakukan pelatihan model, melihat evaluasi model, memasukkan komentar, melakukan klasifikasi komentar, melihat hasil klasifikasi, serta keluar dari sistem.

Interaksi admin terhadap sistem dituangkan dalam bentuk *use case diagram*. Diagram ini menunjukkan peran admin serta fungsi-fungsi utama yang dapat diakses dalam sistem.



Gambar 3.2 Use Case Diagram

Berdasarkan Gambar 3.2, admin dapat mengakses beberapa fitur utama pada sistem, yaitu login, kelola dataset, training model, evaluasi model, simpan dan aktivasi model, input komentar, klasifikasi komentar, lihat hasil klasifikasi, dan logout. Fitur login digunakan sebagai proses awal untuk masuk ke dalam sistem. Fitur pengelolaan dataset digunakan untuk mengatur data komentar yang menjadi bahan dalam proses pelatihan model.

Fitur training model berfungsi untuk membentuk model klasifikasi berdasarkan dataset yang telah disiapkan. Hasil kinerja model setelah proses

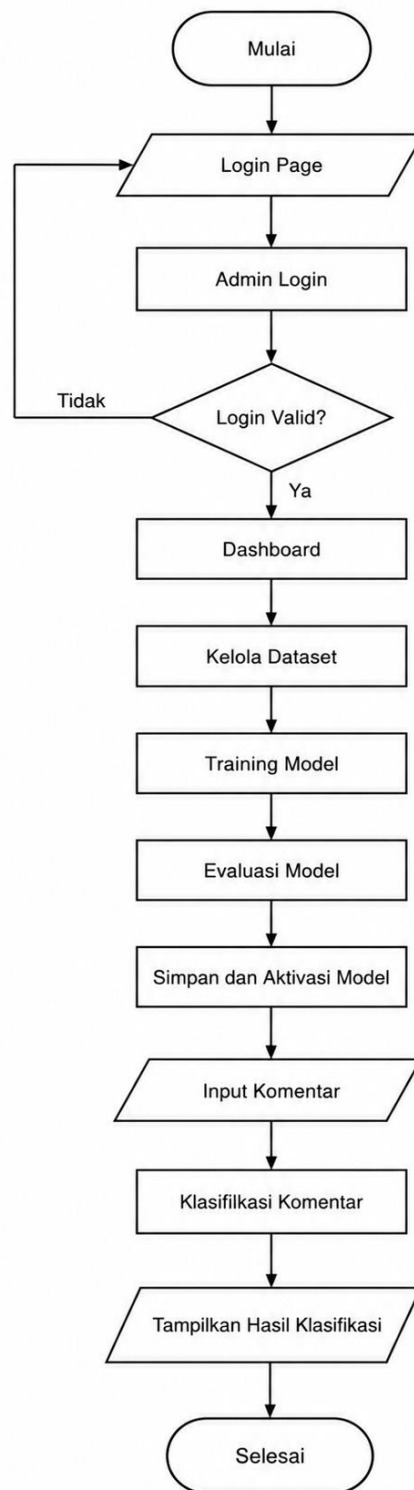
pelatihan dapat dilihat melalui fitur evaluasi model. Model yang telah dinilai layak kemudian dapat disimpan dan diaktifkan agar dapat digunakan dalam proses klasifikasi komentar baru.

Fitur input komentar digunakan untuk memasukkan komentar yang ingin diklasifikasikan. Setelah komentar dimasukkan, sistem menjalankan proses klasifikasi komentar menggunakan model SVM yang telah aktif. Hasil klasifikasi kemudian dapat dilihat melalui fitur lihat hasil klasifikasi. Fitur logout digunakan untuk keluar dari sistem dan mengakhiri sesi penggunaan.

Seluruh fungsi yang dimiliki admin dirancang untuk mendukung proses klasifikasi komentar secara menyeluruh. Setiap fitur memiliki peran yang saling berkaitan, mulai dari pengelolaan data sebagai masukan sistem, proses pembentukan model klasifikasi, hingga penggunaan model untuk mengklasifikasikan komentar baru. Dengan demikian, seluruh aktivitas yang dilakukan admin dapat berlangsung dalam satu sistem yang terintegrasi.

Melalui perancangan ini, hubungan antara admin dengan setiap fungsi sistem dapat diidentifikasi secara jelas. Hal tersebut bertujuan agar implementasi sistem sesuai dengan kebutuhan pengguna serta seluruh fitur dapat digunakan sebagaimana mestinya tanpa adanya proses yang terlewatkan.

Selain *use case diagram*, alur kerja sistem juga digambarkan menggunakan *flowchart*. *Flowchart* digunakan untuk menjelaskan tahapan proses sistem secara berurutan, dimulai dari admin membuka halaman login hingga sistem menampilkan hasil klasifikasi.



Gambar 3.3 Flowchart Sistem

Berdasarkan Gambar 3.3, proses sistem dimulai dari tahap mulai, kemudian admin diarahkan ke halaman login. Pada halaman login, admin memasukkan data

akun untuk dapat mengakses sistem. Sistem akan memeriksa data login yang dimasukkan oleh admin. Apabila data tersebut tidak sesuai, sistem mengembalikan admin ke halaman login. Sebaliknya, apabila data login valid, admin dapat mengakses halaman dashboard.

Pada halaman dashboard, admin dapat mengakses fitur kelola dataset. Dataset yang telah tersedia digunakan untuk proses training model. Pada tahap training model, sistem akan melatih model klasifikasi menggunakan data data komentar yang telah melalui proses pelabelan dan telah melalui proses pengolahan data. Setelah proses training selesai, Sistem kemudian mengevaluasi model untuk mengukur kinerja klasifikasi yang telah dibangun.

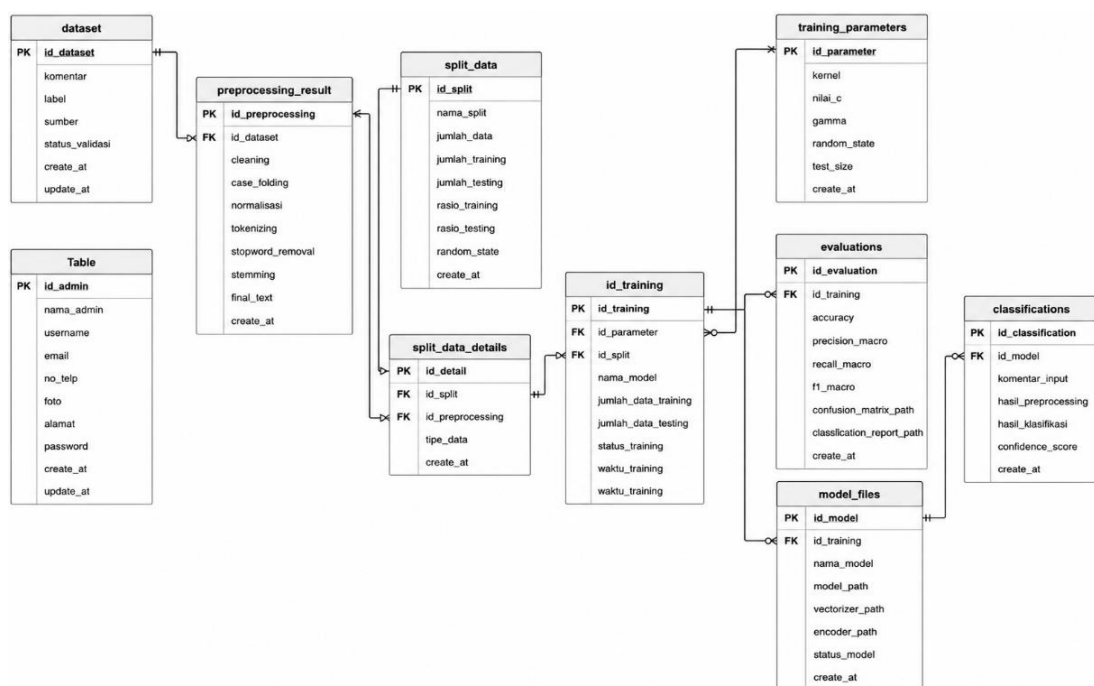
Hasil evaluasi digunakan untuk menilai kelayakan model sebelum diterapkan pada proses klasifikasi. Model yang telah melalui proses evaluasi kemudian disimpan dan diaktifkan agar dapat digunakan untuk mengklasifikasikan komentar baru. Setelah model aktif, admin dapat memasukkan komentar melalui fitur input komentar.

Komentar yang dimasukkan akan diproses oleh sistem melalui *tahapan pre-processing* data, ekstraksi fitur TF-IDF, dan klasifikasi menggunakan SVM. Setelah proses klasifikasi selesai, sistem menampilkan kategori yang diperoleh dari hasil prediksi terhadap komentar. Proses berakhir setelah hasil klasifikasi ditampilkan kepada admin.

Perancangan basis data bertujuan mengatur penyimpanan data yang dibutuhkan untuk mendukung proses klasifikasi dalam sistem. Data yang disimpan meliputi data admin, dataset komentar, label komentar, model yang digunakan, serta hasil klasifikasi. Dengan adanya basis data, informasi pada

sistem dapat dikelola secara lebih terstruktur serta dapat digunakan kembali sesuai kebutuhan.

Struktur dan keterkaitan antarentitas pada basis data digambarkan menggunakan *Entity Relationship Diagram* (ERD). ERD menunjukkan entitas yang terdapat dalam sistem, atribut pada setiap entitas, serta hubungan antarentitas yang mendukung pengelolaan data komentar, *pre-processing* teks, pembagian data, pelatihan model, evaluasi model, penyimpanan file model, hingga klasifikasi topik keresahan karir menggunakan SVM. Penggunaan ERD membantu memberikan gambaran yang lebih jelas mengenai struktur basis data sehingga dapat mendukung proses perancangan dan implementasi sistem. ERD sistem yang diusulkan disajikan pada Gambar 3.4.



Gambar 3.4 Entity Relationship Diagram (ERD)

Berdasarkan Gambar 3.4, sistem memiliki beberapa entitas utama, yaitu *admin*, *dataset*, *preprocessing_result*, *split_data*, *split_data_details*,

training_parameters, *training*, *evaluations*, *model_files*, dan *classifications*. Masing-masing entitas memiliki fungsi yang saling berkaitan dalam mendukung pengelolaan dataset, *pre-processing* komentar, pembagian data, pelatihan dan evaluasi model, penyimpanan model, serta proses klasifikasi topik keresahan karir menggunakan algoritma *Support Vector Machine* (SVM).

Entitas *admin* berfungsi menyimpan data akun pengguna yang memiliki wewenang untuk mengelola sistem. Data yang disimpan mencakup nama admin, *username*, email, nomor telepon, foto, alamat, *password*, serta waktu pembuatan dan pembaruan data. Entitas ini mendukung kebutuhan proses login dan pengelolaan sistem oleh admin.

Entitas *dataset* digunakan untuk menyimpan data komentar TikTok yang menjadi dataset penelitian. Data yang disimpan meliputi isi komentar, label topik, sumber data, status validasi, serta waktu pembuatan dan pembaruan data. Entitas ini menjadi data awal yang akan digunakan dalam proses preprocessing dan pelatihan model klasifikasi.

Entitas *preprocessing_result* digunakan untuk menyimpan hasil pengolahan teks dari setiap komentar pada dataset. Data tahapan yang dicatat meliputi *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, *stemming*, hingga menghasilkan *final text*. Entitas ini berelasi dengan dataset karena setiap komentar yang diproses akan menghasilkan data hasil preprocessing.

Entitas *split_data* digunakan untuk menyimpan informasi pembagian data training dan testing. Informasi yang dicatat meliputi nama *split*, total data, jumlah data latih dan data uji, rasio pembagian, *random state*, serta waktu pembentukan

data. Entitas ini digunakan untuk mencatat skenario pembagian data sebelum proses pelatihan model dilakukan.

Entitas `split_data_details` digunakan untuk menyimpan detail data hasil pembagian dari proses split data. Entitas ini memiliki relasi dengan `split_data` dan `preprocessing_result`. Data yang disimpan meliputi id split, id preprocessing, tipe data, dan waktu pembuatan data. Tipe data digunakan untuk membedakan apakah data tersebut termasuk data training atau data testing.

Entitas `training_parameters` digunakan untuk menyimpan parameter yang digunakan dalam proses pelatihan model SVM. Data yang disimpan meliputi kernel, nilai C, gamma, random state, test size, dan waktu pembuatan data. Entitas ini membantu sistem mencatat konfigurasi model yang digunakan pada proses training.

Entitas `training` digunakan untuk menyimpan riwayat proses pelatihan model klasifikasi. Entitas ini berelasi dengan `training_parameters` dan `split_data`. Data yang disimpan meliputi nama model, jumlah data training, jumlah data testing, status training, waktu training, serta durasi proses training. Entitas ini menjadi pusat pencatatan proses pelatihan model SVM pada sistem.

Entitas `evaluations` digunakan untuk menyimpan hasil evaluasi dari model yang telah dilatih. Data yang disimpan meliputi *accuracy*, *precision macro*, *recall macro*, *F1 macro*, *path confusion matrix*, *path classification report*, dan waktu pembuatan data. Entitas ini berelasi dengan `training` karena setiap hasil evaluasi berasal dari proses training tertentu.

Entitas `model_files` digunakan untuk menyimpan informasi file model yang dihasilkan dari proses training. Data yang disimpan meliputi nama model, path

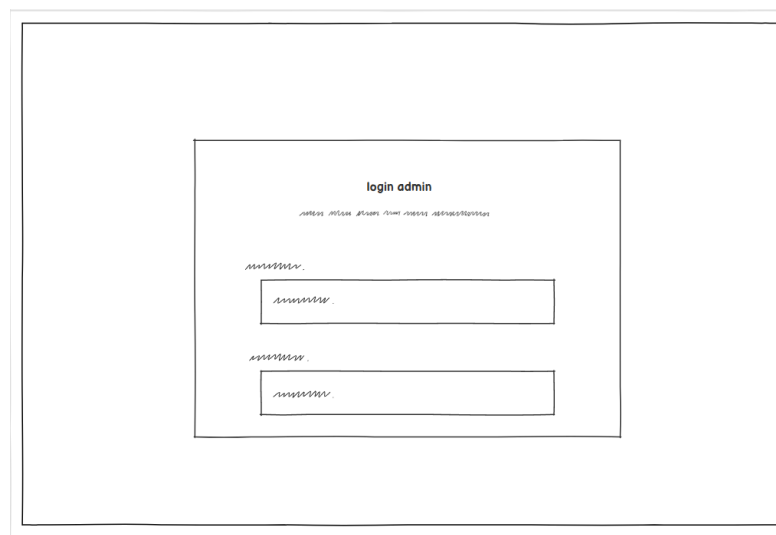
model, path vectorizer, path encoder, status model, dan waktu pembuatan data. Model yang berstatus aktif akan digunakan oleh sistem untuk melakukan klasifikasi komentar baru.

Entitas `classifications` digunakan untuk menyimpan hasil klasifikasi komentar yang dimasukkan ke dalam sistem. Data yang disimpan meliputi komentar input, hasil preprocessing, hasil klasifikasi, dan waktu pembuatan data. Entitas ini berelasi dengan `model_files` karena proses klasifikasi dilakukan menggunakan model yang telah tersimpan dan aktif pada sistem.

3.10.2. Perancangan Antarmuka Sistem

Perancangan antarmuka bertujuan memberikan gambaran awal mengenai tampilan dan susunan fitur pada sistem yang akan dibangun. Perancangan ini dibuat dalam bentuk wireframe untuk menunjukkan struktur halaman, tata letak komponen, dan fungsi dari setiap fitur yang tersedia pada sistem.

a. Wireframe Login



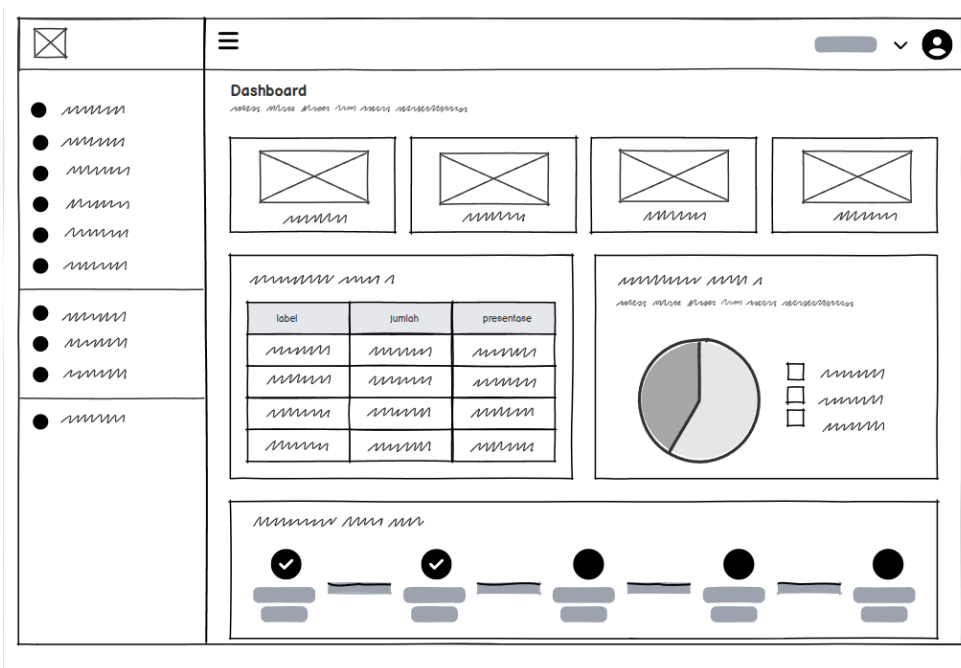
Gambar 3.5 Wireframe Login

Wireframe halaman login dirancang sebagai tampilan awal bagi admin untuk melakukan autentikasi sebelum masuk ke dalam sistem. Halaman ini

menjadi awal sebelum pengguna dapat menggunakan seluruh fitur yang tersedia pada sistem klasifikasi topik keresahan karir. Halaman ini menyediakan form autentikasi yang memuat kolom untuk memasukkan username dan password. Admin diwajibkan memasukkan data akun yang valid untuk dapat masuk ke dalam sistem. Tata letak komponen dibuat sederhana agar proses login dapat dilakukan dengan mudah dan cepat.

Setelah data login dimasukkan, sistem akan melakukan verifikasi. Jika kredensial yang dimasukkan cocok dengan data pada basis data, admin dapat masuk ke halaman *dashboard*. Namun, jika terjadi ketidaksesuaian, sistem memberikan notifikasi kesalahan dan admin perlu melakukan login kembali.

b. Wireframe Dashboard



Gambar 3.6 Wireframe Dashboard

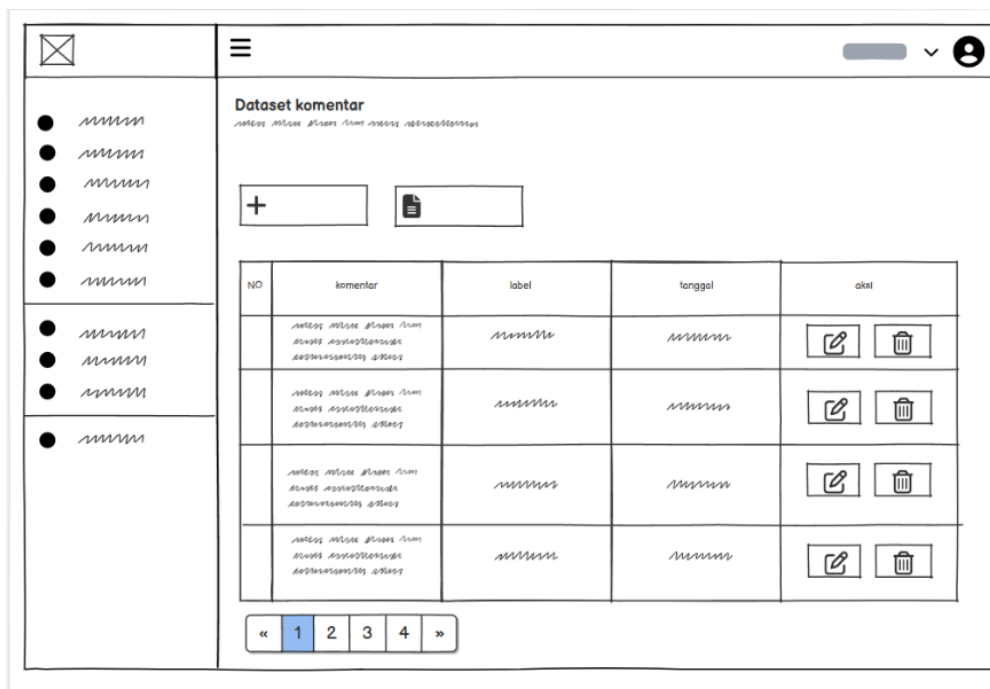
Halaman dashboard menjadi tampilan utama yang dapat diakses admin setelah berhasil melakukan login ke dalam sistem. Halaman ini berfungsi

menampilkan ringkasan informasi sekaligus menyediakan akses ke berbagai fitur yang tersedia.

Pada bagian kiri terdapat sidebar yang berisi menu navigasi sistem. Bagian atas menampilkan informasi akun admin yang sedang aktif. Sementara itu, bagian utama dashboard menampilkan ringkasan data seperti jumlah dataset, jumlah hasil klasifikasi, serta informasi lain yang berkaitan dengan aktivitas sistem.

Selain menampilkan informasi statistik, dashboard juga dilengkapi dengan grafik distribusi kategori topik keresahan karir dan aktivitas terbaru yang dilakukan dalam sistem. Informasi yang ditampilkan memudahkan admin dalam memantau data serta hasil klasifikasi yang diperoleh.

c. Wireframe Dataset

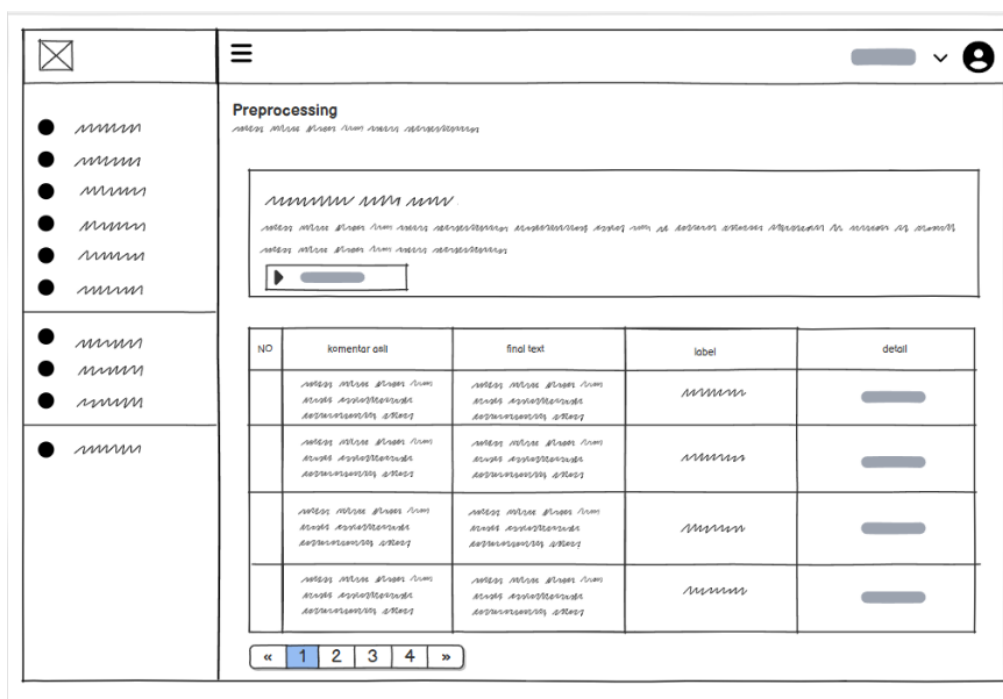


Gambar 3.7 Wireframe Dataset

Halaman dataset digunakan untuk mengelola data komentar TikTok yang menjadi sumber data penelitian. Melalui halaman ini admin dapat melakukan pengelolaan data sebelum digunakan dalam proses pelatihan model. Data

komentar ditampilkan dalam bentuk tabel yang berisi informasi komentar, label kategori, tanggal data, serta tombol aksi. Admin memiliki akses untuk menambah, memperbarui, dan menghapus data sesuai kebutuhan sistem. Selain itu, halaman ini menyediakan fitur pencarian serta pagination yang membantu pengguna dalam menelusuri dan mengelola data dalam jumlah besar sehingga proses pengolahan dataset menjadi lebih efisien.

d. Wireframe Preprocessing



Gambar 3.8 Wireframe Preprocessing

Halaman *preprocessing* digunakan untuk menampilkan hasil pengolahan data teks sebelum dilakukan proses klasifikasi. Tahap ini dilakukan untuk mengolah dan menyiapkan data agar sesuai untuk digunakan dalam proses klasifikasi. Pada halaman tersebut, admin dapat melihat data komentar asli beserta hasil *preprocessing* yang telah dilakukan oleh sistem. Proses *pre-processing* menghasilkan data yang telah melalui *cleansing*, *case folding*, normalisasi,

tokenisasi, *stopword removal*, dan *stemming*. Melalui halaman ini, admin dapat mengetahui perubahan pada data komentar setelah melewati setiap tahapan *preprocessing*, sehingga kualitas data yang digunakan untuk pelatihan model dapat terlihat dengan lebih jelas.

e. Wireframe Pembagian Data

Pembagian Data

contoh data training

NO	final text	label

contoh data testing

NO	final text	label

Gambar 3.9 Wireframe Pembagian Data

Halaman pembagian data digunakan untuk membagi dataset menjadi data *training* dan data *testing* sebelum model dilatih. Pembagian data dilakukan untuk memperoleh penilaian kinerja model yang lebih objektif. Pada halaman ini ditampilkan informasi jumlah data yang digunakan beserta proporsi pembagian data. Data dalam sistem dibagi dengan proporsi 80% sebagai data latih dan 20% sebagai data uji. Selain informasi pembagian data, halaman ini juga menampilkan contoh data yang masuk ke dalam kelompok training dan testing sehingga admin dapat melihat hasil pembagian data yang dilakukan oleh sistem.

f. Wireframe Training Model

training model svm

parameter training

algoritma	~~~~~
fitur	~~~~~
nilai c	~~~~~
fitur	~~~~~

▶

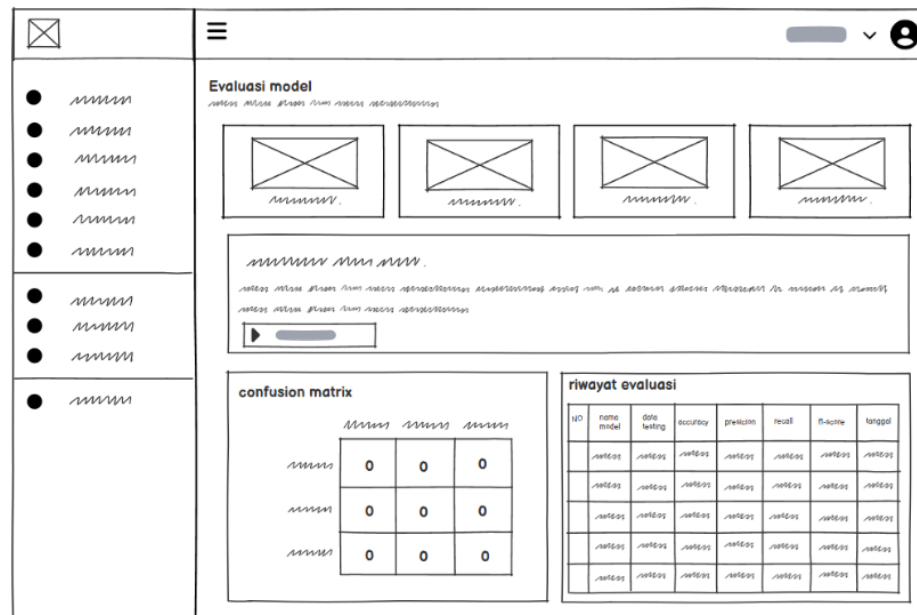
contoh data testing

NO	nama model	data training	data testing	status	waktu training	tanggal
~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~

Gambar 3.10 Wireframe Training Model

Halaman *training model* digunakan untuk menjalankan proses pelatihan model klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Proses tersebut bertujuan menghasilkan model yang dapat digunakan untuk mengklasifikasikan komentar. Pada halaman ini ditampilkan informasi mengenai metode ekstraksi fitur TF-IDF, algoritma yang digunakan, serta parameter pelatihan model. Admin dapat menjalankan proses training melalui tombol yang telah disediakan. Selain itu, sistem juga menampilkan riwayat pelatihan model yang berisi informasi nama model, jumlah data training, jumlah data testing, status model, serta waktu pelatihan yang telah dilakukan.

#### g. Wireframe Evaluasi Model

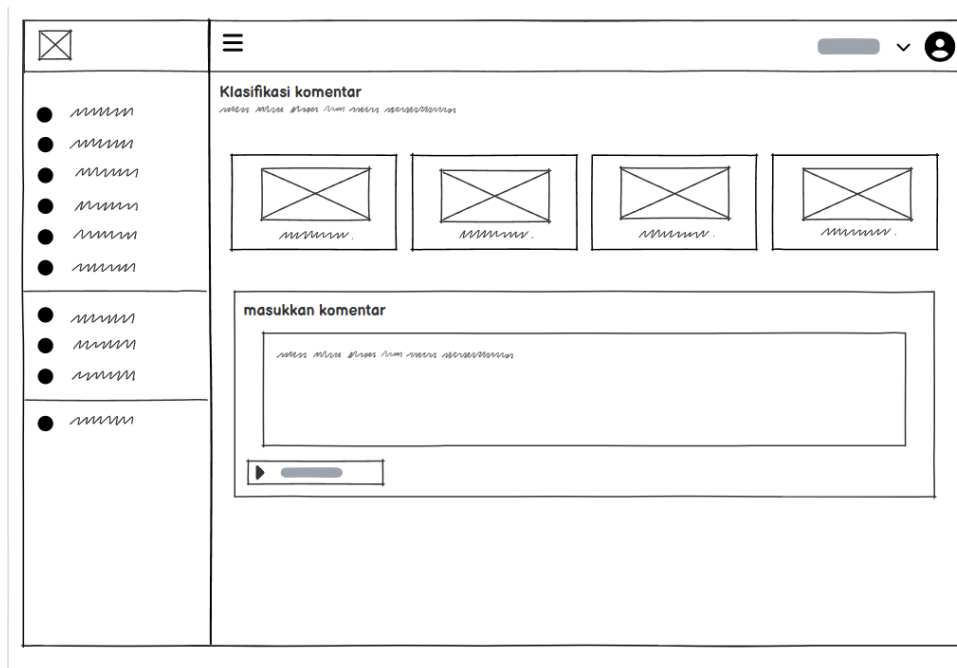


Gambar 3.11 Wireframe Evaluasi Model

Halaman evaluasi model digunakan untuk menampilkan hasil pengujian model setelah proses pelatihan. Halaman ini digunakan untuk menilai kemampuan model dalam mengklasifikasikan data komentar. Pada halaman ini ditampilkan *confusion matrix* yang menggambarkan hasil prediksi model pada setiap kategori. Sistem juga menyajikan nilai *accuracy*, *precision*, *recall*, dan *F1-score* untuk menunjukkan performa model dalam melakukan klasifikasi.

Informasi hasil evaluasi tersebut membantu admin dalam menentukan apakah model yang telah dihasilkan layak digunakan untuk proses klasifikasi komentar baru. Melalui hasil evaluasi tersebut, admin dapat mengetahui tingkat keberhasilan model dalam mengenali masing-masing kategori komentar. Informasi ini menjadi dasar untuk menentukan apakah model telah memiliki performa yang memadai untuk diimplementasikan pada proses klasifikasi komentar baru di dalam sistem.

#### h. Wireframe Klasifikasi Komentar



Gambar 3.12 Wireframe Klasifikasi Komentar

Halaman klasifikasi komentar merupakan fitur utama yang digunakan untuk melakukan prediksi terhadap komentar baru. Halaman ini memungkinkan admin untuk menguji model klasifikasi secara langsung melalui sistem berbasis web. Admin dapat memasukkan komentar ke dalam kolom yang tersedia. Setelah tombol klasifikasi ditekan, sistem memproses komentar melalui tahap *pre-processing*. Data tersebut kemudian diubah menjadi fitur menggunakan TF-IDF dan selanjutnya diklasifikasikan dengan model SVM yang sedang aktif. Hasil klasifikasi kemudian ditampilkan dalam bentuk kategori topik keresahan karir sesuai dengan model yang digunakan.

### i. Wireframe Riwayat Komentar

**Riwayat komentar**

Riwayat komentar

Riwayat komentar

Riwayat komentar

Riwayat komentar

Riwayat komentar

Riwayat komentar

Riwayat komentar

NO	komentar	hasil preprocessing	hasil klasifikasi	confidence	model	tanggal	aksi

Gambar 3.13 Wireframe Riwayat Komentar

Halaman riwayat komentar digunakan untuk menyimpan dan menampilkan seluruh hasil klasifikasi yang pernah dilakukan oleh sistem. Halaman ini berfungsi sebagai dokumentasi hasil prediksi yang dapat diakses kembali oleh admin. Data ditampilkan dalam bentuk tabel yang memuat informasi komentar, hasil preprocessing, kategori hasil klasifikasi, model yang digunakan, serta tanggal klasifikasi. Informasi tersebut memudahkan admin dalam melakukan penelusuran terhadap hasil klasifikasi yang telah dilakukan sebelumnya. Halaman ini juga menyediakan fitur pencarian dan penyaringan data untuk memudahkan admin menemukan riwayat klasifikasi secara lebih cepat.

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1. Hasil Pengumpulan Data**

Sumber data penelitian berasal dari komentar publik di TikTok. Data dikumpulkan dengan menelusuri video yang relevan dengan topik keresahan karir melalui fitur pencarian pada aplikasi tersebut. Kata kunci yang digunakan dalam proses pencarian antara lain “kerja capek”, “beban kerja”, “lingkungan kerja toxic”, “gaji kecil”, “lembur tidak dibayar”, dan kata kunci lain yang berhubungan dengan permasalahan dunia kerja.

Berdasarkan hasil pencarian tersebut, ditemukan beberapa video yang relevan dengan topik penelitian. Salah satu video yang digunakan sebagai sumber data adalah video TikTok dengan tautan <https://vt.tiktok.com/ZSQay3PQy/>. Video tersebut dipilih karena memiliki pembahasan yang sesuai dengan keresahan karir, khususnya mengenai tekanan dan tuntutan dalam pekerjaan.

Komentar yang dikumpulkan terlebih dahulu diseleksi berdasarkan kesesuaiannya dengan topik penelitian. Komentar yang tidak relevan, bersifat spam, promosi, atau tidak memiliki makna yang jelas tidak digunakan dalam dataset. Selain dari video tersebut, data juga diperoleh dari beberapa video lain yang memiliki pembahasan serupa mengenai pengalaman, pendapat, dan keluhan pengguna TikTok terhadap dunia kerja.

Pengumpulan data dilakukan secara bertahap pada periode April 2026 sampai Mei 2026. Dari hasil seleksi, sebanyak 305 komentar terpilih untuk digunakan sebagai dataset penelitian. Data tersebut kemudian disusun dalam

bentuk dataset dan dilanjutkan ke tahap pelabelan serta pre-processing sebelum digunakan dalam proses klasifikasi menggunakan metode SVM..

#### 4.2. Atribut Data

Komentar yang telah terkumpul kemudian disusun dalam bentuk dataset agar lebih mudah dikelola dan digunakan pada tahap penelitian berikutnya. Penyusunan data ke dalam format dataset dilakukan untuk mempermudah proses pelabelan, pre-processing, pelatihan model, serta pengujian model klasifikasi.

Dataset dalam penelitian ini terdiri atas beberapa atribut utama, yaitu nomor data, komentar, dan label. Atribut komentar menjadi data utama yang diproses melalui tahapan *Natural Language Processing*, sedangkan atribut label berfungsi sebagai target klasifikasi dalam proses pelatihan model *Support Vector Machine*. Adapun atribut yang digunakan dalam penelitian ini disajikan pada Tabel 4.1.

Tabel 4.1. Atribut Data

Atribut	Penjelasan
No	Atribut yang digunakan sebagai nomor urut data. Atribut ini berfungsi untuk mempermudah identifikasi setiap komentar dalam dataset.
Komentar	Atribut yang berisi teks komentar pengguna TikTok. Atribut ini merupakan data utama yang akan dianalisis dan diproses dalam penelitian.
Label	Atribut yang berisi kategori dari komentar berdasarkan topik keresahan karir. Atribut ini digunakan sebagai target klasifikasi dalam proses pembangunan model.

Berdasarkan atribut tersebut, data komentar yang telah dikumpulkan akan lebih terstruktur. Atribut komentar akan diproses pada tahap pre-processing data, sedangkan atribut label digunakan sebagai acuan dalam proses pelatihan dan pengujian model klasifikasi.

#### 4.3. Hasil Pelabelan Data

Setelah komentar dikumpulkan dan disusun menjadi dataset, tahap berikutnya adalah pelabelan data. Pelabelan dilakukan secara manual dengan membaca setiap komentar untuk memahami isi serta maksud utama yang disampaikan. Tujuan dari tahap ini adalah untuk menentukan kategori topik keresahan karir yang paling sesuai pada setiap komentar.

Dalam proses pelabelan, penentuan label tidak hanya dilakukan berdasarkan kemunculan kata tertentu, tetapi juga memperhatikan konteks kalimat secara keseluruhan. Kata-kata seperti “capek”, “lembur”, “gaji”, “atasan”, atau “toxic” digunakan sebagai petunjuk awal, namun keputusan akhir tetap dilihat dari makna utama komentar. Hal ini dilakukan agar komentar dapat diberi label sesuai dengan permasalahan yang paling dominan.

Penelitian ini menggunakan tiga kategori label, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Secara umum, komentar yang menunjukkan adanya tekanan pekerjaan, pekerjaan yang terlalu banyak, lembur, target kerja, atau kelelahan akibat pekerjaan diberi label Beban Kerja. Komentar yang berkaitan dengan kondisi tempat kerja, hubungan dengan atasan atau rekan kerja, serta suasana kerja yang tidak nyaman diberi label Lingkungan Kerja. Sementara itu, komentar yang membahas gaji, upah, tunjangan, atau ketidaksesuaian antara imbalan dan pekerjaan yang dilakukan diberi label Kompensasi Kerja.

Apabila dalam satu komentar terdapat lebih dari satu permasalahan, maka label ditentukan berdasarkan topik yang paling dominan. Misalnya, komentar yang membahas pekerjaan yang banyak dan gaji yang kecil akan dilihat dari fokus

utama kalimatnya. Jika komentar lebih menekankan pada banyaknya pekerjaan dan rasa lelah, maka komentar diberi label Beban Kerja. Namun, jika komentar lebih menekankan pada gaji yang tidak sebanding, maka komentar diberi label Kompensasi Kerja.

Hasil proses pelabelan menghasilkan tiga kategori data sesuai dengan kelas yang telah ditetapkan. Rincian jumlah komentar pada setiap kategori ditampilkan pada Tabel 4.2.

Tabel 4.2. Distribusi Label Dataset

Label	Jumlah Data
Beban Kerja	96
Lingkungan Kerja	102
Kompensasi Kerja	107
Total	305

Berdasarkan Tabel 4.2, diperoleh 107 komentar dengan label Kompensasi Kerja, 102 komentar dengan label Lingkungan Kerja, dan 96 komentar dengan label Beban Kerja. Perbedaan jumlah data pada masing-masing kategori menunjukkan bahwa komentar yang berkaitan dengan kompensasi kerja lebih banyak ditemukan dibandingkan kategori lainnya. Kondisi ini menunjukkan bahwa isu mengenai gaji, upah, tunjangan, dan ketidaksesuaian antara beban kerja dengan imbalan merupakan topik yang cukup sering dibahas oleh pengguna TikTok dalam konteks keresahan karir. Sementara itu, jumlah komentar pada kategori Beban Kerja dan Lingkungan Kerja juga tidak berbeda jauh, sehingga distribusi data pada ketiga kategori tergolong cukup seimbang. Kondisi tersebut dapat mendukung proses pelatihan model karena setiap kategori memiliki jumlah data yang cukup untuk dipelajari oleh algoritma *Support Vector Machine*.

#### 4.4. Hasil Pre-processing Data

Setelah tahap pelabelan selesai, data komentar selanjutnya diproses melalui tahap *pre-processing*. Tahap ini bertujuan mengolah komentar mentah menjadi teks yang lebih terstruktur, bersih, dan seragam sebelum digunakan pada proses selanjutnya. Hasil *pre-processing* tersebut kemudian digunakan sebagai masukan sebelum dilakukan pembentukan bobot pada setiap kata menggunakan TF-IDF dan proses klasifikasi dengan metode SVM.

Pada sistem yang dikembangkan, komentar diproses secara bertahap melalui *cleaning*, *case folding*, normalisasi, *stopword removal*, dan *stemming*. Seluruh komentar dalam dataset mengikuti rangkaian proses tersebut hingga diperoleh teks akhir yang lebih ringkas dan siap digunakan pada tahap berikutnya.

##### 4.4.1 Cleaning

Pada tahap *cleaning*, sistem memproses komentar dengan menghapus bagian-bagian teks yang tidak diperlukan. Komentar yang awalnya masih mengandung tanda baca, simbol, emoji, tautan, angka tertentu, atau karakter khusus dibersihkan agar hanya menyisakan kata-kata yang dibutuhkan dalam proses analisis.

Sebagai contoh, komentar yang mengandung tanda seru, emotikon, atau tanda baca berlebihan akan diubah menjadi bentuk teks yang lebih bersih. Tahap ini membantu sistem mengurangi gangguan dari karakter yang tidak berhubungan langsung dengan topik komentar.

##### 4.4.2. Case Folding

Setelah komentar dibersihkan, sistem akan menyeragamkan seluruh huruf pada komentar menjadi huruf kecil agar penulisan kata menjadi konsisten. Pada

tahap ini, kata yang sebelumnya ditulis dengan huruf besar atau campuran akan disamakan bentuknya. Misalnya, kata “Kerja”, “KERJA”, dan “kerja” akan diproses menjadi “kerja”. Dengan demikian, kata yang sama tidak akan dibaca oleh sistem sebagai bentuk kata yang berbeda hanya karena perbedaan antara huruf besar dan kecil.

#### **4.4.3 Normalisasi**

Tahap berikutnya adalah normalisasi. Pada tahap ini, sistem mengganti kata kurang baku, singkatan, serta penulisan yang keliru menjadi bentuk kata yang lebih baku. Proses ini dilakukan karena komentar pada TikTok umumnya menggunakan bahasa informal dan banyak mengandung singkatan.

Contohnya, kata “gk”, “ga”, atau “nggak” diseragamkan menjadi “tidak”. Kata “yg” diubah menjadi “yang”, sedangkan kata yang ditulis berlebihan seperti “lemburr” disesuaikan menjadi “lembur”. Dengan proses ini, kata yang mengandung makna serupa, meskipun dituliskan dalam bentuk yang berbeda dapat diseragamkan penggunaannya.

Perbaikan pada kamus normalisasi juga membantu sistem dalam menyeragamkan variasi kata yang sering muncul pada komentar TikTok. Hal ini membuat hasil *pre-processing* menjadi lebih seragam dan konsisten sebelum dilanjutkan ke tahap pembobotan TF-IDF.

#### **4.4.4 Stopword Removal**

Setelah tahap normalisasi dilakukan, sistem menghilangkan kata-kata yang kurang relevan dalam menentukan kategori komentar. Kata-kata yang bersifat umum dan sering muncul dalam kalimat dihapus sehingga teks yang dihasilkan lebih menekankan kata-kata yang relevan untuk proses klasifikasi.

Pada tahap ini, kata umum seperti “yang”, “di”, “ke”, “dan”, serta “itu” dihilangkan apabila tidak memberikan pengaruh besar terhadap topik komentar. Setelah tahap ini dilakukan, komentar menjadi lebih ringkas dan berisi kata-kata yang lebih berkaitan dengan kategori klasifikasi, seperti “lembur”, “gaji”, “capek”, “atasan”, “toxic”, dan kata penting lainnya.

#### 4.4.5 Stemming

*Stemming* merupakan tahap akhir dalam *pre-processing* yang digunakan untuk mengubah kata berimbuhan menjadi kata dasar. Proses ini dilakukan agar kata-kata yang memiliki dasar sama dapat diseragamkan sehingga lebih mudah diproses pada tahap klasifikasi.

Sebagai contoh, kata “bekerja”, “pekerjaan”, dan “dikerjakan” dapat diproses menjadi bentuk dasar “kerja”. Dengan adanya tahap stemming, hasil *pre-processing* menjadi lebih sederhana karena kata-kata yang bentuknya berbeda tetapi memiliki makna dasar yang sama dapat disatukan.

Setelah seluruh proses *pre-processing* selesai, diperoleh data komentar yang lebih bersih dan terstruktur untuk selanjutnya diproses menggunakan pembobotan TF-IDF. Hasil *pre-processing* tidak selalu berbentuk kalimat yang sempurna, karena tahap ini lebih berfokus pada pengambilan kata-kata penting yang dapat membantu proses klasifikasi. Ringkasan hasil perubahan komentar sebelum dan sesudah *pre-processing* dapat dilihat pada Tabel 4.3.

Tabel 4.3 Hasil *Pre-processing* Data

No	Komentar Awal	Hasil Pre-processing Data
1	atasan nya baik tapi patner kerja nya yg toxic sama ajaa bikin kita ga betah 🤔	atasan baik rekan kerja toxic bikin tidak betah
2	kerja full time + capek fisik ga di gaji lagi 😞	kerja full time capek fisik tidak gaji
3	tp otak ini serasa mau pecah dibawah	otak serasa pecah tekan

	tekanan trs2 peraturan yg berubah2 capek	terus atur ubah capek
4	Gaji kecil tapi tuntutan kerja banyak banget	gaji kecil tuntutan kerja banyak
5	bener banget, beban kerja yg berat banget dan gk ada habisnya 🤖	benar beban kerja berat tidak habis

Berdasarkan Tabel 4.3, dapat diketahui bahwa komentar awal yang masih mengandung kata tidak baku, tanda baca, dan bentuk kata yang belum seragam telah diproses menjadi teks yang lebih sederhana. Meskipun hasil pre-processing tidak selalu membentuk kalimat lengkap, kata-kata utama yang berkaitan dengan topik komentar tetap dipertahankan.

#### 4.5. Hasil Pembobotan TF-IDF

Data komentar yang telah melalui *pre-processing* kemudian masuk ke tahap pembobotan menggunakan TF-IDF. Proses ini mengubah teks menjadi nilai numerik yang dapat digunakan sebagai fitur dalam proses klasifikasi, karena SVM membutuhkan data dalam bentuk numerik.

Pada sistem yang dikembangkan, proses pembobotan dilakukan secara otomatis menggunakan *library TfidfVectorizer* pada Python. Setiap komentar diolah menjadi representasi vektor berdasarkan bobot kata yang terdapat dalam dataset. Penentuan bobot tersebut mempertimbangkan frekuensi kemunculan suatu kata pada komentar (*Term Frequency/TF*) dan tingkat kemunculan kata tersebut pada keseluruhan dokumen (*Inverse Document Frequency/IDF*).

Hasil dari proses TF-IDF berupa matriks fitur yang digunakan sebagai data masukan pada proses pelatihan model SVM. Matriks tersebut tidak ditampilkan secara langsung pada halaman sistem, tetapi digunakan oleh sistem dalam proses pembentukan model klasifikasi. Dengan demikian, TF-IDF berperan sebagai tahap yang menghubungkan hasil preprocessing dengan proses training model.

Untuk memberikan gambaran mengenai proses pembobotan TF-IDF yang dilakukan oleh sistem, berikut disajikan contoh perhitungan manual menggunakan salah satu komentar hasil preprocessing. Perhitungan manual ini hanya digunakan untuk menjelaskan proses pembentukan bobot TF-IDF, sedangkan pada implementasi sistem seluruh proses dilakukan secara otomatis menggunakan library *TfidfVectorizer*. Komentar yang digunakan sebagai contoh:

“atasan baik rekan kerja toxic bikin tidak betah”

Tahap pertama adalah menghitung nilai *Term Frequency* (TF). Karena setiap kata pada komentar tersebut hanya muncul satu kali, maka nilai TF masing-masing kata adalah 1.

Tabel 4.4 Perhitungan Term Frequency (TF)

Kata	TF
atasan	1
baik	1
rekan	1
kerja	1
toxic	1
bikin	1
tidak	1
betah	1

Tahap berikutnya adalah menghitung nilai *Inverse Document Frequency* (IDF). Perhitungan ini menggunakan jumlah seluruh dokumen dalam dataset ( $N = 305$ ) dan jumlah dokumen yang mengandung kata tertentu (*Document Frequency/DF*). Rumus yang digunakan untuk memperoleh nilai IDF adalah sebagai berikut:

$$IDF = \log\left(\frac{N}{DF}\right)$$

Keterangan:

$N$  = jumlah seluruh dokumen dalam dataset, yaitu 305 komentar.

$DF$  jumlah dokumen yang memuat kata tertentu.

Sebagai contoh, kata *betah* muncul pada 4 dokumen, sehingga perhitungannya adalah sebagai berikut.

$$\begin{aligned} IDF &= \log\left(\frac{305}{4}\right) \\ &= \log(76,25) \\ &= 1,8822 \end{aligned}$$

Nilai akhir TF-IDF kemudian diperoleh melalui perhitungan dengan rumus berikut:

$$TF-IDF = TF \times IDF$$

Karena nilai TF kata *betah* adalah 1, maka:

$$TF-IDF = 1 \times 1,8822 = 1,8822$$

Perhitungan yang sama dilakukan terhadap seluruh kata pada komentar tersebut sehingga diperoleh hasil seperti pada Tabel 4.5.

Tabel 4.5 Contoh Perhitungan TF-IDF

Kata	TF	DF	IDF	TF-IDF
atasan	1	20	1,1833	1,1833
baik	1	11	1,4425	1,4425
rekan	1	7	1,6399	1,6399
kerja	1	187	0,2123	0,2123
toxic	1	64	0,6782	0,6782
bikin	1	15	1,3082	1,3082
tidak	1	125	0,3874	0,3874
betah	1	4	1,8822	1,8822

Berdasarkan Tabel 4.5, dapat dilihat bahwa setiap kata memiliki bobot TF-IDF yang berbeda. Kata *betah* memiliki bobot paling tinggi karena hanya muncul pada sedikit dokumen sehingga dianggap lebih spesifik dalam merepresentasikan

isi komentar. Sebaliknya, kata kerja memiliki bobot yang lebih rendah karena muncul pada banyak dokumen sehingga tingkat pembeda antar kategori relatif lebih kecil. Hasil pembobotan TF-IDF selanjutnya menjadi representasi numerik dari setiap komentar yang digunakan sebagai masukan dalam proses klasifikasi menggunakan *Support Vector Machine* (SVM).

#### 4.6. Hasil Pembagian Data dan Training Model SVM

Setelah proses *pre-processing* dan pembobotan TF-IDF selesai, dataset dipisahkan menjadi data *training* dan data *testing*. Tahap pembagian ini dilakukan sebelum model *Support Vector Machine* (SVM) dilatih.

Sebanyak 305 komentar digunakan dalam penelitian ini dengan proporsi pembagian 80:20. Sebanyak 80% data ditetapkan sebagai data *training* untuk mempelajari pola berdasarkan label yang tersedia, sedangkan 20% lainnya digunakan sebagai data *testing* untuk menguji kemampuan model pada data yang tidak digunakan saat pelatihan. Rincian hasil pembagian dataset ditampilkan pada Tabel 4.6.

Tabel 4.6 Pembagian Data Training dan Testing

Jenis Data	Jumlah Data	Persentase
Data Training	244	80%
Data Testing	61	20%
Total	305	100%

Berdasarkan Tabel 4.6, dari 305 komentar yang tersedia diperoleh 244 data *training* dan 61 data *testing*. Pembagian tersebut memungkinkan model memperoleh jumlah data yang memadai untuk mempelajari karakteristik dari setiap kategori, sementara data *testing* digunakan untuk melihat kemampuan model pada data yang tidak dilibatkan dalam proses pembelajaran.

Tahap berikutnya adalah melakukan *training* menggunakan 244 data *training* yang telah diproses melalui *pre-processing* dan pembobotan TF-IDF. Data tersebut telah berbentuk nilai numerik sehingga dapat digunakan sebagai masukan bagi algoritma *Support Vector Machine* (SVM). Proses pelatihan dilakukan dengan tujuan agar model dapat mengenali karakteristik komentar pada tiga kategori, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Parameter yang digunakan dalam proses *training* ditampilkan pada Tabel 4.7.

Tabel 4.7 Parameter Training Model SVM

Parameter	Nilai
Algoritma	Support Vector Machine
Kernel	Linear
Nilai C	1.0
Random State	42
Data Training	244
Data Testing	61

Berdasarkan Tabel 4.7, model klasifikasi menggunakan *Support Vector Machine* dengan *kernel* linear. Parameter  $C = 1.0$  diterapkan untuk mengatur keseimbangan antara pembentukan batas pemisah dan kesalahan klasifikasi. Sementara itu, *random state* 42 digunakan agar hasil pembagian data tetap konsisten saat proses pelatihan diulang. Penggunaan parameter tersebut menghasilkan model yang selanjutnya dapat digunakan pada tahap evaluasi dan klasifikasi komentar baru.

Proses *training* menghasilkan model klasifikasi yang selanjutnya dapat digunakan untuk menentukan kategori komentar baru. Model yang telah berhasil dilatih kemudian disimpan oleh sistem agar dapat digunakan pada proses evaluasi dan klasifikasi komentar.

#### 4.7. Hasil Evaluasi Model

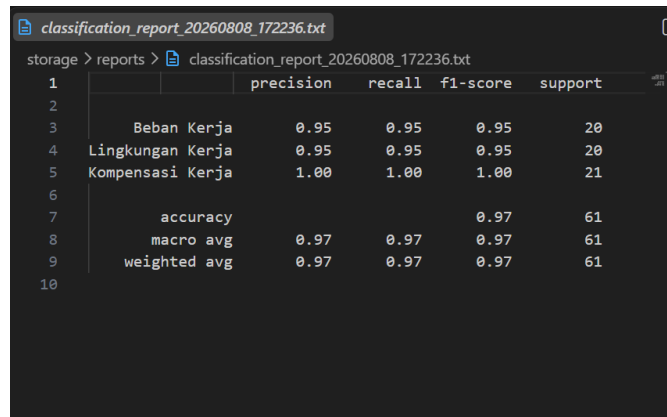
Evaluasi dilakukan setelah model selesai dilatih untuk mengukur kemampuan *Support Vector Machine* (SVM) dalam menentukan kategori komentar TikTok. Tiga kategori yang digunakan dalam pengujian adalah Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja.

Pengujian menggunakan 61 data testing dengan membandingkan hasil prediksi model terhadap label aktual. Model yang dievaluasi adalah *svm_tfidf_20260628_140438*. Ringkasan hasil evaluasi ditampilkan pada Tabel 4.8.

Tabel 4.8 Hasil Evaluasi Model SVM

<b>Metrik Evaluasi</b>	<b>Nilai</b>
<i>Accuracy</i>	96,72%
<i>Precision Macro</i>	96,67%
<i>Recall Macro</i>	96,67%
<i>F1-score Macro</i>	96,67%

Berdasarkan Tabel 4.8, model menghasilkan *accuracy* sebesar 96,72%. Hasil ini menunjukkan bahwa model mampu memberikan prediksi yang benar pada sebagian besar data testing. *Precision macro* sebesar 96,67% menunjukkan tingkat ketepatan prediksi yang baik pada setiap kategori. Nilai *recall macro* sebesar 96,67% menunjukkan bahwa model dapat mengenali sebagian besar data sesuai dengan label aktualnya. Sementara itu, *F1-score macro* yang juga mencapai 96,67% menunjukkan bahwa kemampuan model dalam menghasilkan *precision* dan *recall* berada pada tingkat yang seimbang.



```

classification_report_20260808_172236.txt
storage > reports > classification_report_20260808_172236.txt
1
2
3
4
5
6
7
8
9
10
precision    recall  f1-score   support
Beban Kerja    0.95    0.95    0.95        20
Lingkungan Kerja    0.95    0.95    0.95        20
Kompensasi Kerja    1.00    1.00    1.00        21
accuracy              0.97        61
macro avg              0.97    0.97    0.97        61
weighted avg           0.97    0.97    0.97        61

```

Gambar 4.1. Hasil Classification Report Menggunakan Python

Hasil tersebut menunjukkan bahwa model dapat mengklasifikasikan ketiga kategori komentar dengan baik dan memberikan prediksi yang konsisten pada data yang tidak digunakan saat pelatihan. Selain hasil pengukuran kinerja secara keseluruhan, evaluasi model juga disajikan dalam bentuk *confusion matrix*. *Confusion matrix* digunakan untuk membandingkan label sebenarnya dengan hasil prediksi model. Hasil perbandingan tersebut ditampilkan pada Tabel 4.9.

Tabel 4.9 Confusion Matrix Model SVM

Aktual \ Prediksi	Beban Kerja	Lingkungan Kerja	Kompensasi Kerja
Beban Kerja	19	1	0
Lingkungan Kerja	1	19	0
Kompensasi Kerja	0	0	21

Berdasarkan Tabel 4.9, jumlah prediksi benar berada pada bagian diagonal, yaitu 19 data Beban Kerja, 19 data Lingkungan Kerja, dan 21 data Kompensasi Kerja. Dengan demikian, terdapat 59 data yang berhasil diklasifikasikan dengan benar dari total 61 data testing. Sementara itu, terdapat 2 data yang mengalami kesalahan prediksi.

Pada label Beban Kerja, dari 20 data aktual terdapat 19 data yang berhasil diprediksi dengan tepat sebagai Beban Kerja, sedangkan 1 data lainnya diprediksi sebagai Lingkungan Kerja. Pada label Lingkungan Kerja, sebanyak 20 data

tercatat sebagai data aktual, dengan 19 data berhasil diprediksi secara tepat sebagai Lingkungan Kerja dan 1 data diprediksi sebagai Beban Kerja. Sementara itu, label Kompensasi Kerja memiliki 21 data aktual dan seluruh data tersebut berhasil diprediksi dengan tepat sebagai Kompensasi Kerja.

Kesalahan prediksi hanya terjadi antara kategori Beban Kerja dan Lingkungan Kerja. Kondisi tersebut dapat terjadi karena adanya kemiripan karakteristik kata pada kedua kategori, sehingga model masih mengalami kesulitan dalam menentukan kelas yang tepat untuk beberapa komentar. Sebaliknya, kategori Kompensasi Kerja dapat dikenali dengan sangat baik karena tidak terdapat kesalahan klasifikasi pada kategori tersebut.



Gambar 4.2. Hasil Confusion Matrix Menggunakan Python

Gambar tersebut menunjukkan bahwa hasil confusion matrix yang dihasilkan oleh Python sesuai dengan nilai confusion matrix yang disajikan pada Tabel 4.9. Metrik evaluasi pada masing-masing label dapat dilihat pada Tabel 4.10.

Tabel 4.10 Metrik Evaluasi per Label

Label	Precision	Recall	F1- Score	Support
Beban Kerja	95,00%		95,00%	20

		95,00%		
Lingkungan Kerja	95,00%	95,00%	95,00%	20
Kompensasi Kerja	100,00%	100,00%	100,00%	21

Berdasarkan Tabel 4.10, kategori Kompensasi Kerja memperoleh nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 100,00%. Artinya, seluruh data *testing* pada kategori tersebut berhasil diprediksi dengan benar. Sementara itu, kategori Beban Kerja dan Lingkungan Kerja mendapatkan nilai 95,00% pada ketiga metrik evaluasi. Hasil ini menunjukkan bahwa performa model pada kedua kategori tersebut sudah baik, meskipun masih terdapat beberapa prediksi yang kurang tepat.

Perbedaan nilai evaluasi antar kategori menunjukkan bahwa komentar pada kategori Kompensasi Kerja cenderung memiliki karakteristik kata yang lebih spesifik, seperti 'gaji', 'upah', atau 'bonus', sehingga lebih mudah dibedakan oleh model dibandingkan dua kategori lainnya.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *Support Vector Machine* mampu memberikan kinerja yang baik dalam mengklasifikasikan komentar TikTok berdasarkan topik keresahan karir. Nilai *accuracy* sebesar 96,72% serta *F1-score* macro sebesar 96,67% menunjukkan bahwa model mampu menghasilkan prediksi yang akurat dan seimbang pada ketiga kategori. Dengan performa tersebut, model menunjukkan kemampuan yang baik dalam mendukung proses klasifikasi komentar pada sistem yang dibangun.

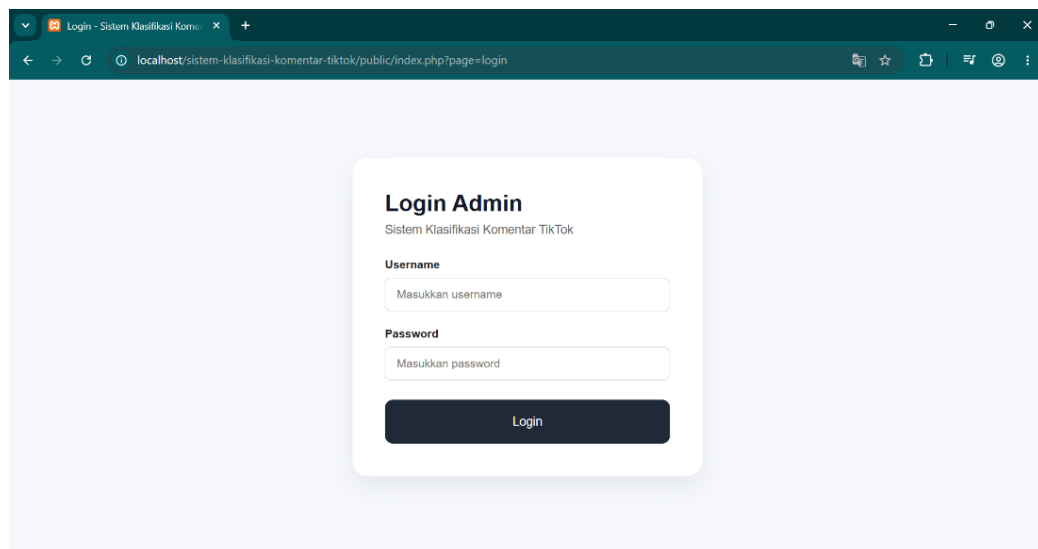
#### 4.8. Hasil Implementasi Sistem Web

Tahap implementasi sistem dilakukan untuk menerapkan proses klasifikasi komentar TikTok ke dalam bentuk sistem berbasis web. Sistem ini dibangun agar

proses pengelolaan dataset, pre-processing data, pembagian data, training model, evaluasi model, klasifikasi komentar, riwayat klasifikasi, dan manajemen model dapat dilakukan melalui satu sistem yang terstruktur.

Sistem yang dibangun digunakan oleh admin sebagai pengguna utama. Admin dapat mengelola data komentar, menjalankan proses pre-processing, membagi data training dan testing, melakukan training model Support Vector Machine, melihat hasil evaluasi model, serta melakukan klasifikasi terhadap komentar baru. Berikut merupakan hasil implementasi dari setiap halaman utama pada sistem.

#### 4.8.1. Halaman Login

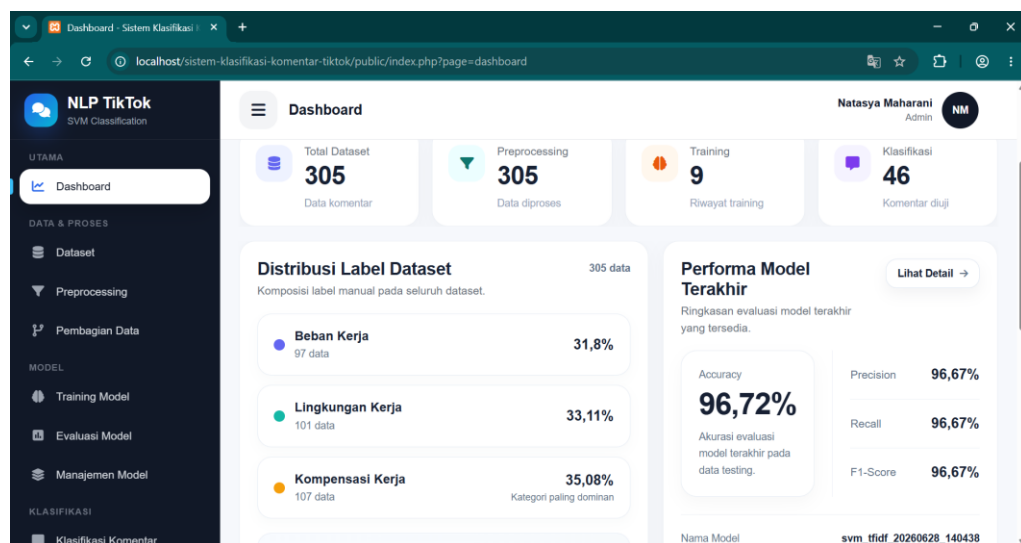


Gambar 4.3 Halaman Login

Halaman login menjadi tampilan pertama yang muncul ketika sistem diakses. Halaman ini berfungsi sebagai proses autentikasi untuk memastikan akses ke sistem klasifikasi komentar TikTok hanya dapat dilakukan oleh admin yang telah memiliki akun. Pada halaman tersebut tersedia dua kolom input, yaitu *username* dan *password*, yang digunakan untuk proses autentikasi. Setelah

username dan password diisi, admin dapat memilih tombol Login untuk masuk ke sistem. Data tersebut kemudian diperiksa oleh sistem. Jika sesuai dengan akun yang tersimpan, admin diarahkan ke halaman *dashboard*. Jika terjadi ketidaksesuaian, sistem menampilkan pesan kesalahan dan admin dapat mengulangi proses login dengan data yang benar.

#### 4.8.2. Halaman Dashboard



Gambar 4.4 Halaman Dashboard

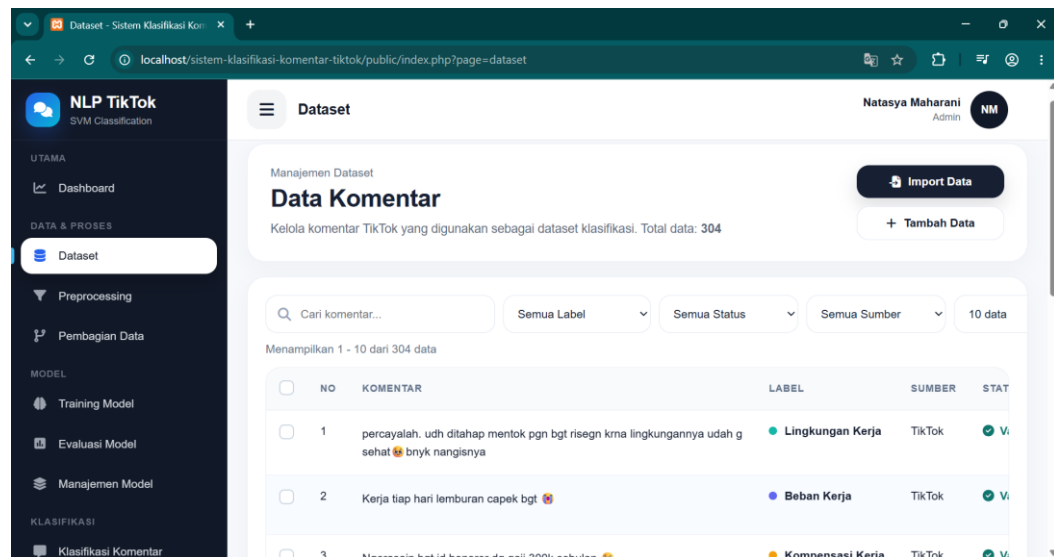
Setelah proses login berhasil, admin akan diarahkan ke halaman *dashboard* sebagai tampilan utama sistem. Pada halaman ini, admin dapat melihat ringkasan informasi terkait data dan proses klasifikasi komentar TikTok yang disajikan secara sederhana agar mudah dipahami.

Pada halaman dashboard, sistem menampilkan beberapa kartu informasi yang berisi jumlah dataset, hasil pre-processing, model yang digunakan, serta ringkasan aktivitas sistem. Selain itu, ditampilkan visualisasi distribusi label untuk memperlihatkan persebaran data pada masing-masing kategori klasifikasi.

Dashboard juga menampilkan nilai accuracy, precision, recall, dan F1-score dari model yang aktif sehingga admin dapat mengetahui performa model tanpa

harus membuka halaman evaluasi. Berdasarkan hasil implementasi, halaman dashboard mampu menyajikan informasi penting mengenai kondisi dataset dan performa model secara ringkas, sehingga memudahkan admin dalam memantau proses klasifikasi dan pengelolaan sistem secara lebih efisien.

#### 4.8.3. Halaman Dataset



Gambar 4.5. Halaman Dataset

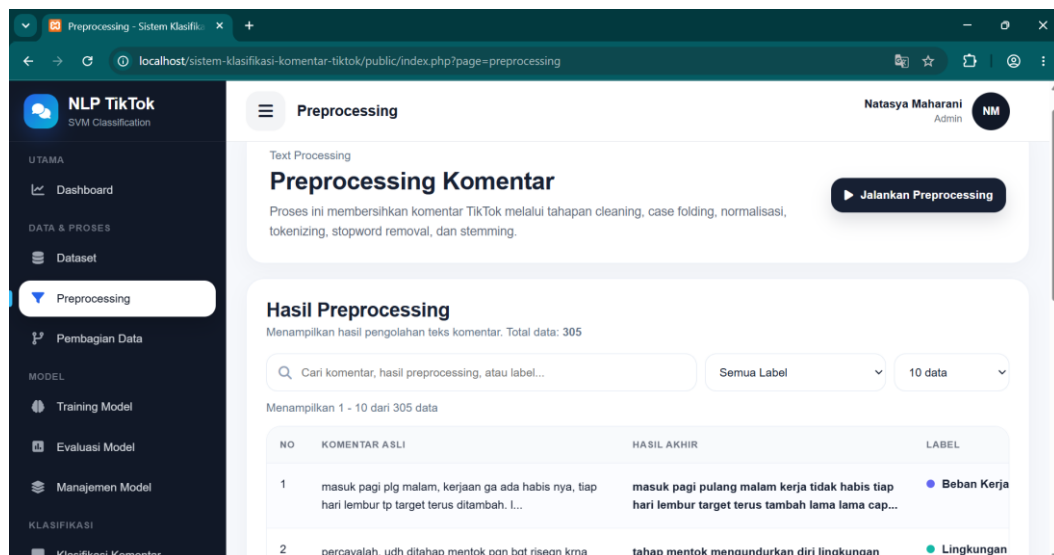
Halaman dataset digunakan untuk mengelola data komentar TikTok yang akan dijadikan sebagai dataset penelitian. Seluruh data disajikan dalam bentuk tabel sehingga admin dapat melihat dan mengelola informasi yang tersedia dengan lebih mudah.

Pada halaman ini, sistem menyediakan fitur pencarian untuk mempermudah admin dalam menemukan komentar tertentu. Selain itu, terdapat fitur filter berdasarkan label, status, dan sumber data sehingga proses pencarian menjadi lebih cepat dan terarah. Sistem juga menyediakan pengaturan jumlah data yang dapat ditampilkan pada setiap halaman. Pengaturan ini membantu menjaga tampilan tabel agar tetap teratur dan nyaman untuk dibaca.

Pada halaman dataset tersedia fitur *import* yang dapat digunakan untuk mengunggah data dalam format CSV atau XLSX. Admin juga dapat menambahkan komentar secara manual melalui fitur tambah data. Setiap data dilengkapi dengan fitur edit yang berfungsi untuk memperbarui informasi komentar serta fitur hapus untuk menghapus data yang sudah tidak diperlukan.

Melalui halaman dataset, admin dapat melakukan seluruh proses pengelolaan data komentar sebelum data diproses ke tahap berikutnya, yaitu *pre-processing*, pembagian data, pelatihan model, dan evaluasi model.

#### 4.8.4. Halaman Preprocessing



Gambar 4.6. Halaman *Pre-processing*

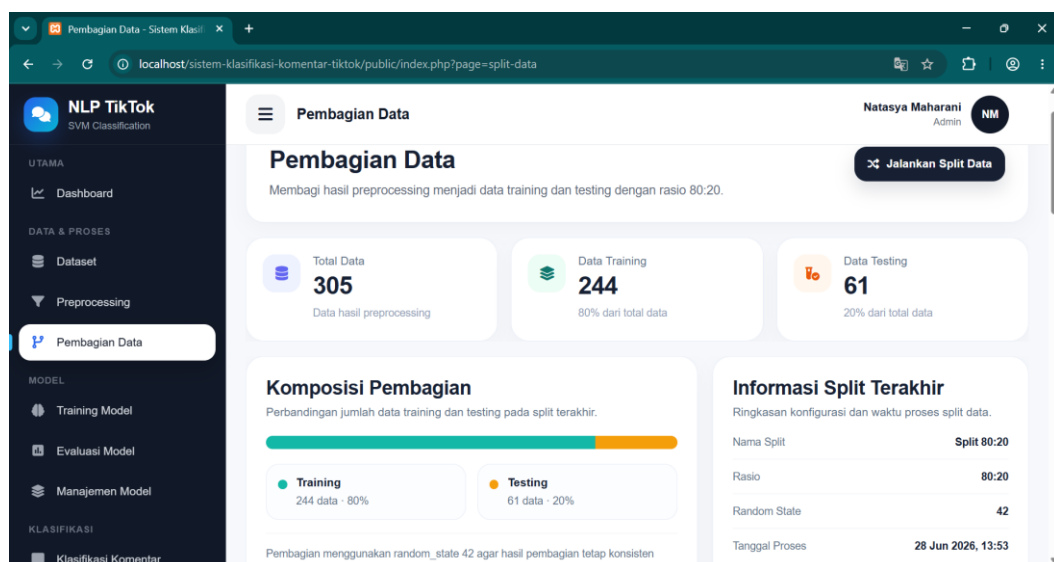
Halaman *pre-processing* digunakan untuk mengolah teks komentar sebelum masuk ke tahap pembobotan TF-IDF dan pelatihan model. Halaman ini digunakan untuk menjalankan proses *preprocessing* pada data komentar. Setiap komentar akan diolah secara bertahap melalui *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, hingga *stemming*.

Hasil *pre-processing* ditampilkan dalam bentuk tabel yang memuat komentar asli, hasil akhir pre-processing, serta label komentar. Tampilan tersebut

memudahkan admin untuk melihat perubahan komentar setelah melalui setiap tahapan pengolahan teks.

Untuk memudahkan pengelolaan data, halaman ini menyediakan fitur pencarian dan penyaringan berdasarkan label. Admin juga dapat menentukan jumlah data yang ingin ditampilkan pada setiap halaman tabel. Melalui halaman ini, admin dapat memastikan bahwa data telah diproses dengan baik sebelum digunakan pada tahap pembagian data, pelatihan model, dan evaluasi model.

#### 4.8.5. Halaman Pembagian Data



Gambar 4.7. Halaman Pembagian Data

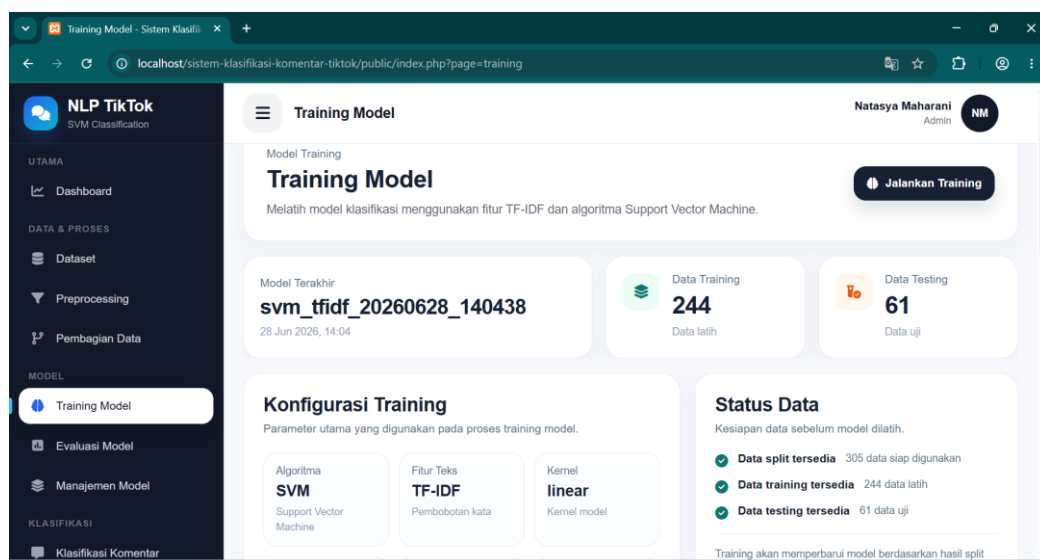
Fitur pembagian data berfungsi untuk memisahkan data yang telah melalui tahap *pre-processing* ke dalam dua kelompok, yaitu data *training* sebagai data latih dan data *testing* sebagai data pengujian. Pada halaman ini, admin dapat menjalankan proses pembagian data melalui tombol *Jalankan Split Data*.

Halaman ini menampilkan informasi pembagian data, seperti rasio pembagian, komposisi data training dan data testing, serta informasi split terakhir yang meliputi random state dan waktu proses pembagian data. Selain itu, sistem

menampilkan distribusi label pada data *training* dan data *testing* untuk memastikan bahwa setiap kategori tetap terwakili setelah proses pembagian data.

Pada bagian bawah halaman, tersedia tabel detail hasil pembagian data yang berisi tipe data, komentar, hasil pre-processing, dan label. Fitur ini memungkinkan admin memeriksa hasil pemisahan dataset ke dalam data *training* dan *testing*. Data tersebut selanjutnya digunakan sebagai dasar dalam proses pelatihan serta pengujian model.

#### 4.8.6. Halaman Training Model



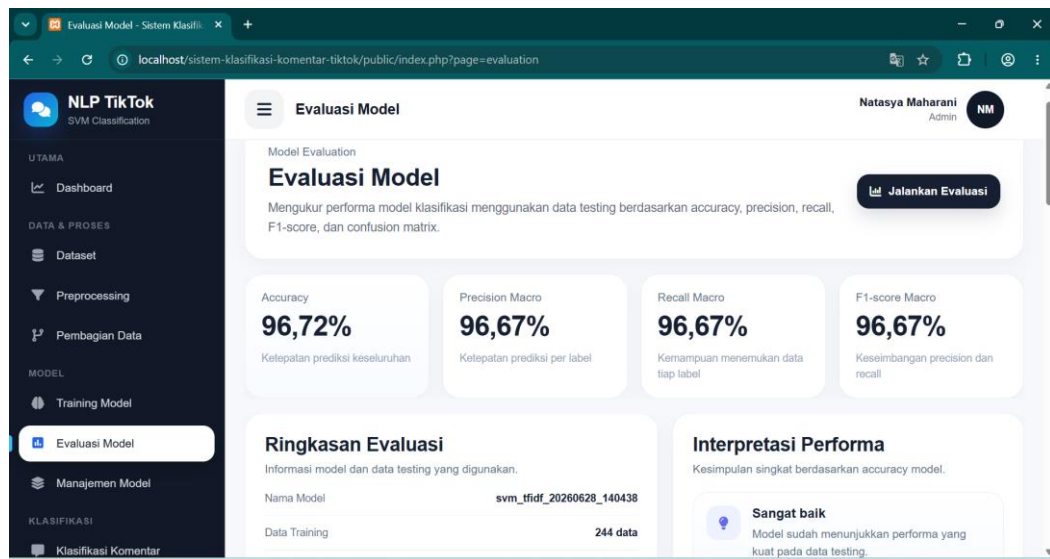
Gambar 4.8. Halaman Training Model

Halaman *training model* digunakan untuk membangun model klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Informasi yang disajikan pada halaman ini mencakup model yang digunakan, jumlah data *training* dan *testing*, serta parameter yang diterapkan selama proses pelatihan.

Selain itu, halaman ini juga menampilkan status kesiapan data sebelum proses pelatihan dilakukan, sehingga admin dapat memastikan bahwa data training dan data testing telah tersedia. Setelah proses training selesai, sistem akan

menyimpan model yang dihasilkan untuk digunakan pada tahap evaluasi maupun proses klasifikasi komentar. Proses *training* dapat dijalankan langsung oleh admin melalui halaman sistem yang tersedia. Dengan fitur ini, admin tidak perlu menjalankan kode program secara terpisah untuk melatih model.

#### 4.8.7. Halaman Evaluasi Model



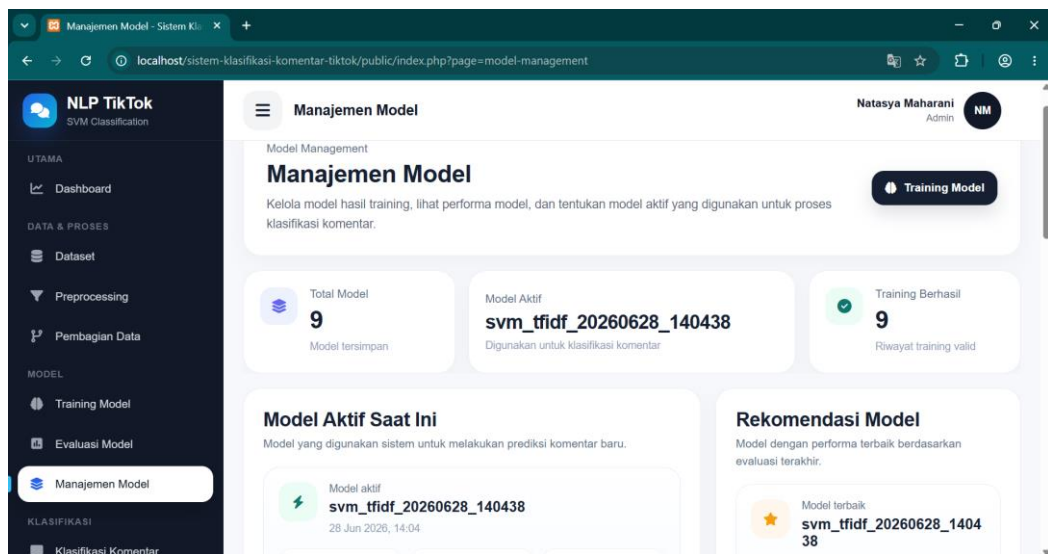
Gambar 4.9. Halaman Evaluasi Model

Halaman evaluasi model digunakan untuk menyajikan hasil evaluasi berdasarkan prediksi model terhadap data *testing*. Pada halaman ini, admin dapat menjalankan evaluasi melalui tombol Jalankan Evaluasi untuk mengetahui kinerja model yang telah dilatih.

Halaman ini menampilkan ringkasan hasil evaluasi yang diperoleh model berdasarkan hasil klasifikasi data *testing*. Penilaian kinerja model ditunjukkan melalui *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, sistem menampilkan informasi mengenai model yang dievaluasi serta interpretasi kinerjanya untuk memberikan gambaran tentang kualitas model.

Selain ringkasan evaluasi, halaman ini menampilkan confusion matrix dan metrik evaluasi untuk setiap kategori, sehingga admin dapat melihat hasil prediksi model secara lebih rinci. Sistem juga menyediakan riwayat evaluasi yang berisi daftar hasil evaluasi model yang pernah dilakukan sebagai dokumentasi proses pengujian.

#### 4.8.8. Halaman Manajemen Model



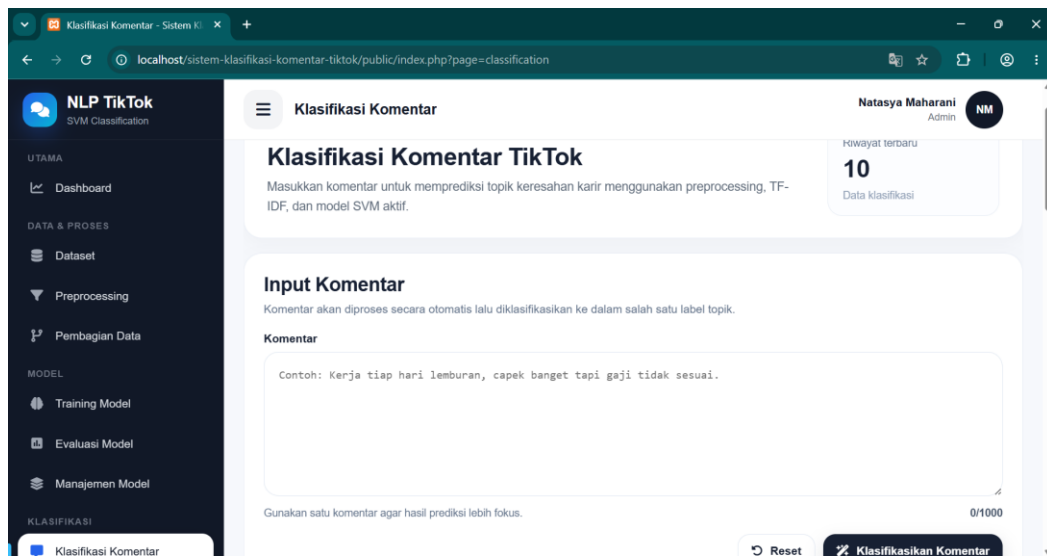
Gambar 4.10. Halaman Manajemen Model

Halaman manajemen model digunakan untuk mengatur model klasifikasi yang telah tersimpan setelah proses *training*. Melalui halaman ini, admin dapat memeriksa daftar model, mengetahui model yang sedang digunakan, serta melihat status dari masing-masing model.

Selain itu, sistem menampilkan rekomendasi model terbaik berdasarkan hasil evaluasi sehingga admin dapat menentukan model yang akan digunakan pada proses klasifikasi komentar. Halaman ini juga menyediakan daftar model yang berisi informasi konfigurasi dan performa setiap model yang telah dilatih.

Melalui halaman manajemen model, admin dapat mengaktifkan model yang diinginkan untuk digunakan pada proses klasifikasi tanpa perlu melakukan pelatihan ulang atau mengubah kode program secara manual.

#### 4.8.9. Halaman Klasifikasi Komentar



Gambar 4.11. Halaman Klasifikasi Komentar

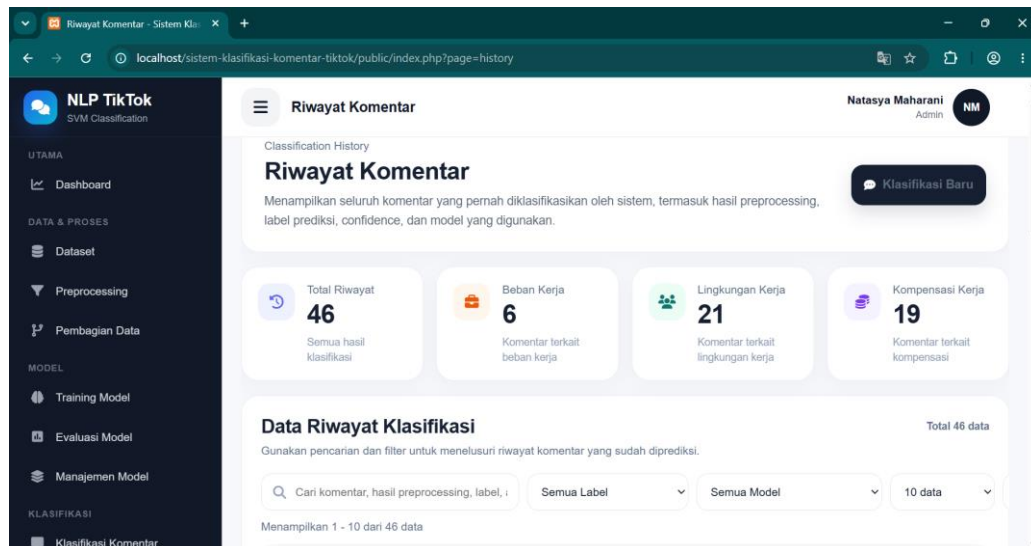
Halaman klasifikasi komentar digunakan untuk melakukan prediksi terhadap komentar baru yang dimasukkan oleh admin. Pada halaman ini, tersedia kolom input komentar dan tombol Klasifikasikan Komentar untuk menjalankan proses klasifikasi.

Setiap komentar yang diberikan oleh admin diproses secara bertahap, dimulai dari *pre-processing* dan pembobotan TF-IDF. Hasil pengolahan tersebut kemudian digunakan sebagai masukan bagi model *Support Vector Machine* yang aktif untuk menentukan kategori komentar. Hasil klasifikasi kemudian ditampilkan dalam bentuk kategori berdasarkan hasil prediksi.

Selain itu, halaman ini juga menampilkan riwayat klasifikasi terbaru yang berisi komentar yang telah diproses, hasil pre-processing, kategori prediksi, model

yang digunakan, dan waktu klasifikasi. Melalui halaman ini, admin dapat menguji komentar baru dan melihat hasil klasifikasinya secara langsung.

#### 4.8.10. Halaman Riwayat Klasifikasi



Gambar 4.12. Halaman Riwayat Klasifikasi

Halaman riwayat komentar digunakan untuk menampilkan komentar yang telah melalui proses klasifikasi oleh sistem. Di halaman ini, admin dapat mengetahui kembali hasil klasifikasi yang telah dilakukan sebelumnya sebagai dokumentasi proses klasifikasi.

Halaman ini dilengkapi fitur pencarian dan filter berdasarkan label maupun model yang digunakan. Riwayat klasifikasi ditampilkan dalam bentuk tabel yang memuat komentar, hasil pre-processing, hasil klasifikasi, model yang digunakan, tanggal klasifikasi, serta detail data.

Melalui halaman riwayat komentar, admin dapat memantau hasil klasifikasi, menelusuri kembali komentar yang telah diproses, serta melihat informasi klasifikasi secara lebih lengkap.

#### 4.9. Hasil Pengujian Sistem

Pengujian sistem bertujuan memastikan setiap fitur yang tersedia telah berjalan sesuai dengan rancangan dan kebutuhan sistem klasifikasi komentar TikTok berbasis SVM. Metode yang digunakan adalah *black-box testing*, yaitu pengujian yang menilai keluaran dan fungsi sistem berdasarkan input yang diberikan tanpa meninjau kode atau struktur internal program.

Hasil pengujian menunjukkan bahwa seluruh fitur dalam sistem dapat berjalan dengan baik sesuai dengan fungsinya, mulai dari fitur login, pengelolaan dataset komentar, *pre-processing* teks, klasifikasi komentar menggunakan SVM, hingga fitur riwayat hasil klasifikasi dan *dashboard*. Sistem juga mampu menampilkan hasil klasifikasi sesuai dengan kategori yang telah ditentukan, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja.

Tabel 4.11. Hasil Pengujian Sistem

No	Fitur yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Keterangan
1	Login Sistem	Admin memasukkan data <i>username</i> dan <i>password</i> .	Sistem berhasil melakukan proses login dan mengarahkan admin ke halaman <i>dashboard</i> .	Sesuai
2	Dashboard	Menampilkan ringkasan data sistem	Statistik dan informasi tampil	Sesuai
3	Dataset Komentar	Menambah, mengubah, serta menghapus data komentar.	Data tersimpan dan terupdate di database	Sesuai
4	Pre-processing Teks	Sistem melakukan preprocessing terhadap dataset	Data menjadi bentuk siap proses	Sesuai
5	Pembagian Data	Membagi dataset menjadi data <i>training</i> dan <i>testing</i>	Data terbagi sesuai rasio yang ditentukan	Sesuai
6	Training Model	Melatih model SVM menggunakan data	Model terbentuk dengan baik	Sesuai

		latih		
7	Evaluasi Model	Menguji performa model	Menampilkan akurasi, precision, recall, f1-score	Sesuai
8	Manajemen Model	Menyimpan, memilih, dan mengaktifkan model	Model tersimpan dan dapat digunakan	Sesuai
9	Klasifikasi Komentar	Input komentar untuk klasifikasi	Sistem menampilkan hasil kategori	Sesuai
10	Riwayat Komentar	Menampilkan hasil klasifikasi sebelumnya	Data riwayat tampil lengkap	Sesuai
11	Manajemen Pengguna	Mengelola data pengguna dengan menambah, mengubah, dan menghapus data	Data pengguna berhasil disimpan, diperbarui, dan dihapus dari sistem	Sesuai

Berdasarkan hasil *black-box testing*, seluruh fitur pada sistem telah berjalan dengan baik dan sesuai dengan kebutuhan fungsional yang ditetapkan. Seluruh proses dari pengelolaan data, pelatihan model, hingga klasifikasi komentar dapat berfungsi dengan baik.

Hasil pengujian juga menunjukkan bahwa integrasi antara aplikasi web, database, dan model SVM berjalan dengan stabil tanpa kendala yang signifikan. Dengan demikian, sistem yang dikembangkan dapat digunakan sesuai tujuan penelitian, yaitu untuk mengklasifikasikan komentar TikTok ke dalam kategori Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja.

## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan *Natural Language Processing* dalam klasifikasi topik keresahan karir berdasarkan komentar TikTok berhasil dilakukan. Data komentar yang digunakan dalam penelitian ini terdiri atas 305 data yang kemudian diklasifikasikan ke dalam tiga kategori, yaitu Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja. Data tersebut kemudian melalui proses pelabelan secara manual agar setiap komentar memiliki kategori yang sesuai dengan topik utama yang dibahas.

Proses *pre-processing* dilakukan melalui tahapan *cleaning*, *case folding*, normalisasi, *stopword removal*, dan *stemming*. Hasilnya berupa teks yang lebih bersih dan seragam, kemudian digunakan dalam proses pembobotan TF-IDF. Nilai TF-IDF tersebut selanjutnya menjadi fitur masukan untuk proses klasifikasi menggunakan algoritma *Support Vector Machine*.

Model SVM dengan kernel linear berhasil mengklasifikasikan komentar TikTok ke dalam tiga kategori yang telah ditentukan. Dari 305 data, sebanyak 244 digunakan sebagai data *training* dan 61 sebagai data *testing*. Hasil pengujian memperoleh *accuracy* sebesar 96,72%, *precision macro* 96,67%, *recall macro* 96,67%, dan *F1-score macro* 96,67%. Nilai tersebut menunjukkan bahwa model memiliki kinerja yang baik dalam mengklasifikasikan komentar berdasarkan topik keresahan karir.

Sistem berbasis web yang dibangun juga berhasil mendukung proses klasifikasi komentar TikTok secara lebih terstruktur. Sistem ini menyediakan

beberapa fitur utama, seperti pengelolaan dataset, pre-processing data, pembagian data, training model, evaluasi model, manajemen model, klasifikasi komentar, dan riwayat komentar. Dengan adanya sistem ini, proses pengolahan data dan klasifikasi komentar dapat dilakukan melalui antarmuka web sehingga lebih mudah digunakan oleh admin.

## 5.2. Saran

Berdasarkan penelitian yang telah dilakukan, pengembangan sistem selanjutnya dapat dilakukan dengan menambah jumlah dataset agar model dapat mempelajari lebih banyak variasi komentar. Penggunaan dataset yang lebih besar dan beragam diharapkan mampu meningkatkan kemampuan model dalam mengenali berbagai pola komentar yang lebih luas pada media sosial.

Penelitian selanjutnya juga dapat menambahkan kategori label lain yang masih berkaitan dengan keresahan karir. Penambahan kategori dapat membantu hasil klasifikasi menjadi lebih luas dan lebih detail, sehingga topik keresahan karir yang dianalisis tidak hanya terbatas pada Beban Kerja, Lingkungan Kerja, dan Kompensasi Kerja.

Selain itu, sistem dapat dikembangkan dengan menambahkan fitur pengambilan komentar TikTok secara otomatis. Dengan adanya fitur tersebut, proses pengumpulan data tidak perlu dilakukan secara manual sehingga dapat menghemat waktu dan mempermudah proses pembaruan dataset.

Penelitian selanjutnya dapat mencoba metode klasifikasi lain, seperti *Naive Bayes*, *Random Forest*, dan *IndoBERT*. Hasil dari metode tersebut dapat dibandingkan dengan SVM untuk mengetahui algoritma yang memberikan kinerja terbaik dalam mengklasifikasikan komentar TikTok terkait keresahan karir.

## DAFTAR PUSTAKA

- Abdurrazik, & Wirawana, I. M. W. (2025). Analisis Klasifikasi Tweet Berdasarkan Topik Sosial Menggunakan SVM. *JNATIA*, 3(4).
- Agustian, A. A., & Bisri, A. (2019). Data Mining Optimization Using Sample Bootstrapping and Particle Swarm Optimization in the Credit Approval Classification. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(1), 18–27. <https://doi.org/10.24014/ijaidm.v2i1.6299>
- Alhaqq, R. I., Putra, I. M. K., & Ruldeviyani, Y. (2022). Sentiment Analysis toward the Use of MySAPK BKN Application in Google Play Store. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 11(2), 105–113.
- Anggini, Y. C., & Syarqawi, A. (2023). Tingkat Kecemasan Menghadapi Dunia Kerja pada Mahasiswa Akhir Prodi Bimbingan Konseling Pendidikan Islam UIN Sumatera Utara Medan. *EDUKASIA: Jurnal Pendidikan Dan Pembelajaran*, 4(2), 1273–1278. <https://doi.org/10.62775/edukasia.v4i2.426>
- Astuti, K. C. (2024). Implementasi Text Mining Untuk Analisis Sentimen Masyarakat Terhadap Ulasan Aplikasi Digital Korlantas Polri pada Google Play Store. *REMIK: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 8(1), 383–394. <https://www.jurnal.polgan.ac.id/index.php/remik/article/view/13421>
- Aulia, T. M. P., Arifin, N., & Mayasari, R. (2021). Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19. *SINTECH (Science and Information Technology) Journal*, 4(2), 139–145. <https://doi.org/10.31598/sintechjournal.v4i2.762>
- Bahram, M. (2023). *TRANSFORMASI MASYARAKAT DI ERA DIGITAL:*

*MENJAGA KAIDAH HUKUM SEBAGAI LANDASAN UTAMA*. 2(5), 1733–1746.

Bakker, A. B., & Demerouti, E. (2017). Job Demands-Resources Theory : Taking Stock and Looking Forward Job Demands – Resources Theory : Taking Stock and Looking Forward. *Journal of Occupational Health Psychology*, 22, 273–285.

Budiman, B. (2021). Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru. *Nuansa Informatika*, 15(2), 37–52. <https://doi.org/10.25134/nuansa.v15i2.4162>

Bur, R., Ayuningtyas, F., & Muqsith, M. A. (2023). Jurnal Komunikasi Nusantara. *Jurnal Komunikasi Nusantara*, 5(2), 189–198. <https://doi.org/https://10.33366/jkn.v%vi%i.260>

Fide, S., Suparti, & Sudarno. (2021). *ANALISIS SENTIMEN ULASAN APLIKASI TIKTOK DI GOOGLE PLAY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN ASOSIASI*. 10(3), 346–358. <https://ejournal3.undip.ac.id/index.php/gaussian/>

HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE*, 15(5), 1–22. <https://doi.org/10.1371/journal.pone.0232525>

Hadi, A. P., & Rokhman, F. A. (2020). Implementasi Website Sebagai Media Informasi Dan Promosi Pada Pondok Pesantren Putra-Putri Addainuriyah 2 Semarang. *Pixel :Jurnal Ilmiah Komputer Grafis*, 13(1), 39–49. <https://doi.org/10.51903/pixel.v13i1.190>

Hendriyanto, M. D., & Sari, B. N. (2022). Penerapan Algoritma K-Nearest

- Neighbor Dalam Klasifikasi Judul Berita Hoax. *Jurnal Ilmiah Informatika*, 10(02), 80–84. <https://doi.org/10.33884/jif.v10i02.5477>
- Herlinda, V. A., Aditya, C. S. K., & Chandranegara, D. R. (2024). Analisis Sentimen Masyarakat Terhadap Generasi Z Dalam Dunia Kerja Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes. *REPOSITOR*, 6(4), 405–414.
- Hilmi, N. F., & Irwiensyah, F. (2024). Analisis Sentimen Terhadap Aplikasi Tiktok Dari Ulasan Pada Google Playstore Menggunakan Metode Naïve Bayes. *Smatika Jurnal*, 14(01), 146–156. <https://doi.org/10.32664/smatika.v14i01.1210>
- Ilmawan, L. B., & Mude, M. A. (2025). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Alibaba Di Google Play Store. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 3050–3057. <https://doi.org/10.36040/jati.v9i2.13269>
- Ilpa Hasanah. (2026). Pengaruh Media Sosial sebagai Tempat Curhat terhadap Kesehatan Mental Mahasiswa Stain Madina. *Jurnal Manajemen Dan Pendidikan Agama Islam*, 4(1), 65–73. <https://doi.org/10.61132/jmpai.v4i1.1727>
- Indrayuni, E., Nurhadi, A., & Kristiyanti, D. A. (2021). Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc. *Faktor Exacta*, 14(2), 64. <https://doi.org/10.30998/faktorexacta.v14i2.9697>
- Indriyani, F. A., Ahmad, F., & Sutan, F. (2023). Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine.

*TEKNOSAINS : Jurnal Sains, Teknologi Dan Informatika*, 10(2), 176–184.

<https://doi.org/10.37373/tekno.v10i2.419>

Indriyani, L. S., & Dihan, F. N. (2025). Analisis Faktor-Faktor Penyebab Stres Kerja terhadap Kepuasan Kerja Karyawan Divisi Rangka Atas UPT Balai Yasa Yogyakarta. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 6672–6684. <https://doi.org/10.31004/riggs.v4i2.1703>

Juliana, N., & Gunawan, W. (2021). Uji Validitas dan Reliabilitas South African Career Interest Inventory Short. *Jurnal Psikologi*, 17(2), 191. <https://doi.org/10.24014/jp.v17i2.12676>

Krippendorff, K. (2018). Content Analysis Chpt1and Chpt4. In *Content Analysis: An introduction to its Methodology*.

Kurniawan, E. P., S, I. B. H., & Irwansyah. (2021). Pengungkapan Diri di Situs Media Sosial Youtube. *Jurnal Lensa Mutiara Komunikasi*, 5(1), 42–51. <https://doi.org/10.51544/jlmk.v5i1.1526>

Kusumawati, I. (2017). *ANALISA MANAGEMENT DAN CAREER OPORTUNITY UNTUK MENINGKATKAN KINERJA ORGANISASI*. 1–21.

Luwei, W., & Huimin, M. (2024). From jobs to careers: drivers and barriers to career development in emerging labor markets. *Frontiers in Sociology*, 9(November). <https://doi.org/10.3389/fsoc.2024.1486871>

Mardianto, D. (2023). Komunikasi Ekspresif Penggunaan Media Sosial TikTok (Studi Kasus Generasi Z Usia 18-23 Tahun). *IKOMIK: Jurnal Ilmu Komunikasi Dan Informasi*, 3(2), 50–58. <https://doi.org/10.33830/ikomik.v3i2.6481>

- Merdiansah, R., Siska, S., & Ali Ridha, A. (2024). Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 7(1), 221–228. <https://doi.org/10.55338/jikomsi.v7i1.2895>
- Mulyono, H. S., & Saprudin, U. (2025). Efektivitas Logistic Regression dalam Analisis Sentimen Berbahasa Indonesia pada Komentar YouTube tentang Isu Ketenagakerjaan. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 6(3), 1547–1555. <https://doi.org/10.63447/jimik.v6i3.1481>
- Mutawalli, L., Zaen, M. T. A., & Bagye, W. (2019). KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto). *Jurnal Informatika Dan Rekayasa Elektronik*, 2(2), 43. <https://doi.org/10.36595/jire.v2i2.117>
- Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informatika*, 4(1), 61–72. <https://doi.org/10.54082/jiki.168>
- Pratama, E. E., & Trilaksono, B. R. (2015). Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM). *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 1(2). <https://doi.org/10.26418/jp.v1i2.11023>
- Putra, T. I. Z. M., Suprpto, & Bukhori, A. F. (2022). Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia. *Jurnal Ilmu Siber Dan Teknologi Digital*, 1(1), 1–28. <https://doi.org/10.35912/jisted.v1i1.1509>

- Rahma, I. A., & Suadaa, L. H. (2023). Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(6), 1329–1340. <https://doi.org/10.25126/jtiik.2023107325>
- Rahmi, E., Yumami, E., & Hidayasari, N. (2023). Analisis Metode Pengembangan Sistem Informasi Berbasis Website: A Literatur Review. *Riset Dan E-Jurnal Manajemen Informatika Komputer*, 7(1), 821–834. <https://doi.org/http://doi.org/10.33395/remik.v7i1.12177>
- Rifaldi, D., Fadlil, A., & Herman. (2023). DECODE: Jurnal Pendidikan Teknologi Informasi TEKNIK PREPROCESSING PADA TEXT MINING MENGGUNAKAN DATA TWEET “MENTAL HEALTH.” *DECODE: Jurnal Pendidikan Teknologi Informasi*, 3(2), 161–171.
- Rivaldi, R. C., Wismarini, T. D., Lomba, J. T., & Semarang, J. (2024). Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP) (Studi Kasus Zalika Store 88 Shopee). *Elkom*, 17(1), 120–128.
- Schmitt, A., Heihal, T. I., & Zacher, H. (2024). Financial Worries, Health Complaints, and Career Exploration: The Role of Action Crises. *Occupational Health Science*, 8(3), 613–635. <https://doi.org/10.1007/s41542-024-00182-2>
- Sintaro, S. (2022). Permodelan Sistem Informasi Pembelian dan Penjualan Berbasis Website. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 1(1), 25–32. <https://doi.org/10.58602/jima-ilkom.v1i1.5>
- Sull, D., Sull, C., & Zweig, B. (2022). Toxic culture is driving the Great Resignation. *MIT Sloan Management Review*, 63(2), 1–9.

<https://sloanreview.mit.edu/article/toxic-culture-is-driving-the-great-resignation/>

Wibowo, I. S., Witanti, A., & Susilawati, I. (2024). Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF. *Jurnal Teknik Informatika Dan Sistem Informasi*, 11(1), 99–111.

Wicaksana, D. G. A., & Rudy, D. G. (2021). Perlindungan Konsumen “Curhat Online” dalam Platform Media Sosial. *Jurnal Magister Hukum Udayana (Udayana Master Law Journal)*, 10(3), 644. <https://doi.org/10.24843/jmhu.2021.v10.i03.p15>

Yolanda, A. M., & Mulya, R. T. (2024). Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store. *Journal of Statistics and Its Application on Teaching and Research*, 6(2), 76–83. <https://doi.org/10.35580/variensiunm258>

Yudana, F. R., Suyanto, M., & Nasiri, A. (2023). Model Klasifikasi Untuk Menentukan Kesiapan Kerja Mahasiswa Dan Kelulusan Tepat Waktu Dengan Metode Machine Learning. *IJITECH: Indonesian Journal of Information Technology*, 1(1), 1–12. <https://doi.org/10.71155/vfxg8n73>

## Lampiran SK Dosen Pembimbing



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH

# UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

## FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 174/SK/BAN-PT/AR/PT/11/2014  
 Pusat Administrasi: Jalan Mukhlis Bani No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6623474 - 6631003  
 Email: [umsu@umsu.ac.id](mailto:umsu@umsu.ac.id) Website: [umsu.ac.id](http://umsu.ac.id)

---

**KEPUTUSAN DEKAN**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
 Nomor : 1100/KEP/IL.3.AU/UMSU-09/F/2025

Tentang :  
**PENGANGKATAN DOSEN PEMBIMBING SKRIPSI**  
**PRODI SISTEM INFORMASI**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, setelah:

Menimbang : 1. Bahwa sehubungan dengan pelaksanaan Tugas Akhir Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara;

2. Bahwa untuk memenuhi maksud tersebut perlu diterbitkan surat keputusan Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.

Mengingat : 1. Undang-undang Republik Indonesia No. 12 Tahun 2012 tentang Pendidikan Tinggi;

2. Peraturan Pemerintah Nomor 4 Tahun 2014 tentang Penyelenggaraan Pendidikan Tinggi dan Pengelolaan Perguruan Tinggi;

3. Pedoman Perguruan Tinggi Muhammadiyah;

4. Statuta Universitas Muhammadiyah Sumatera Utara

5. Keputusan Rektor No. 1001/KEP/IL.3-AU/UMSU/D/2025 tanggal 25 Oktober 2025 tentang Dekan Fakultas Ilmu Komputer dan Teknologi Informasi UMSU Masa Jabatan 2025-2029

6. Keputusan Rektor No. 624/KEP/IL.3.AU/UMSU/D/2025 tentang Prodi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU Masa Jabatan 2025-2029

**MEMUTUSKAN**

Menetapkan : Menetapkan nama-nama dalam lampiran surat keputusan ini sebagai Dosen Pembimbing Prodi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU.

KESATU : Menetapkan

KEDUA : Keputusan ini berlaku sejak tanggal ditetapkan, dan akan ditinjau serta diperbaiki kembali jika terdapat kekeliruan di kemudian hari.

Ditetapkan di : Medan  
 Pada tanggal : 02 Rajab 1447 H  
 22 Desember 2025 M

  
 Dekan  
 Prodi Sistem Informasi, M.Kom  
 NIDN : 0127099201



**Tembusan:**

1. Yth. Wakil Dekan I & III UMSU di Medan,-
2. Ka. Prodi SI dan Sek. Prodi SI,-
3. Peringgal.





MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No 174/SK/BAN-PT/Ak.I/II/2014

Pusat Administrasi: Jalan Mukhtar Dairi No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

 <http://www.umsu.ac.id>  [umsu.ac.id](#)  [umsu.ac.id](#)  [umsu.ac.id](#)  [umsu.ac.id](#)

Lampiran Dosen Pembimbing Prodi Sistem Informasi

Nomor : 1100/KEP/II.3.AU/UMSU-09/F/2025

Tanggal : 02 Rajab 1447 H /22 Desember 2025 M

**PENGANGKATAN DOSEN PEMBIMBING SKRIPSI  
PRODI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

NO	NAMA	NPM	JUDUL	DOSEN
1	Firza Nasifa Harahap	2209010259	"Implementasi Kriptografi AES untuk Keamanan Pengiriman Data Teks Sensitif (Chat dan Informasi Administrasi) Antar Komputer dalam Jaringan Lokal"	Mhd. Basri, S.Si. M.Kom
2	Shintya Anggie Syafitri	2209010128	Penerapan Metode Digital Forensic NIST untuk Investigasi Serangan Brute Force pada Sistem Login Website Berbasis PHP Melalui Analisis Log Server.	Andi Zulherry, S.Kom.,M.Kom.
3	Cynthia Haliza	2209010282	Aplikasi Pengamanan Ijazah Digital Berformat PDF Menggunakan Algoritma AES (Advanced Encryption Standard)	Mhd. Basri, S.Si. M.Kom
4	Ekham Difa Nasution	2209010197	Sistem Pendukung Keputusan Penentuan Prioritas Lokasi Hunian Bebas Banjir di Kawasan Tembung, Kecamatan Percut Sei Tuan Menggunakan Metode AHP Berbasis Geographic Information System (GIS).	Mhd. Basri, S.Si. M.Kom
5	Natasya Maharani	2209010079	Pemanfaatan Natural Language Processing (NLP) untuk Analisis dan Klasifikasi Topik Keresahan Karir Berdasarkan Komentar Publik di TikTok Berbasis Website Menggunakan Algoritma Support Vector Machine (SVM)	Yohanni Syahra, S.Si.,M.Kom.
6	Ummi Kalsum	2209010143	Sistem Klasifikasi Sidik Jari Berdasarkan Pola Whorl, Loop, dan Arch Menggunakan Metode K-Nearest Neighbor (KNN)	Mhd. Basri, S.Si. M.Kom
7	Vivi Damayanti	2209010116	Sistem Pendukung Keputusan Prioritas Kecamatan Penerima Program Makan Bergizi Gratis (MBG) Berdasarkan Indeks Kemiskinan Menggunakan Metode Multi-Criteria Decision Analysis (MCDA) Berbasis GIS di Kota Medan	Mahardika Prawira Tanjung, S.Kom.,M.Kcm.



## STARS

Agensi Kelayakan Malaysia  
Malaysian Qualifications Agency

## Lampiran Berita Acara Bimbingan



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**  
 UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019  
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003  
<https://fki.umsu.ac.id> [fki@umsu.ac.id](mailto:fki@umsu.ac.id) [umsu.medan](https://www.facebook.com/umsu.medan) [umsu.medan](https://www.instagram.com/umsu.medan) [umsu.medan](https://www.linkedin.com/company/umsu.medan) [umsu.medan](https://www.youtube.com/channel/UC...)

### Berita Acara Pembimbingan Proposal

Nama Mahasiswa : Natasya Maharani Program Studi : Sistem Informasi  
 NPM : 2209010079 Judul Penelitian : Pemanfaatan Natural  
 Language Processing (NLP) Untuk  
 Analisis Dan Klasifikasi Topik Keresahan  
 Karir Berdasarkan Komentar Tiktok  
 Berbasis Web Menggunakan Support  
 Vector Machine (SVM)

Nama Dosen Pembimbing : Yohanni Syahra, S.Si., M.Kom

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
07-01-2026	Bimbingan terkait judul	Yd.
02-02-2026	Penambahan bab 1, 2, dan 3	Yd.
25-02-2026	Perbaikan format penulisan bab 1, 2, dan 3	Yd.
04/03-2026	ACC Seminar Proposal	Yd.
03-06-2026	Bimbingan revisi Seminar Proposal	Yd.
11-06-2026	Penambahan bab 4 dan 5	Yd.
01-07-2026	Perbaikan bab 4 dan 5	Yd.
08-07-2026	All BAB OK. ACC sidang Skripsi	Yd.

Medan, 07 Januari 2026

Diketahui oleh :

Ketua Program Studi

  
 (Mardika Abdi Prawira Tanjung, S.Kom., M.Kom)

Disetujui oleh :

Dosen Pembimbing

  
 (Yohanni Syahra, S.Si., M.Kom)



## Lampiran Hasil Turnitin

PEMANFAATAN NATURAL LANGUAGE PROCESSING (NLP) UNTUK ANALISIS DAN KLASIFIKASI TOPIK KERESAHAN KARIR BERDASARKAN KOMENTAR TIKTOK BERBASIS WEB MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)

### ORIGINALITY REPORT

<b>28%</b> SIMILARITY INDEX	<b>25%</b> INTERNET SOURCES	<b>18%</b> PUBLICATIONS	<b>18%</b> STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	------------------------------

### PRIMARY SOURCES

<b>1</b>	<b>repository.umsu.ac.id</b> Internet Source	<b>3%</b>
<b>2</b>	<b>Submitted to UPN Veteran Yogyakarta</b> Student Paper	<b>1%</b>
<b>3</b>	<b>repository.uin-suska.ac.id</b> Internet Source	<b>1%</b>
<b>4</b>	<b>Submitted to Universitas Muhammadiyah Sumatera Utara</b> Student Paper	<b>1%</b>
<b>5</b>	<b>repository.unissula.ac.id</b> Internet Source	<b>1%</b>
<b>6</b>	<b>Submitted to Universitas Sangga Buana YPKP</b> Student Paper	<b>1%</b>
<b>7</b>	<b>etheses.uin-malang.ac.id</b> Internet Source	<b>1%</b>
<b>8</b>	<b>dspace.uui.ac.id</b> Internet Source	<b>1%</b>
<b>9</b>	<b>Akbar Supia Dirja, Lana Aulia Salsabila, M. Akhdan Putra Hermawan, Cika Jelita, Hanif Zaidan Sinaga. "Analisis Sentimen MBG Di Media Sosial X Menggunakan Support Vector Machine Dan Naive Bayes", Indonesian Journal of Multidisciplinary on Social and Technology, 2026</b> Publication	<b>&lt;1%</b>
<b>10</b>	<b>repository.unsri.ac.id</b> Internet Source	<b>&lt;1%</b>
<b>11</b>	<b>Dela Puspita Helmi, Jumadil Nangi, Asa Hari Wibowo. "Sistem Informasi Analisis Kebutuhan Fasilitas Publik Berdasarkan Laporan Masyarakat Menggunakan Metode Text Mining dan Support Vector Machine", RIGGS: Journal of Artificial Intelligence and Digital Business, 2026</b> Publication	<b>&lt;1%</b>
<b>12</b>	<b>repository.upi.edu</b> Internet Source	