Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

Analisis Sentimen Masyarakat terhadap Profesionalisme Generasi Z di Dunia Kerja Menggunakan Support Vector Machine (SVM)

Bulan Kirana Subrata[1], Yuwan Jumaryadi[2*], Febryo Ponco Sulistyo[3], Sarwati Rahayu[4], Inge Handriani[5], Ratna Mutu Manikam[6]
[1,2,3,4,5,6]Fakultas Ilmu Komputer, Universitas Mercu Buana, Indonesia
[1]41822010058@student.mercubuana.ac.id, [2*]yuwan.jumaryadi@mercubuana.ac.id,
[3]febryo.ponco@mercubuana.ac.id, [4]sarwati@mercubuana.ac.id,
[5]inge.handriani@mercubuana.ac.id, [6]ratna_mutumanikam@mercubuana.ac.id

Abstrak: Generasi Z yang lahir pada rentang tahun 1997–2012 telah menjadi bagian penting dari angkatan kerja modern. Karakteristik generasi ini yang berbeda dibandingkan generasi sebelumnya sering memunculkan berbagai persepsi dan diskusi terkait profesionalisme di lingkungan kerja, terutama melalui media sosial. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap profesionalisme Generasi Z di dunia kerja berdasarkan data yang diperoleh dari platform X. Dataset penelitian terdiri atas 2.095 tweet yang dikumpulkan melalui proses *crawling*. Pelabelan sentimen dilakukan menggunakan pendekatan berbasis leksikon yang menghasilkan 1.092 tweet (52,12%) berkategori negatif, 855 tweet (40,81%) berkategori positif, dan 145 tweet (6,92%) berkategori netral. Hasil tersebut menunjukkan bahwa persepsi masyarakat terhadap profesionalisme Generasi Z cenderung didominasi oleh sentimen negatif. Selanjutnya, data diproses melalui tahapan *text preprocessing* dan ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM) dengan tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Hasil pengujian menunjukkan bahwa model SVM memperoleh performa terbaik pada rasio pembagian data 90:10 dengan nilai akurasi sebesar 69,52%, presisi 66%, dan *recall* 70%. Temuan penelitian ini memberikan gambaran empiris mengenai persepsi publik terhadap profesionalisme Generasi Z di dunia kerja serta menunjukkan bahwa SVM mampu digunakan untuk mengklasifikasikan sentimen pada data media sosial dengan tingkat performa yang cukup baik.
Kata Kunci: Analisis Sentimen; Generasi Z; Profesionalisme Kerja; Support Vector Machine; Media Sosial X

Abstract: Generation Z, born between 1997 and 2012, has become an important part of the modern workforce. Their distinctive characteristics, which differ from those of previous generations, have sparked various public perceptions and discussions regarding their professionalism in the workplace, particularly on social media platforms. This study aims to analyze public sentiment in Indonesia toward Generation Z's professionalism in the workplace using data collected from the X platform. The research dataset consists of 2,095

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

tweets obtained through a crawling process. Sentiment labeling was conducted using a lexicon-based approach, resulting in 1,092 negative tweets (52.12%), 855 positive tweets (40.81%), and 145 neutral tweets (6.92%). These findings indicate that public perceptions of Generation Z's professionalism are predominantly negative. Subsequently, the data were processed through text preprocessing and feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) method. Sentiment classification was then performed using the Support Vector Machine (SVM) algorithm with three data-splitting scenarios: 70:30, 80:20, and 90:10. The experimental results show that the SVM model achieved its best performance with a 90:10 train-test split ratio, obtaining an accuracy of 69.52%, a precision of 66%, and a recall of 70%. The findings provide empirical insights into public perceptions of Generation Z's professionalism in the workplace and demonstrate that SVM can effectively classify sentiment in social media data with satisfactory performance.
Keywords: Sentiment Analysis; Generation Z; Professionalism; Support Vector Machine; Social Media X

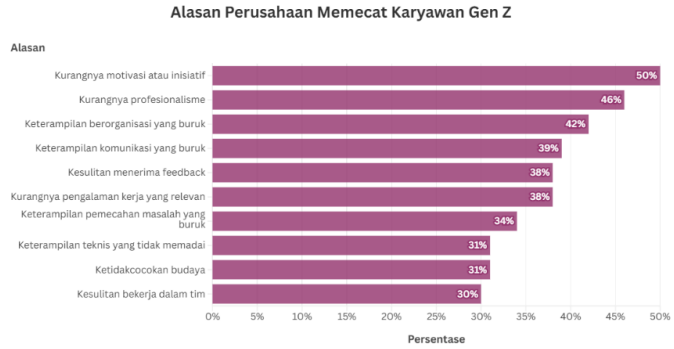
1. PENDAHULUAN

Perkembangan teknologi digital telah mengubah cara masyarakat berinteraksi, berbagi informasi, dan menyampaikan opini melalui berbagai platform media sosial [1], [2]. Media sosial tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga menjadi sumber data yang kaya untuk memahami persepsi publik terhadap berbagai isu sosial yang berkembang di masyarakat [3], [4]. Salah satu platform yang banyak digunakan untuk menyampaikan opini secara terbuka adalah X (sebelumnya Twitter). Karakteristik X yang berbasis teks memungkinkan pengguna untuk mengekspresikan pandangan, kritik, maupun pengalaman terkait berbagai fenomena sosial secara real-time. Berdasarkan laporan We Are Social tahun 2025, jumlah pengguna X di Indonesia telah melampaui 25 juta pengguna, sehingga platform ini memiliki potensi yang besar sebagai sumber data untuk menganalisis opini publik [5], [6].

Salah satu topik yang banyak menjadi perhatian dalam diskursus publik adalah profesionalisme Generasi Z di dunia kerja. Generasi Z, yang lahir pada rentang tahun 1997–2012, merupakan generasi yang tumbuh bersama perkembangan internet, media sosial, dan berbagai teknologi digital. Paparan teknologi sejak usia dini membentuk karakteristik yang berbeda dibandingkan generasi sebelumnya, baik dalam cara berkomunikasi, memperoleh informasi, maupun menjalankan aktivitas profesional. Berbagai penelitian menunjukkan bahwa Generasi Z cenderung mengutamakan fleksibilitas kerja, keseimbangan antara kehidupan pribadi dan pekerjaan (*work-life balance*), serta pekerjaan yang memiliki nilai dan makna bagi diri merek [7]. Di sisi lain, karakteristik tersebut juga memunculkan beragam persepsi dari masyarakat dan pelaku industri, mulai dari pandangan positif terkait kemampuan adaptasi teknologi hingga kritik terhadap komitmen, etika kerja, dan profesionalisme di lingkungan kerja [8].

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329



Gambar 1. Survey 2024 Alasan Perusahaan Memecat Karyawan Generasi Z [9].

Persepsi terhadap profesionalisme Generasi Z semakin mengemuka seiring meningkatnya partisipasi generasi ini dalam pasar tenaga kerja. Berdasarkan survei yang dipublikasikan oleh GoodStats pada tahun 2024, sebanyak 79% perusahaan menyatakan bahwa karyawan yang dianggap menunjukkan perilaku kurang profesional dimasukkan ke dalam program peningkatan kinerja, sedangkan 60% di antaranya berakhir pada pemutusan hubungan kerja [9]. Temuan tersebut menunjukkan bahwa isu profesionalisme generasi muda menjadi perhatian bagi berbagai organisasi dan memunculkan beragam respons di ruang publik. Oleh karena itu, diperlukan pendekatan yang mampu mengidentifikasi dan mengukur persepsi masyarakat secara sistematis berdasarkan data yang dihasilkan melalui media social

Analisis sentimen merupakan salah satu pendekatan yang banyak digunakan untuk mengidentifikasi opini, emosi, dan sikap masyarakat terhadap suatu topik berdasarkan data teks yang tersedia pada media sosial [10], [11]. Pendekatan ini memanfaatkan teknik *Natural Language Processing* (NLP) dan *machine learning* untuk mengklasifikasikan opini ke dalam kategori sentimen positif, negatif, maupun netral [12], [13]. Namun, karakteristik data media sosial yang tidak terstruktur, mengandung bahasa informal, singkatan, serta variasi kosakata yang tinggi menjadikan proses klasifikasi sentimen sebagai tantangan tersendiri. Oleh karena itu, pemilihan algoritma klasifikasi yang mampu menangani data berdimensi tinggi dan kompleks menjadi aspek penting dalam penelitian analisis sentimen.

Support Vector Machine (SVM) merupakan salah satu algoritma yang banyak digunakan dalam klasifikasi teks karena kemampuannya dalam membangun hyperplane optimal untuk memisahkan kelas data pada ruang fitur berdimensi tinggi [14], [15]. Berbagai penelitian terdahulu menunjukkan bahwa SVM mampu menghasilkan performa yang kompetitif dibandingkan metode klasifikasi lainnya, seperti Naïve Bayes dan Random Forest. Penelitian sebelumnya melaporkan bahwa SVM memperoleh akurasi sebesar 82%, lebih tinggi dibandingkan Random Forest sebesar 81% dan Naïve Bayes sebesar 71% [9]. Hasil serupa juga ditunjukkan oleh penelitian lain yang memperoleh akurasi sebesar 83% menggunakan SVM dan 79% menggunakan Naïve Bayes [10]. Meskipun demikian, sebagian besar penelitian sebelumnya berfokus pada analisis sentimen terhadap produk, layanan, atau isu sosial tertentu, sedangkan kajian yang secara khusus menganalisis persepsi masyarakat terhadap profesionalisme Generasi Z di dunia kerja masih relatif terbatas. Kondisi ini menunjukkan adanya kesenjangan penelitian (research gap) yang perlu dieksplorasi lebih lanjut.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap profesionalisme Generasi Z di dunia kerja menggunakan data yang diperoleh dari platform X selama periode Januari hingga November 2025. Penelitian menerapkan metode Support Vector Machine (SVM) untuk

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral. Selain itu, penelitian ini juga mengevaluasi pengaruh variasi rasio pembagian data terhadap performa model klasifikasi. Hasil penelitian diharapkan dapat memberikan kontribusi empiris dalam memahami persepsi publik terhadap profesionalisme Generasi Z sekaligus memperkaya kajian analisis sentimen berbasis media sosial pada konteks ketenagakerjaan dan hubungan antargenerasi di era digital.

Analisis sentimen merupakan suatu metode yang digunakan untuk mengetahui dan menilai opini atau pandangan seseorang terhadap suatu isu atau kejadian tertentu yang disampaikan dalam bentuk teks, seperti komentar maupun pendapat yang ditulis di media sosial [12]. Analisis sentimen dilakukan dengan memanfaatkan pendekatan NLP untuk memproses data berbentuk teks sebelum dilakukan proses klasifikasi [12]. Pada penelitian ini digunakan bagaimana mengetahui persepsi masyarakat terhadap bagaimana Generasi Z yang ada di dunia kerja melalui platform X, dimana data yang dianalisis umumnya bersifat tidak terstruktur, memiliki ragam kosakata yang luas, serta sering mengandung bahasa tidak baku dan singkatan. Pemilihan metode *Support Vector Machine* (SVM) dalam penelitian ini didasarkan pada karakteristik data opini masyarakat di platform X yang bersifat tidak terstruktur, berdimensi tinggi, serta memiliki pola bahasa yang kompleks. SVM memiliki keunggulan dalam menangani data teks hasil ekstraksi fitur seperti TF-IDF karena mampu membentuk hyperplane optimal dengan margin maksimum sehingga dapat memisahkan kelas sentimen secara lebih akurat. Selain itu, SVM relatif tahan terhadap *noise* dan ketidakseimbangan data yang umum ditemukan pada data media sosial, sehingga memiliki kemampuan generalisasi yang lebih baik dibandingkan metode lain seperti *Naïve Bayes*. Pemilihan SVM juga diperkuat oleh hasil penelitian terdahulu menunjukkan bahwa SVM memperoleh nilai akurasi lebih tinggi dibandingkan metode klasifikasi lain, sehingga dinilai lebih efektif dan relevan untuk digunakan dalam analisis sentimen persepsi masyarakat terhadap profesionalisme Generasi Z di dunia kerja [16].

Dalam menganalisis sentimen, model SVM telah digunakan oleh beberapa peneliti terdahulu, seperti [17] yang melakukan perbandingan terhadap tiga metode, yaitu SVM, Naïve Bayes, dan Random Forest, dengan nilai akurasi masing-masing sebesar 82% untuk SVM, 81% untuk Random Forest, dan 71% untuk Naïve Bayes. Dari hasil tersebut dapat dilihat bahwa metode SVM memperoleh nilai akurasi paling tinggi dibandingkan metode lain, sehingga dinilai lebih efektif dalam mengklasifikasikan sentimen komentar atau opini. Lalu penelitian [18] dengan penelitian mereka menunjukkan bahwa penggunaan dua metode SVM dan *Naive Bayes*, menunjukkan SVM 83% menunjukkan performa superior dalam menangani data dengan margin yang lebih tinggi dan mampu membedakan sentimen dengan lebih akurat. Sementara *Naive Bayes* 79% memberikan hasil yang cukup baik namun kurang optimal pada data dengan distribusi kompleks.

Periode penelitian pengambilan data dari 1 Januari 2025 sampai 30 bulan November 2025 karena pada rentang waktu tersebut semakin banyak Generasi Z yang mulai memasuki dunia kerja, baik melalui proses rekrutmen maupun program magang. Kondisi ini menimbulkan berbagai interaksi antara Generasi Z dengan lingkungan kerja, sehingga memunculkan opini dan pandangan masyarakat di media sosial. Rentang waktu penelitian yang cukup panjang juga memungkinkan peneliti memperoleh data untuk menggambarkan sentimen masyarakat terhadap profesionalisme Generasi Z, baik dari sudut pandang rekan kerja lintas generasi maupun dari Generasi Z sendiri yang sedang beradaptasi dengan tuntutan dan budaya kerja profesional.

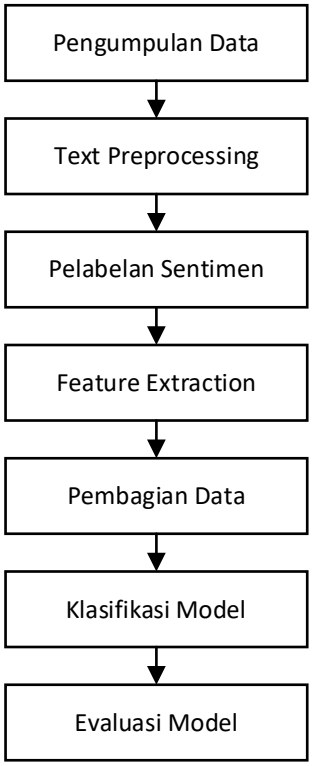
2. METODE PENELITIAN

Tahapan penelitian yang digunakan dalam penelitian ini disusun secara sistematis untuk menghasilkan model analisis sentimen yang mampu mengklasifikasikan opini masyarakat terhadap profesionalisme Generasi Z di dunia kerja. Proses penelitian meliputi

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

pengumpulan data, *text preprocessing*, pelabelan data, ekstraksi fitur, pembagian data, klasifikasi menggunakan algoritma Support Vector Machine (SVM), serta evaluasi kinerja model. Alur tahapan penelitian yang digunakan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian.

Berdasarkan Gambar 1, penelitian diawali dengan proses pengumpulan data melalui teknik *crawling* pada platform X untuk memperoleh data tweet yang relevan dengan topik penelitian. Data yang diperoleh kemudian melalui tahap *preprocessing* untuk meningkatkan kualitas data sebelum dilakukan pelabelan sentimen. Selanjutnya, data diproses pada tahap ekstraksi fitur dan dibagi menjadi data latih serta data uji untuk membangun model klasifikasi menggunakan algoritma Support Vector Machine (SVM). Tahap terakhir adalah evaluasi model untuk mengukur kinerja klasifikasi berdasarkan metrik pengujian yang digunakan. Adapun penjelasan dari masing-masing tahapan penelitian dijelaskan sebagai berikut.

1. **Pengumpulan Data**
   Data dikumpulkan melalui *crawling data* dari platform X dengan mendapatkan *tweet* publik yang mengandung opini atau pernyataan masyarakat terkait Generasi Z dalam dunia kerja. *Crawling* dilakukan menggunakan *Google Colaboration* dengan bahasa pemrograman *Python* dengan periode yang telah ditentukan, yaitu 01 Januari hingga 30 November 2025.
2. ***Text Preprocessing***
   Tweet yang diperoleh dari hasil *crawling* umumnya masih mengandung banyak unsur yang tidak relevan yang mengganggu akurasi analisis, seperti emoji, simbol, tanda baca berlebihan, URL, tagar, atau mention. Oleh karena itu, dilakukan tahap *text preprocessing* yang meliputi:
   1. *Cleansing*: Proses pembersihan data, seperti menghapus duplikasi dan menghilangkan karakter-karakter yang dianggap tidak selaras.

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

2. *Case Folding*: Proses untuk mengubah seluruh huruf dalam teks menjadi huruf kecil dengan tujuan menyamakan format penulisan
3. *Tokenization*: Proses pemecahan teks menjadi unit-unit yang lebih kecil yang disebut token, seperti kata, frasa, atau simbol tertentu, dengan tujuan mempermudah proses analisis teks.
4. *Normalization* : Proses mengganti kata-kata tidak baku menjadi bentuk baku serta mengubah akronim atau singkatan menjadi kata lengkap sesuai dengan ejaan yang benar
5. *Stopword Removal*: Proses menghilangkan kata-kata umum yang sering muncul dalam teks namun tidak memiliki makna signifikan terhadap konteks analisis.
6. *Stemming*: proses mengubah kata yang memiliki imbuhan (prefix, suffix, infix, atau konfiks) menjadi bentuk kata dasarnya menggunakan algoritma Sastrawi untuk Bahasa Indonesia.

3. Labeling Data

Data yang telah melalui tahap *preprocessing* selanjutnya diberikan label sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Proses pelabelan dilakukan menggunakan pendekatan berbasis leksikon dengan menentukan polaritas sentimen berdasarkan skor yang diperoleh dari kamus sentimen.

4. Feature Extraction

Pada tahap feature extraction, penelitian ini menggunakan metode *Term Frequency - Inverse Document Frequency* (TF-IDF) untuk mengubah data teks menjadi representasi numerik. Metode TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan tingkat kepentingannya di seluruh dokumen, sehingga kata yang sering muncul namun jarang ditemukan pada dokumen lain akan memiliki bobot yang lebih tinggi [19].

5. Pembagian Data

Setelah melalui tahapan *preprocessing*, pelabelan sentimen, dan ekstraksi fitur menggunakan TF-IDF, dataset dibagi menjadi data latih (*training data*) dan data uji (*testing data*) untuk membangun serta mengevaluasi model klasifikasi. Pada penelitian ini digunakan tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10, yang bertujuan untuk menganalisis pengaruh proporsi data latih terhadap performa algoritma Support Vector Machine (SVM). Dari total 2.095 tweet, rasio 70:30 menghasilkan 1.466 data latih dan 629 data uji, rasio 80:20 menghasilkan 1.676 data latih dan 419 data uji, sedangkan rasio 90:10 menghasilkan 1.885 data latih dan 210 data uji. Hasil dari ketiga skenario tersebut kemudian dibandingkan berdasarkan nilai akurasi, presisi, dan *recall* untuk menentukan konfigurasi pembagian data yang memberikan kinerja terbaik.

6. Klasifikasi Model

Dataset yang telah melalui tahap *preprocessing* dan pelabelan selanjutnya diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM). Representasi fitur yang dihasilkan oleh metode TF-IDF digunakan sebagai masukan (*input*) model klasifikasi. Pada penelitian ini, SVM diimplementasikan menggunakan pustaka Scikit-learn dengan kernel linear. Pemilihan kernel linear didasarkan pada karakteristik data teks yang memiliki dimensi fitur tinggi dan bersifat *sparse*, sehingga kernel linear dinilai lebih efisien dan mampu memberikan performa yang baik pada tugas klasifikasi sentimen. Parameter regularisasi (*C*) ditetapkan sebesar 1,0 menggunakan nilai bawaan (*default parameter*) dari Scikit-learn. Penelitian ini tidak menerapkan proses *hyperparameter tuning*, seperti *Grid Search* atau *Random Search*, karena penelitian berfokus pada evaluasi performa dasar algoritma SVM pada berbagai skenario pembagian data.

7. Hasil Akurasi dan Analisis

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

Setelah model selesai, tahap terakhir adalah hasil kinerja model klasifikasi sentimen yang telah dilakukan dengan tingkat akurasi serta menganalisis hasil evaluasi menggunakan confusion matrix serta classification report. Hasil evaluasi ini digunakan untuk mengetahui sejauh mana model mampu mengklasifikasikan data sentimen.

3. HASIL DAN PEMBAHASAN

Pengumpulan Data

Pada penelitian ini, data-data yang digunakan berasal dari platform X yaitu, tweets beberapa kata kunci, yaitu 'gen z kerja', 'etika gen z', 'karyawan gen z', 'profesional gen z'. Data yang terkumpul sebanyak 2154 *tweets* dengan pengambilan periode yang telah ditentukan, yaitu 01 Januari 2025 hingga 30 November 2025 dan menggunakan bahasa pemprograman *python*.

Text Preprocessing

Pada tahap ini, dataset yang telah diperoleh melalui proses *crawling* dilanjutkan untuk diolah lebih lanjut dengan tahap *pre-processing*. Proses ini dilakukan agar *machine learning* dapat mengenali pola dengan lebih mudah sehingga proses terhadap dataset menjadi lebih optimal. Adapun tahapan *pre-processing* yang dilakukan adalah sebagai berikut:

1. Cleansing

Pada tahap ini, proses yang dilakukan meliputi penghapusan data duplikat serta penghilangan karakter-karakter yang tidak selaras, seperti tanda baca (koma, titik, tanda tanya, tanda seru) maupun berbagai simbol. *Teks pada kolom full_text* dihapus karakter-karakter yang tidak memiliki makna khusus serta numerik menggunakan beberapa fungsi, diantaranya *'remove_number'*, *'remove_tweet_special'*, *'remove_punctuation'*, *'remove_singl_char'*, dan *'remove_whitespace_LT'* pada Python.

Tabel 1. *Cleansing*

No	*full_text*	*full_text_cleansing*
1	Mungkinkan dia gen z? Etika profesinya dimana?	Mungkinkan dia gen Etika profesinya dimana
2	@noodlexcheese @worksfess lah itu lu ngatain goblok, jadi apa bedanya etika lu sama gen z selain umur?	lah itu lu ngatain goblok jadi apa bedanya etika lu sama gen selain umur
3	@worksfess GUA HRD JUGA GABAKAL LADENIN MODELAN GINI, ETIKA DALAM MEMBALAS DAN IZIN MENGKONFIRMASI COBA CARI LG DI GOOGLE DAH, GUA GEN Z KOK LULUS 2024 KMRN GA GNIÂ² AMAT	GUA HRD JUGA GABAKAL LADENIN MODELAN GINI ETIKA DALAM MEMBALAS DAN IZIN MENGKONFIRMASI COBA CARI LG DI GOOGLE DAH GUA GEN KOK LULUS KMRN GA GNI AMAT
4	@sbyfess Menyebalkan punya karyawan dan atau kerja dng gen z, hen z tdk disiplin, skill minim tapi sok hebat, kerja sesukanya, tdk mau menerima masukan, adab/etika minus, paling parah komunikasi nya gak cocok didunia profesional atau kerja.	Menyebalkan punya karyawan dan atau kerja dng gen hen tdk disiplin skill minim tapi sok hebat kerja sesukanya tdk mau menerima masukan adabetika minus paling parah komunikasi nya gak cocok didunia profesional atau kerja

2. Case Folding

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

Tahapan ini merupakan proses untuk mengubah seluruh huruf dalam teks menjadi huruf kecil dengan tujuan menyamakan format penulisan agar lebih konsisten. Pada tahap case folding ini, teks pada kolom *full_text_cleansing* diubah seluruhnya menjadi huruf kecil menggunakan fungsi *'.str.lower()'* pada Python. Prosesnya dilakukan dengan menyalin dataset terlebih dahulu, lalu membuat kolom baru *'full_text_Case_Folding'* yang berisi teks yang sudah dikonversi ke huruf kecil. Tujuannya agar format penulisan menjadi lebih rapi, dan konsisten sehingga memudahkan proses pengolahan data pada tahap selanjutnya.

Tabel 2. *Case Folding*

No	*full_text_cleansing*	*full_text_Case_Folding*
1	Mungkinkan dia gen Etika profesinya dimana	mungkinkan dia gen etika profesinya dimana
2	lah itu lu ngatain goblok jadi apa bedanya etika lu sama gen selain umur	lah itu lu ngatain goblok jadi apa bedanya etika lu sama gen selain umur
3	GUA HRD JUGA GABAKAL LADENIN MODELAN GINI ETIKA DALAM MEMBALAS DAN IZIN MENGKONFIRMASI COBA CARI LG DI GOOGLE DAH GUA GEN KOK LULUS KMRN GA GNI AMAT	gua hrd juga gabakal ladenin modelan gini etika dalam membalas dan izin mengkonfirmasi coba cari lg di google dah gua gen kok lulus kmrn ga gni amat
4	Menyebalkan punya karyawan dan atau kerja dng gen hen tdk disiplin skill minim tapi sok hebat kerja sesukanya tdk mau menerima masukan adabetika minus paling parah komunikasi nya gak cocok didunia profesional atau kerja	menyebalkan punya karyawan dan atau kerja dng gen hen tdk disiplin skill minim tapi sok hebat kerja sesukanya tdk mau menerima masukan adabetika minus paling parah komunikasi nya gak cocok didunia profesional atau kerja

3. Tokenizing

Pada tahap tokenizing ini, teks dipecah menjadi bagian-bagian yang lebih kecil agar lebih mudah dianalisis. Prosesnya dilakukan dengan menyalin dataset, memastikan kolom teks sudah dalam bentuk *string*, lalu menerapkan fungsi *'word_tokenize'* pada kolom *'full_text_Case_Folding* untuk memisahkan kalimat menjadi potongan-potongan kata. Hasil pemecahan tersebut kemudian disimpan pada kolom *'full_text_tokenizing'*. Dengan demikian, setiap teks berubah menjadi deretan token (kata) sehingga lebih mudah diproses pada tahapan analisis selanjutnya.

Tabel 3. *Tokenizing*

No	*full_text_Case_Folding*	*full_text_tokenizing*
1	mungkinkan dia gen etika profesinya dimana	['mungkinkan', 'dia', 'gen', 'etika', 'profesinya', 'dimana']
2	lah itu lu ngatain goblok jadi apa bedanya etika lu sama gen selain umur	['lah', 'itu', 'lu', 'ngatain', 'goblok', 'jadi', 'apa', 'bedanya', 'etika', 'lu', 'sama', 'gen', 'selain', 'umur']
3	gua hrd juga gabakal ladenin modelan gini etika dalam membalas dan izin mengkonfirmasi coba cari lg di google dah gua gen kok lulus kmrn ga gni amat	['gua', 'hrd', 'juga', 'gabakal', 'ladenin', 'modelan', 'gini', 'etika', 'dalam', 'membalas', 'dan', 'izin', 'mengkonfirmasi', 'coba', 'cari', 'lg', 'di', 'google', 'dah', 'gua', 'gen', 'kok', 'lulus', 'kmrn', 'ga', 'gni', 'amat']

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

4	menyebalkan punya karyawan dan atau kerja dng gen hen tdk disiplin skill minim tapi sok hebat kerja sesukanya tdk mau menerima masukan adabetika minus paling parah komunikasi nya gak cocok didunia profesional atau kerja	['menyebalkan', 'punya', 'karyawan', 'dan', 'atau', 'kerja', 'dng', 'gen', 'hen', 'tdk', 'disiplin', 'skill', 'minim', 'tapi', 'sok', 'hebat', 'kerja', 'sesukanya', 'tdk', 'mau', 'menerima', 'masukan', 'adabetika', 'minus', 'paling', 'parah', 'komunikasi', 'nya', 'gak', 'cocok', 'didunia', 'profesional', 'atau', 'kerja']

4. Normalization

Pada tahap normalisasi, proses yang dilakukan adalah mengganti kata-kata tidak baku, singkatan, maupun bahasa gaul menjadi bentuk kata yang baku sesuai ejaan yang benar. Sistem terlebih dahulu memuat kamus normalisasi dari kemudian setiap kata pada teks dibandingkan dengan kamus tersebut. Jika ditemukan padanan katanya, maka kata tersebut diganti dengan bentuk bakunya. Hasil normalisasi disimpan pada kolom *'tweet_normalized'*, sehingga teks menjadi lebih standar, konsisten, dan mudah dipahami pada proses analisis selanjutnya.

Tabel 4. *Normalization*

No	*full_text_tokenizing*	*tweet_normalized*
1	['mungkinkan', 'dia', 'gen', 'etika', 'profesinya', 'dimana']	['mungkinkan', 'dia', 'gen', 'etika', 'profesinya', 'dimana']
2	['lah', 'itu', 'lu', 'ngatain', 'goblok', 'jadi', 'apa', 'bedanya', 'etika', 'lu', 'sama', 'gen', 'selain', 'umur']	['lah', 'itu', 'kamu', 'mengatai', 'goblok', 'jadi', 'apa', 'bedanya', 'etika', 'kamu', 'sama', 'gen', 'selain', 'umur']
3	['gua', 'hrd', 'juga', 'gabakal', 'ladenin', 'modelan', 'gini', 'etika', 'dalam', 'membalas', 'dan', 'izin', 'mengkonfirmasi', 'coba', 'cari', 'lg', 'di', 'google', 'dah', 'gua', 'gen', 'kok', 'lulus', 'kmrn', 'ga', 'gni', 'amat']	['gua', 'hrd', 'juga', 'tidak bakal', 'meladeni', 'modelan', 'begini', 'etika', 'dalam', 'membalas', 'dan', 'izin', 'mengkonfirmasi', 'coba', 'cari', 'lagi', 'di', 'google', 'sudah', 'gua', 'gen', 'kok', 'lulus', 'kemarin', 'tidak', 'begini', 'amat']
4	['menyebalkan', 'punya', 'karyawan', 'dan', 'atau', 'kerja', 'dng', 'gen', 'hen', 'tdk', 'disiplin', 'skill', 'minim', 'tapi', 'sok', 'hebat', 'kerja', 'sesukanya', 'tdk', 'mau', 'menerima', 'masukan', 'adabetika', 'minus', 'paling', 'parah', 'komunikasi', 'nya', 'gak', 'cocok', 'didunia', 'profesional', 'atau', 'kerja']	['menyebalkan', 'punya', 'karyawan', 'dan', 'atau', 'kerja', 'dengan', 'gen', 'hen', 'tidak', 'disiplin', 'skill', 'minim', 'tapi', 'sok', 'hebat', 'kerja', 'sesukanya', 'tidak', 'mau', 'menerima', 'masukan', 'adabetika', 'minus', 'paling', 'parah', 'komunikasi', 'ya', 'tidak', 'cocok', 'didunia', 'profesional', 'atau', 'kerja']

5. Stopword Removal

Proses menghilangkan kata-kata umum yang sering muncul dalam teks namun tidak memiliki makna terhadap analisis. Pada tahap *stopword removal* ini, dilakukan penghapusan kata-kata umum yang sering muncul namun tidak memiliki makna penting bagi analisis. Daftar *stopword* diambil dari NLTK bahasa Indonesia, kemudian ditambah dengan kata-kata slang. Setelah itu, fungsi *'stopwords_removal'* digunakan untuk menyaring token sehingga hanya menyisakan kata-kata yang relevan. Hasilnya disimpan pada kolom *'full_text_tokens_Stopword'*, sehingga teks menjadi lebih fokus dan siap untuk tahap analisis berikutnya.

Tabel 5. *Stopword Removal*

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

No	*tweet_normalized*	*full_text_tokens_Stopword*
1	['mungkinkan', 'dia', 'gen', 'etika', 'profesinya', 'dimana']	['mungkinkan', 'gen', 'etika', 'profesinya', 'dimana']
2	['lah', 'itu', 'kamu', 'mengatai', 'goblok', 'jadi', 'apa', 'bedanya', 'etika', 'kamu', 'sama', 'gen', 'selain', 'umur']	['mengatai', 'goblok', 'bedanya', 'etika', 'gen', 'umur']
3	['gua', 'hrd', 'juga', 'tidak bakal', 'meladeni', 'modelan', 'begini', 'etika', 'dalam', 'membalas', 'dan', 'izin', 'mengkonfirmasi', 'coba', 'cari', 'lagi', 'di', 'google', 'sudah', 'gua', 'gen', 'kok', 'lulus', 'kemarin', 'tidak', 'begini', 'amat']	['gua', 'hrd', 'tidak bakal', 'meladeni', 'modelan', 'etika', 'membalas', 'izin', 'mengkonfirmasi', 'coba', 'cari', 'google', 'gua', 'gen', 'lulus', 'kemarin']
4	['menyebalkan', 'punya', 'karyawan', 'dan', 'atau', 'kerja', 'dengan', 'gen', 'hen', 'tidak', 'disiplin', 'skill', 'minim', 'tapi', 'sok', 'hebat', 'kerja', 'sesukanya', 'tidak', 'mau', 'menerima', 'masukan', 'adabetika', 'minus', 'paling', 'parah', 'komunikasi', 'ya', 'tidak', 'cocok', 'didunia', 'profesional', 'atau', 'kerja']	['menyebalkan', 'karyawan', 'kerja', 'gen', 'hen', 'disiplin', 'skill', 'minim', 'sok', 'hebat', 'kerja', 'sesukanya', 'menerima', 'masukan', 'adabetika', 'minus', 'parah', 'komunikasi', 'cocok', 'didunia', 'profesional', 'kerja']

6. Stemming

Pada tahap stemming ini, setiap kata yang memiliki imbuhan diubah menjadi bentuk dasarnya menggunakan algoritma stemming dari Sastrawi. Proses ini dilakukan agar kata-kata yang sebenarnya memiliki makna sama namun berbeda bentuk penulisannya dapat diseragamkan. Dengan demikian, representasi kata menjadi lebih konsisten dan analisis teks dapat dilakukan dengan lebih efektif karena variasi bentuk kata dapat diminimalkan.

Tabel 6. *Stemming*

No	*full_text_tokens_Stopword*	*tweet_tokens_stemmed*
1	['mungkinkan', 'gen', 'etika', 'profesinya', 'dimana']	['mungkin', 'gen', 'etika', 'profesi', 'mana']
2	['mengatai', 'goblok', 'bedanya', 'etika', 'gen', 'umur']	['katai', 'goblok', 'beda', 'etika', 'gen', 'umur']
3	['gua', 'hrd', 'tidak bakal', 'meladeni', 'modelan', 'etika', 'membalas', 'izin', 'mengkonfirmasi', 'coba', 'cari', 'google', 'gua', 'gen', 'lulus', 'kemarin']	['gua', 'hrd', 'tidak bakal', 'laden', 'model', 'etika', 'balas', 'izin', 'konfirmasi', 'coba', 'cari', 'google', 'gua', 'gen', 'lulus', 'kemarin']
4	['menyebalkan', 'karyawan', 'kerja', 'gen', 'hen', 'disiplin', 'skill', 'minim', 'sok', 'hebat', 'kerja', 'sesukanya', 'menerima', 'masukan', 'adabetika', 'minus', 'parah', 'komunikasi', 'cocok', 'didunia', 'profesional', 'kerja']	['sebal', 'karyawan', 'kerja', 'gen', 'hen', 'disiplin', 'skill', 'minim', 'sok', 'hebat', 'kerja', 'suka', 'terima', 'masuk', 'adabetika', 'minus', 'parah', 'komunikasi', 'cocok', 'dunia', 'profesional', 'kerja']

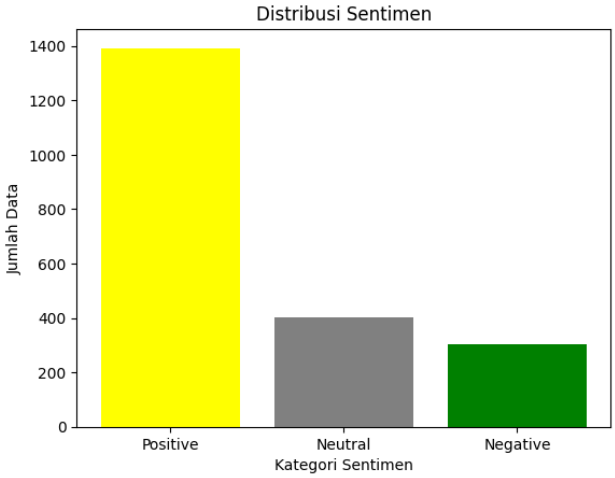
4.1 Pelabelan dataset

Pelabelan sentimen pada penelitian ini dilakukan menggunakan pendekatan berbasis leksikon (*lexicon-based labeling*). Hasil pelabelan menunjukkan bahwa dari 2.095 tweet yang dianalisis, sebanyak 1.392 tweet dikategorikan sebagai sentimen positif, 401 tweet sebagai sentimen netral, dan 302 tweet sebagai sentimen negatif, sebagaimana

Yuwan Jumaryadi: * Penulis Korespondensi

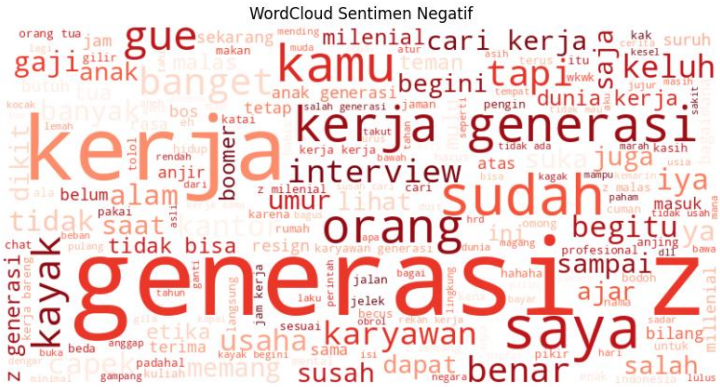
Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

ditunjukkan pada Gambar 2. Distribusi sentimen tersebut menunjukkan dominasi sentimen positif dalam dataset yang diperoleh. Hasil pelabelan menggunakan pendekatan leksikon sangat dipengaruhi oleh keberadaan kata-kata yang memiliki polaritas positif maupun negatif dalam teks. Oleh karena itu, proses penentuan sentimen dilakukan berdasarkan skor polaritas yang dihasilkan dari kamus sentimen yang digunakan tanpa mempertimbangkan konteks linguistik yang lebih kompleks, seperti ironi, sarkasme, atau makna implisit dalam suatu kalimat.



Gambar 2. Distribusi Sentimen Pelabelan Lexicon.

Wordcloud merupakan bentuk visualisasi dari hasil *pre-processing* yang telah dilakukan sebelumnya, yang digunakan untuk melihat kata-kata apa saja yang paling sering muncul dalam dataset. Pembagian wordcloud juga terbagi sesuai dengan jumlah sentimen, yaitu negatif, positif dan netral. Melalui tampilan ini, kata dengan frekuensi yang tinggi akan terlihat lebih banyak sehingga memudahkan dalam memahami gambaran umum isi data.



Gambar 2. *Wordcloud* Sentimen Negatif.

Berdasarkan visualisasi WordCloud pada sentimen negatif pada Gambar 3, terlihat bahwa kata-kata yang muncul antara lain "generasi", "gen", "z", "kerja", "orang", "susah", "karyawan", "etika", "usaha", "gaji", "keluh", "interview", "resign", dan "milenial". Dominasi kata-kata tersebut menunjukkan bahwa sentimen negatif masyarakat terhadap profesionalisme Generasi Z di dunia kerja banyak berkaitan dengan persepsi mengenai

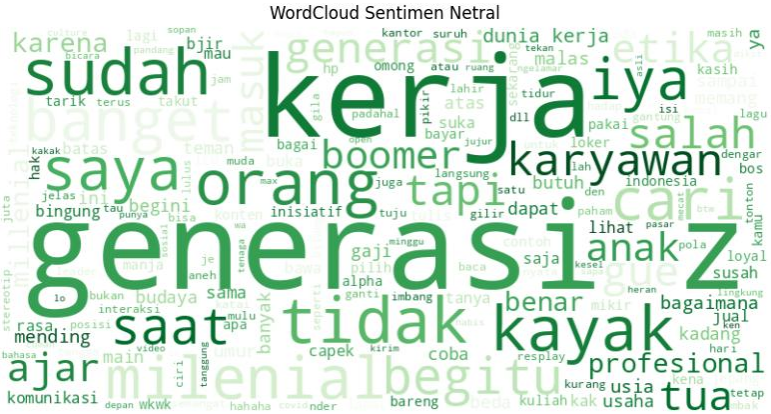
Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

sikap kerja, etika profesional, serta kemampuan beradaptasi di lingkungan kerja. Kemunculan kata seperti "susah", "tidak", "keluh", dan "resign" mengindikasikan adanya pandangan bahwa Generasi Z dianggap kurang memiliki ketahanan kerja, mudah mengeluh, serta cenderung tidak bertahan lama dalam pekerjaan. Selain itu, kata "etika" dan "karyawan" menunjukkan adanya kritik terhadap sikap profesional dan perilaku kerja Generasi Z yang dinilai belum sesuai dengan ekspektasi dunia kerja. Sementara itu, kata "gaji" dan "usaha" mencerminkan adanya perbedaan persepsi antara harapan Generasi Z terhadap kompensasi dan realitas dunia kerja, yang turut memicu sentimen negatif dari masyarakat.

Gambar 4. Wordcloud Sentimen Positif

Berdasarkan hasil visualisasi *WordCloud* pada kelas sentimen positif Gambar 4, kata-kata yang muncul antara lain "generasi", "gen", "z", "kerja", "karyawan", "profesional", "usaha", "kantor", "gaji", "bos", "orang", dan "milenial". Dominasi kata-kata tersebut menunjukkan bahwa sentimen positif masyarakat terhadap profesionalisme Generasi Z di dunia kerja berkaitan dengan persepsi mengenai kemampuan kerja, profesionalisme, serta kontribusi Generasi Z dalam lingkungan kerja.



Gambar 5. *Wordcloud* Sentimen Netral

Berdasarkan hasil visualisasi *WordCloud* pada kelas sentimen netral Gambar 5, Kata-kata yang muncul an'tara lain 'generasi', kerja', 'orang', 'saya', 'kayak', 'tidak', 'anak', 'karyawan', 'milenial', dan 'boomer'. Kemunculan kata-kata tersebut menunjukkan bahwa pembahasan netral tetap berfokus pada Generasi Z dalam konteks dunia kerja, dengan

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

banyak opini yang disampaikan berdasarkan pengalaman pribadi maupun pengamatan terhadap perilaku generasi muda di lingkungan kerja.

4.2 *Feature Extraction*

Pada tahap *feature extraction* menggunakan TF-IDF ini, setiap kata (*term*) dalam dataset diberi bobot berdasarkan seberapa penting kata tersebut dalam keseluruhan dokumen. Kata dengan skor IDF tertinggi menunjukkan bahwa kata tersebut jarang muncul di seluruh dokumen namun kuat kemunculannya pada dokumen tertentu, sehingga dianggap lebih unik dan informatif. Sebaliknya, kata dengan skor IDF rendah berarti kata tersebut sering muncul di banyak dokumen dan kurang memiliki kontribusi dalam membedakan dokumen. Dengan adanya proses ini, kita dapat mengetahui kata mana yang paling berpengaruh dalam proses analisis sentimen dan dapat membantu model dalam memahami pola penting pada teks secara lebih efektif.

Tabel 7. Pembobotan *TF-IDF.*

Kata *(term)*	*Skor TF-IDF*
generasi	0. 0.073135
kerja	0.072347
saya	0.028417
sudah	0.024537
karyawan	0.018819
….	…..
batal	0.0
telaah	0.0
pkl	0.0
remuk	0.0
ninja	0.0

4.3 *Pemodelan*

Pada tahap pemodelan, dataset yang telah melalui proses *preprocessing* selanjutnya digunakan untuk membangun model klasifikasi sentimen. Proses diawali dengan pembagian dataset menjadi data latih (*training data*) dan data uji (*testing data*) menggunakan tiga skenario rasio pembagian, yaitu 70:30, 80:20, dan 90:10 dari total 2.095 data tweet. Penggunaan beberapa rasio pembagian data bertujuan untuk menganalisis pengaruh proporsi data latih terhadap kinerja model klasifikasi yang dihasilkan.

Sebelum proses klasifikasi dilakukan, data teks dipersiapkan menggunakan pustaka Python, seperti Pandas dan NumPy. Data yang telah melalui tahapan *text preprocessing* dan pelabelan sentimen kemudian direpresentasikan ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Representasi TF-IDF digunakan karena mampu mengubah data teks menjadi vektor numerik yang dapat diproses oleh algoritma pembelajaran mesin serta memberikan bobot yang lebih besar pada kata-kata yang dianggap penting dalam dokumen.

Proses klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) yang diimplementasikan melalui pustaka Scikit-learn. Model SVM dibangun menggunakan kernel linear dengan parameter regularisasi (*C*) sebesar 1,0. Pemilihan kernel linear didasarkan pada karakteristik data teks hasil ekstraksi TF-IDF yang umumnya memiliki dimensi fitur tinggi dan bersifat *sparse*, sehingga kernel linear dinilai lebih efektif dan efisien dibandingkan kernel non-linear dalam tugas klasifikasi teks. Selain itu, nilai parameter *C* sebesar 1,0 digunakan sebagai konfigurasi standar (*default*) untuk menjaga keseimbangan antara kemampuan model dalam meminimalkan kesalahan klasifikasi pada

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

data latih dan kemampuan generalisasi terhadap data baru. Penelitian ini tidak menerapkan proses optimasi parameter (*hyperparameter tuning*), seperti *Grid Search* atau *Random Search*, sehingga hasil yang diperoleh mencerminkan kinerja dasar algoritma SVM dengan konfigurasi parameter standar pada dataset yang digunakan.

Tahap terakhir adalah evaluasi dan analisis kinerja model. Evaluasi dilakukan menggunakan *confusion matrix* dan *classification report* untuk memperoleh nilai akurasi, presisi (*precision*), *recall*, dan *F1-score*. Hasil evaluasi tersebut digunakan untuk mengukur kemampuan model dalam mengklasifikasikan sentimen positif, negatif, dan netral, serta untuk menganalisis pengaruh rasio pembagian data terhadap performa algoritma SVM dalam melakukan klasifikasi sentimen.

4.4 Evaluasi

Evaluasi kinerja model dilakukan untuk mengukur kemampuan algoritma dalam mengklasifikasikan sentimen masyarakat terhadap profesionalisme Generasi Z di dunia kerja. Pengujian dilakukan menggunakan tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Selain menggunakan algoritma Support Vector Machine (SVM) sebagai metode utama, penelitian ini juga membandingkan performa SVM dengan algoritma Naive Bayes dan Random Forest. Evaluasi dilakukan menggunakan *confusion matrix* dan *classification report* yang menghasilkan metrik akurasi, presisi (*precision*), *recall*, dan *F1-score*.

Hasil pengujian menunjukkan bahwa performa model cenderung meningkat seiring bertambahnya proporsi data latih. Pada rasio pembagian data 70:30, SVM menghasilkan akurasi sebesar 48,81%, sedangkan Naive Bayes dan Random Forest masing-masing memperoleh akurasi sebesar 58,18% dan 59,62%. Pada skenario ini, Random Forest menunjukkan performa terbaik dibandingkan dua algoritma lainnya. Namun, seluruh model masih mengalami kesulitan dalam mengklasifikasikan kelas sentimen netral yang memiliki jumlah data relatif lebih sedikit dibandingkan kelas positif dan negatif. Kondisi tersebut terlihat dari rendahnya nilai *recall* dan *F1-score* pada kelas netral.

Ketika proporsi data latih ditingkatkan menjadi 80%, performa SVM mengalami peningkatan yang signifikan dengan akurasi mencapai 64,92%. Nilai tersebut lebih tinggi dibandingkan Naive Bayes dan Random Forest yang masing-masing memperoleh akurasi sebesar 63,35% dan 59,19%. Peningkatan ini menunjukkan bahwa SVM mampu memanfaatkan tambahan data latih untuk membentuk batas keputusan (*decision boundary*) yang lebih optimal dalam ruang fitur TF-IDF. Meskipun demikian, ketidakseimbangan distribusi data masih memengaruhi kemampuan model dalam mengenali sentimen netral secara konsisten.

Performa terbaik diperoleh pada rasio pembagian data 90:10. Pada skenario ini, SVM menghasilkan akurasi tertinggi sebesar 69,52%, diikuti oleh Naive Bayes sebesar 66,67% dan Random Forest sebesar 64,29%. Selain menghasilkan akurasi yang lebih tinggi, SVM juga menunjukkan keseimbangan yang lebih baik antara nilai *precision*, *recall*, dan *F1-score* dibandingkan model pembanding. Temuan ini mengindikasikan bahwa SVM memiliki kemampuan yang lebih baik dalam mengidentifikasi pola sentimen pada data teks media sosial, khususnya ketika jumlah data latih yang tersedia lebih besar.

Secara keseluruhan, hasil penelitian menunjukkan bahwa SVM merupakan algoritma yang paling efektif untuk klasifikasi sentimen pada dataset yang digunakan. Peningkatan performa yang konsisten pada setiap skenario pembagian data mengindikasikan bahwa jumlah data latih berpengaruh terhadap kemampuan model dalam mempelajari karakteristik sentimen. Namun demikian, rendahnya performa pada kelas netral menunjukkan bahwa ketidakseimbangan distribusi data masih menjadi tantangan utama dalam proses klasifikasi. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penerapan teknik penyeimbangan data, seperti *Synthetic Minority Over-sampling*

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

Technique (SMOTE), atau penggunaan model berbasis *deep learning* dan *transformer* untuk meningkatkan kemampuan klasifikasi pada kelas minoritas.

Tabel 8. Hasil Akurasi Model Klasifikasi

Model	70:30	80:20	90:10
SVM	48,81%	64,92%	69,52%
Naïve Bayes	58,81%	63,35%	66,67%
Random Forest	59.62%	59,19%	64,29%

Berdasarkan Tabel 8, performa model klasifikasi cenderung meningkat seiring bertambahnya proporsi data latih. Algoritma SVM menunjukkan peningkatan akurasi yang paling signifikan, yaitu dari 48,81% pada rasio 70:30 menjadi 69,52% pada rasio 90:10. Sementara itu, Naive Bayes dan Random Forest juga mengalami peningkatan performa, meskipun tidak sebesar SVM. Hasil ini menunjukkan bahwa penambahan jumlah data latih berkontribusi terhadap kemampuan model dalam mempelajari pola sentimen dan meningkatkan akurasi klasifikasi.

4. KESIMPULAN

Penelitian ini berhasil menganalisis sentimen masyarakat Indonesia terhadap profesionalisme Generasi Z di dunia kerja menggunakan data yang diperoleh dari platform X selama periode Januari–November 2025. Hasil analisis menunjukkan bahwa persepsi masyarakat cenderung didominasi oleh sentimen negatif dibandingkan sentimen positif dan netral. Temuan ini mengindikasikan bahwa profesionalisme Generasi Z masih menjadi isu yang memunculkan berbagai tanggapan di ruang publik digital, khususnya terkait etos kerja, komitmen profesional, dan kemampuan beradaptasi di lingkungan kerja.

Penerapan algoritma Support Vector Machine (SVM) menunjukkan bahwa model mampu mengklasifikasikan sentimen masyarakat dengan performa yang cukup baik, dengan hasil terbaik diperoleh pada rasio pembagian data 90:10. Temuan ini menunjukkan bahwa peningkatan proporsi data latih berkontribusi terhadap kemampuan model dalam mengenali pola sentimen pada data media sosial. Namun demikian, hasil evaluasi juga menunjukkan bahwa model masih mengalami keterbatasan dalam mengklasifikasikan kelas sentimen yang memiliki jumlah data relatif sedikit, khususnya kelas netral. Kondisi ini mengindikasikan bahwa ketidakseimbangan distribusi data dan kompleksitas bahasa pada media sosial masih menjadi tantangan dalam proses klasifikasi sentimen.

Penelitian ini memberikan kontribusi empiris dalam memahami persepsi publik terhadap profesionalisme Generasi Z melalui pendekatan analisis sentimen berbasis pembelajaran mesin. Selain itu, hasil penelitian menunjukkan bahwa media sosial dapat dimanfaatkan sebagai sumber data yang relevan untuk mengidentifikasi kecenderungan opini masyarakat terhadap isu-isu ketenagakerjaan dan hubungan antargenerasi. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada jumlah dan distribusi dataset yang belum seimbang serta belum diterapkannya proses optimasi parameter (hyperparameter tuning) pada model SVM. Oleh karena itu, penelitian selanjutnya dapat memanfaatkan dataset yang lebih besar dan representatif, menerapkan teknik penyeimbangan data, melakukan optimasi parameter model, serta mengeksplorasi metode berbasis deep learning dan transformer, seperti Long Short-Term Memory (LSTM) dan Bidirectional Encoder Representations from Transformers (BERT), untuk meningkatkan kinerja klasifikasi sentimen.

5. REFERENCES

[1] U. Salamah, Y. Jumaryadi, and B. Priambodo, "EDUKASI PENGOLAHAN DATA

Yuwan Jumaryadi: * Penulis Korespondensi

STATISTIK MENGGUNAKAN EXCEL UNTUK STAFF DAN GURU SD," *J. Pasopati*, vol. 5, no. 1, pp. 44–50, 2023.

[2] R. I. Kesuma and A. Iqbal, "Penerapan Content-Boosted Collaborative Filtering untuk Meningkatkan Kemampuan Sistem Rekomendasi Penyedia Jasa Acara Pernikahan," *J. Ilm. FIFO*, vol. 12, no. 1, p. 112, 2020, doi: 10.22441/fifo.2020.v12i1.009.

[3] Y. Jumaryadi, R. Fajriah, U. Salamah, B. Priambodo, and A. Lystha, "Machine Learning Approaches to Sentiment Analysis of Mental Health Discussions on Platform X," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 13, no. 2, pp. 235–246, 2025, doi: 10.33558/piksel.v13i2.11350.

[4] Y. Jumaryadi, R. Meiyanti, R. Fajriah, A. N. Mahsyar, and P. S. Anggraeni, "Implementasi Algoritma Random Forest untuk Analisis Sentimen Ulasan Pengguna Aplikasi Merdeka Mengajar," *Bull. Comput. Sci. Res.*, vol. 5, no. 4, pp. 813–820, 2025, doi: 10.47065/bulletincsr.v5i4.530.

[5] L. Afuan, N. Hidayat, and S. Nurhayati, "Aplikasi untuk Mengenerate dan Pengiriman Sertifikat Webinar di Masa Pandemi Corona Virus Disease 19," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 4, p. 735, 2021, doi: 10.25126/jtiik.2021844984.

[6] S. Kemp, "Digital 2025: Indonesia," DataReportal – Global Digital Insights. Accessed: Jan. 01, 2026. [Online]. Available: https://datareportal.com/reports/digital-2025-indonesia

[7] M. D. Al Fahreza, A. Luthfiarta, M. Rafid, and M. Indrawan, "Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 16–25, 2024.

[8] P. K. Putri, "GEN Z DI DUNIA KERJA: Kepribadian dan Motivasi Jadi Penentu Produktivitas Kerja," *Akad. J. Mhs. Ekon. Bisnis*, vol. 4, no. 1, pp. 30–38, 2024, doi: 10.37481/jmeb.v4i1.650.

[9] E. C. D. A. Nanda, "Benarkah Gen Z Problematik di Dunia Kerja?," GoodStats. Accessed: Feb. 02, 2026. [Online]. Available: https://goodstats.id/article/gen-z-problematik-di-dunia-kerja-benarkah-iQzMV

[10] L. Hakim, M. V. Dalimunthe, C. Danuputri, and D. Widyaningrum, "Sentimen Analisis Mengenai Polusi Udara Menggunakan Algoritma Support Vector Machine dan Random Forest," *J. Ilm. FIFO*, vol. 15, no. 2, pp. 91–101, 2023, doi: 10.22441/fifo.2023.v15i2.001.

[11] S. Sandiwarno, D. I. Sensuse, H. Budi Santoso, D. Sumirat Hidayat, and A. S. Nyamawe, "SEMAR: A Multi-Task Term Weighting Approach for Sentiment and Emotion-Based E-Learning Users' Satisfaction Analysis," *IEEE Access*, vol. 13, no. September, pp. 187584–187601, 2025, doi: 10.1109/ACCESS.2025.3627657.

[12] S. N. Awwal and D. Gunawan, "Analisis sentimen terhadap kompetensi kerja generasi z menggunakan model naive bayes," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 4, pp. 4123–4134, 2025.

[13] N. Zelina and A. Afiyati, "Analisis Sentimen Ulasan Pengguna Aplikasi M-Banking Menggunakan Algoritma Support Vector Machine dan Decision Tree," *J. Linguist. Komputasional*, vol. 7, no. 1, pp. 31–37, 2024, doi: 10.26418/jlk.v7i1.169.

[14] K. A. Qureshi, R. A. S. Malick, M. Sabih, and H. Cherifi, "Deception detection on social media: A source-based perspective," *Knowledge-Based Syst.*, vol. 256, p. 109649, 2022, doi: 10.1016/j.knosys.2022.109649.

[15] T. H. Pinem and Z. P. Putra, "Evaluasi Kinerja Algoritma Klasifikasi Deep Learning dalam Prediksi Diabetes," *J. Ilm. FIFO*, vol. 17, no. 1, p. 17, 2025, doi: 10.22441/fifo.2025.v17i1.003.

[16] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura J. Math.*, vol. 5, no. 2, pp. 431–445, 2023, doi: 10.34312/jjom.v5i2.20860.

Yuwan Jumaryadi: * Penulis Korespondensi

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.279
Page 314-329

[17] D. A. TARIGAN, Z. Situmorang, and R. Rosnelly, "Analisis Sentimen Aplikasi Playstore Sirekap 2024 Pasca Pilpres Dengan Perbandingan Metode Support Vector Machine (SVM), Naïve Bayes Classifier Dan Random Forest.," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 3, pp. 661–670, 2025, doi: 10.25126/jtiik.22025129608.

[18] M. F. A. Shidiq and D. Alita, "Analisis Sentimen Masyarakat Terhadap Kasus Judi Online Menggunakan Data dari Media Sosial X Pendekatan Naive Bayes dan SVM," *J. Sist. Inf. dan Inform.*, vol. 8, no. 1, pp. 24–35, 2025.

[19] E. S. Aritonang and Y. Jumaryadi, "Analisis Sentimen Ulasan Pengguna QRIS pada Aplikasi GoPay : Studi Komparatif Algoritma Support Vector Machine dan Decision Tree Berbasis TF – IDF," *TIN Terap. Inform. Nusant.*, vol. 6, no. 12, pp. 2323–2330, 2026, doi: 10.47065/tin.v6i12.9582.

Yuwan Jumaryadi: * Penulis Korespondensi