TRANSMISI: JURNAL ILMIAH TEKNIK ELEKTRO, 28, (4), OKTOBER 2026
p-ISSN 1411-0814 e-ISSN 2407-6422

SENTIMEN ANALISIS MASYARAKAT PLATFORM X TERHADAP PILIHAN JALUR KARIR MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*

Nazwa Rafilla dan Esi Putri Silmina

Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas 'Aisyiyah Yogyakarta
Jl. Siliwangi (Ring Road Barat) No.63 Nogotirto, Gamping, Sleman, Yogyakarta

*)*Penulis korespondensi: nazwararafilla@gmail.com, esiputrisilmina@unisayogya.ac.id*

Abstrak

Pilihan jalur karir merupakan keputusan strategis yang dipengaruhi oleh faktor internal dan eksternal, seperti nilai pribadi, kondisi sosial, serta persepsi terhadap stabilitas dan peluang kerja. Perkembangan media sosial memungkinkan analisis opini publik dilakukan melalui pendekatan analisis sentimen. Penelitian ini bertujuan menganalisis sentimen masyarakat di Platform X terhadap tiga pilihan jalur karir, yaitu Pegawai Negeri, Pegawai Swasta, dan Wirausaha menggunakan metode SVM. Data dikumpulkan melalui proses *scraping* pada periode 1 Januari 2017 hingga 30 November 2025 dengan total 14.212 data. Tahapan penelitian meliputi *preprocessing*, *labeling*, ekstraksi fitur, serta pemisahan data menggunakan rasio 70:30, 80:20, dan 90:10. Evaluasi model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa skema 90:10 memberikan kinerja tertinggi dengan *accuracy* 75,95% pada Pegawai Negeri, 74,18% pada Pegawai Swasta, dan 73,65% pada Wirausaha. Distribusi sentimen menunjukkan dominasi kategori positif pada ketiga jalur karir. Temuan ini menunjukkan kecenderungan persepsi masyarakat yang positif terhadap pilihan karir dan memberikan gambaran berbasis data untuk mendukung perumusan kebijakan ketenagakerjaan.

Kata kunci: pilihan karir, platform x, sentimen analisis, SVM

Abstract

The choice of career path is a strategic decision that is influenced by internal and external factors, such as personal values, social conditions, and perceptions of stability and job opportunities. The development of social media allows public opinion analysis to be carried out through a sentiment analysis approach. This study aims to analyze public sentiment on Platform X regarding three career paths, namely Civil Servants, Private Employees, and Entrepreneurs, using the Support Vector Machine (SVM) method. Data was collected through a scraping process in the period 1 January 2017 to 30 November 2025 with a total of 14,212 data. The research stages include preprocessing, labeling, feature extraction, and data splitting using ratios of 70:30, 80:20, and 90:10. Model evaluation was carried out using a confusion matrix with accuracy, precision, recall and F1-score metrics. The test results show that the 90:10 scheme provides the highest performance with an accuracy of 75.95% for Civil Servants, 74.18% for Private Employees, and 73.65% for Entrepreneurs. The sentiment distribution was dominated by the positive category across all three career paths. These findings show a trend in positive public perceptions of career choices and provide a data-based picture to support employment policies.

Keywords: career choices, platform x, sentiment analysis, SVM

1. Pendahuluan

Keberhasilan atau kegagalan dalam karir memiliki kaitan erat dengan pembentukan konsep diri, identitas, serta tingkat kepuasan individu. Individu yang merasa memiliki jaminan terhadap karirnya cenderung lebih puas dan menunjukkan komitmen tinggi dalam bekerja [1]. Perencanaan karir merupakan proses yang dilakukan seseorang untuk mengenali keterampilan, minat, nilai, serta peluang dan tantangan yang ada, guna menetapkan sasaran karir dan menyusun strategi untuk mencapainya [2]. Keputusan seseorang dalam memilih karir tidak semata-mata ditentukan oleh bakat atau kemampuan, melainkan juga dipengaruhi oleh berbagai faktor emosional dan sosial yang turut membentuk arah pilihan karirnya [3]. Faktor internal seperti nilai pribadi, harapan keluarga, dan pandangan terhadap kesuksesan turut memengaruhi keputusan karir [4]. Penelitian menemukan bahwa sebagian besar masyarakat lebih memilih jalur karir sebagai pegawai negeri atau pegawai swasta karena dianggap memberikan kepastian masa depan melalui penghasilan yang stabil [5]. Berdasarkan data yang dirilis

oleh Kementerian Koperasi dan Usaha Kecil dan Menengah pada tahun 2023, tingkat kewirausahaan di Indonesia tercatat masih berada pada angka 3,47% [6], yang mencerminkan masih rendahnya partisipasi masyarakat dalam memilih jalur karir sebagai wirausaha, sehingga penting untuk memahami bagaimana persepsi publik terhadap berbagai pilihan karir tersebut.

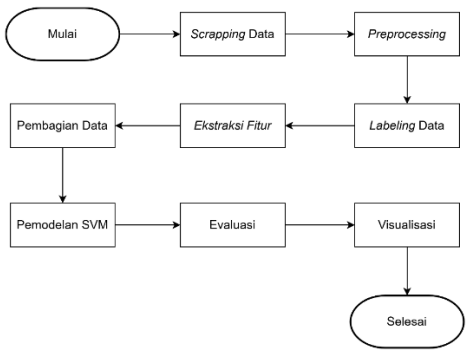
Platform X memungkinkan pengguna berbagi pengalaman singkat dalam bentuk teks, foto, maupun video [7]. Pengguna menyampaikan pandangan, ide, serta respons mereka secara terbuka [8]. Di antara berbagai platform media sosial, X termasuk yang paling sering menjadi objek penelitian karena kemampuannya menyediakan data secara cepat, dalam jumlah besar, serta dilengkapi dengan metadata yang lengkap. Data yang diperoleh melalui API X tersedia dalam format yang mudah diproses untuk analisis [9]. Unggahan yang dipublikasikan dapat dimanfaatkan sebagai sumber data untuk mengidentifikasi opini maupun sentimen pengguna terhadap berbagai isu, seperti evaluasi program PPPK, program Kartu Prakerja, serta jalur karir. Data opini yang diperoleh dari Platform X digunakan sebagai dasar dalam menganalisis pandangan masyarakat terhadap pilihan jalur karir.

Analisis sentimen pada Platform X menjadi salah satu metode yang diterapkan untuk mengidentifikasi dan mengkaji pandangan publik yang berkembang di media sosial [10]. Analisis sentimen memanfaatkan pemrosesan teks dan teknik *Natural Language Processing* untuk proses klasifikasi dalam suatu ulasan [11]. Melalui analisis sentimen, polaritas opini yang ada dapat dikumpulkan untuk menggambarkan suasana publik atau pandangan netizen, baik positif maupun negatif [12]. Algoritma *Support Vector Machine* (SVM) telah banyak digunakan dalam penelitian analisis sentimen pada tahap klasifikasi data. Hasil pengujian menunjukkan SVM mencapai akurasi 82,62%, lebih tinggi dibandingkan *Naive Bayes* yang memperoleh 79,27%, sehingga menegaskan keunggulan performa SVM. Kelebihan tersebut dipengaruhi oleh kemampuannya dalam mengatur parameter secara fleksibel, termasuk pemilihan jenis kernel dan penggunaan representasi fitur berbasis TF-IDF [13]. Hasil penelitian yang membandingkan SVM, *Naïve Bayes*, dan *Random Forest* menunjukkan bahwa SVM memperoleh performa terbaik dengan akurasi sebesar 82%. *Random Forest* dengan akurasi 81%, sedangkan *Naïve Bayes* mencatat nilai akurasi terendah yaitu 71% [14].
Berdasarkan penelitian mengenai penerapan algoritma *Naïve Bayes* dan SVM pada data *tweet* terkait program PPPK menunjukkan performa klasifikasi yang optimal, di mana SVM memperoleh akurasi tertinggi sebesar 95% pada pengujian setelah ketidakseimbangan kelas ditangani menggunakan metode *SMOTE*. Penerapan tersebut juga meningkatkan deteksi sentimen negatif dan memberikan masukan berbasis data bagi kebijakan publik [15]. Penelitian lain pada kasus program Kartu Prakerja menggunakan metode SVM menghasilkan dominasi sentimen netral sebesar 98,34%, dengan akurasi model mencapai 98,67% pada kernel linear dan 98,34% pada kernel RBF [16]. Penelitian lain yang menggunakan algoritma SVM pada analisis sentimen terkait fenomena *quarter-life crisis* di platform X mampu menggambarkan 59,3% sentimen negatif yang mencerminkan kekhawatiran terhadap masa depan dan karir, disertai 40,7% sentimen positif berupa dukungan sosial dan strategi *coping*, serta membuktikan efektivitas SVM dalam menganalisis dinamika psikososial di media sosial [17]. Penelitian lain menggunakan algoritma SVM mampu mengklasifikasikan sentimen terhadap Platform Ruang Guru, Zenius, dan Quipper dengan baik, menghasilkan akurasi tertinggi sebesar 99% pada Ruang Guru dan rata-rata akurasi yang tinggi pada keseluruhan data [18]. Penelitian lainnya menunjukkan efektivitas SVM diterapkan dalam analisis sentimen pada layanan internet seperti IndiHome, Biznet, dan Starlink. Hasil penelitian menunjukkan bahwa IndiHome mencapai performa klasifikasi yang baik dan relatif seimbang, dengan akurasi tertinggi hingga 90%, serta peningkatan performa yang cukup signifikan setelah penerapan *Random Oversampling* [19].

Berbagai penelitian sebelumnya telah menerapkan algoritma SVM untuk analisis sentimen pada berbagai isu di media sosial, seperti program PPPK, program Kartu Prakerja, *quarter-life crisis*, dan layanan digital lainnya. Penelitian ini berfokus pada analisis sentimen masyarakat terhadap tiga pilihan jalur karir, yaitu Pegawai Negeri, Pegawai Swasta, dan Wirausaha, dengan menerapkan tiga skema pembagian data, yaitu 70:30, 80:20, dan 90:10, untuk mengevaluasi performa klasifikasi menggunakan algoritma SVM. Analisis dilakukan terhadap data pada Platform X dengan mengelompokkan sentimen ke dalam kategori positif, negatif, dan netral. Penelitian ini diharapkan dapat memberikan gambaran mengenai dinamika persepsi masyarakat terhadap pilihan karir serta menyediakan landasan berbasis data yang dapat dimanfaatkan oleh pemerintah dan pemangku kepentingan dalam penyusunan kebijakan ketenagakerjaan dan strategi pengembangan sumber daya manusia.

2. Metode

2.1. Alur Penelitian



Gambar 1. Alur Penelitian

Tahapan penelitian yang ditampilkan pada Gambar 1 diawali dengan proses *scraping*, *preprocessing*, *labeling* data, serta ekstraksi fitur. Data hasil pengolahan kemudian dibagi menjadi data latih dan data uji. Kinerja model diukur menggunakan *confusion matrix* dan hasil pengolahannya ditampilkan melalui visualisasi data.

2.2. Scraping Data

Scraping merupakan teknik pengumpulan data dari sumber online dengan mengekstraksi informasi secara teknis, dan banyak digunakan dalam penelitian sosial digital berbasis pemrosesan data [20]. Proses tersebut dijalankan menggunakan bahasa pemrograman *Python* dengan dukungan alat *Tweet-Harvest* untuk memperoleh data [21] dengan memanfaatkan kata kunci dan dalam rentang waktu tertentu sebagai parameter pencarian [22].

2.3. *Preprocessing*

Preprocessing adalah tahap awal dalam klasifikasi teks yang bertujuan menyiapkan dan mengubah data teks menjadi format yang lebih sistematis agar memiliki kualitas yang lebih baik serta siap untuk dianalisis pada tahap berikutnya [23]. Proses ini mencakup sejumlah teknik, seperti data *cleaning, case folding, normalisasi, tokenisasi, stopword removal* dan *stemming*.

a. Data *Cleaning* merupakan tahapan dalam analisis sentimen yang bertujuan mengurangi *noise* pada dataset, seperti menghapus URL, simbol "@" pada *mention* seperti kata '@aku', serta karakter lain yang tidak berkaitan dengan proses analisis. Proses ini juga mencakup penghilangan data atau baris yang terduplikasi [24].
b. *Case folding* berfungsi untuk menyeragamkan penulisan teks dengan mengonversi seluruh huruf kapital menjadi huruf kecil agar konsistensi teks tetap terjaga [25].
c. *Normalisasi* dilakukan untuk menyesuaikan kata-kata yang memiliki makna serupa tetapi ditulis dalam variasi yang berbeda atau kurang tepat [26].
d. *Tokenisasi* merupakan proses segmentasi teks menjadi unit-unit kata berdasarkan susunannya. Tahapan ini dilaksanakan setelah pembersihan data dan penyeragaman huruf, dengan tujuan menguraikan kalimat menjadi elemen kata tunggal [27].
e. *Stopword removal* dilakukan untuk menghapus kata-kata yang sering muncul dalam teks tetapi tidak memiliki peran penting dalam penyampaian makna [28].
f. *Stemming* merupakan proses mengubah kata yang memiliki imbuhan ke bentuk dasar dengan cara menghilangkan awalan, akhiran, atau sisipan yang melekat pada kata tersebut. Hasil dari tahap ini berupa kata dasar yang dapat digunakan dalam pengolahan teks [29].

2.4. *Labeling* Data

Tahapan selanjutnya dilakukan dengan memberikan label pada data teks melalui penentuan kategori sentimen, yaitu positif, negatif, atau netral. Pelabelan ini berfungsi sebagai acuan dalam proses pelatihan model klasifikasi agar dapat mengidentifikasi pola sentimen secara tepat [30].

2.5. Ekstraksi Fitur

Ekstraksi fitur pada teks dokumen merupakan langkah penting dalam proses klasifikasi menggunakan metode bahasa alami dan pembelajaran mesin. Proses ekstraksi fitur dilakukan untuk menangkap kata, frasa, dan informasi penting yang terkandung dalam sebuah dokumen sehingga proses klasifikasi teks dapat dilakukan secara efektif. Ekstraksi fitur dalam penelitian ini menerapkan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) merupakan metode pemberian bobot kata dengan memperhitungkan seberapa sering suatu istilah muncul dalam dokumen serta tingkat kepentingan istilah tersebut dalam keseluruhan kumpulan data [31].

2.6. Pembagian Data

Pembagian data dilakukan sebagai tahap pemisahan dataset yang bertujuan untuk kebutuhan pelatihan dan evaluasi model. Data latih digunakan untuk membentuk model SVM agar dapat mengenali dan mempelajari pola sentimen yang terdapat pada data. Data uji berperan dalam mengevaluasi performa model saat mengklasifikasikan data baru yang belum pernah dianalisis sebelumnya [32]. Dataset dipisahkan ke dalam data latih dan data uji dengan tiga variasi rasio, yaitu 70:30, 80:20, dan 90:10, untuk menentukan komposisi pembagian yang menghasilkan tingkat klasifikasi paling baik.

2.7. Pemodelan SVM

Algoritma SVM berfokus pada pencarian *hyperplane* optimal sebagai batas keputusan untuk memisahkan data. *Hyperplane* berperan sebagai pemisah antarkelas berdasarkan karakteristik fitur yang dimiliki setiap data. Pendekatan ini bertujuan memperoleh batas pemisah terbaik sehingga data teks dapat diklasifikasikan secara akurat ke dalam kategori sentimen negatif, netral, dan positif.

Proses pengelompokan dilakukan dengan membentuk *hyperplane* linier sebagaimana ditunjukkan pada Persamaan (1). Fungsi $f(x)$ digunakan untuk memprediksi kelas target y berdasarkan fitur masukan. Parameter W menyatakan vektor pembentuk *hyperplane*, W^T merupakan *transpose* dari W dan X menunjukkan vektor fitur input sedangkan b berfungsi sebagai nilai *offset* [33].

Perhitungan pada SVM *linear* juga dapat dinyatakan pada Persamaan (2). Kernel *linear* menghitung kemiripan antara dua vektor fitur melalui operasi perkalian titik (*dot product*) secara langsung pada ruang fitur asli tanpa memerlukan transformasi ke dimensi yang lebih tinggi [32].

$$f(x) = W^T X + b = 0 \quad (1)$$

$$K(x, y) = (x.y) \quad (2)$$

2.8. Evaluasi

Tahap evaluasi dilakukan untuk mengetahui tingkat kinerja model dalam proses klasifikasi sentimen. Penilaian performa model menggunakan *accuracy*, *precision*, *recall*, *F1 – Score* sebagaimana tercantum pada Persamaan (3), (4), (5), dan (6). *Confusion matrix* yang disajikan pada Tabel 1 digunakan sebagai pendekatan untuk mengidentifikasi hasil prediksi model. Matriks ini berbentuk tabel 2×2 yang menunjukkan jumlah hasil prediksi yang sesuai maupun tidak sesuai dengan label sebenarnya, sehingga dapat digunakan untuk menilai ketepatan model dalam mengidentifikasi kelas positif, negatif, maupun netral [34].

Tabel 1. ***Confusion Matrix***

Predicted Values	*Actual Values*	
	Positive	*Negative*
Predicted Positive	TP	FP
Predicted Negative	FN	TN

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - score = \frac{2.\ recall\ .\ precision}{recall + precision} \quad (6)$$

Keterangan:

- *True Positive* (TP) menunjukkan jumlah data yang berhasil diprediksi sebagai positif dan hasil tersebut sesuai dengan label yang sebenarnya.
- *True Negative* (TN) menggambarkan banyaknya data yang diprediksi sebagai negatif dan sesuai dengan kondisi sebenarnya.
- *False Positive* (FP) menyatakan banyaknya data yang diprediksi positif padahal sebenarnya termasuk dalam kategori negatif.
- *False Negative* (FN) menunjukkan jumlah data yang dikategorikan negatif padahal label sebenarnya berada pada kelas positif.

2.9. Visualisasi

Tahap akhir dalam penelitian ini berupa penyajian data secara visual menggunakan Grafana, yaitu platform monitoring *open-source* yang berfungsi untuk memvisualisasikan, mengelola, dan menganalisis data. Visualisasi dilakukan untuk menampilkan distribusi sentimen positif, negatif, dan netral melalui berbagai bentuk tampilan seperti grafik, diagram, dan *gauge* meter sehingga hasil analisis dapat dipahami [35].

3. Hasil dan Pembahasan

Data penelitian ini dikumpulkan dari opini pengguna Platform X menggunakan metode *scraping*. Proses ini memanfaatkan pencarian berbasis kata kunci (*keyword*) untuk tiga jalur karir, yaitu Pegawai Negeri, Pegawai Swasta, dan Wirausaha. Data diambil dalam rentang waktu 1 Januari 2017 sampai dengan 30 November 2025 untuk memperoleh gambaran mengenai dinamika sentimen masyarakat terhadap pilihan jalur karir. Periode ini mencakup fase sebelum pandemi COVID-19, masa pandemi, serta periode pemulihan ekonomi, yang masing-masing berpotensi memengaruhi persepsi masyarakat terhadap pekerjaan dan kewirausahaan. Hasil pengumpulan data berdasarkan kategori serta kata kunci disajikan pada Tabel 2.

Tabel 2. Hasil ***Scraping***

Kategori Jalur Karir	Kata Kunci Pencarian	Jumlah Data
Pegawai Negeri	'Pegawai Negeri', 'Pegawai Pemerintah', 'Jadi ASN', 'Jadi PNS'	5576
Pegawai Swasta	'Pegawai Swasta', 'Karir di Swasta', 'Karyawan Swasta', 'Kerja Perusahaan Swasta'	4367
Wirausaha	'Wirausaha', 'Kewirausahaan', 'Ngurus bisnis', 'Ngurus umkm'	4730
Total		**14.212**

Tahap *preprocessing* yang dilakukan yaitu meliputi *cleaning, case folding, normalisasi, tokenisasi, stopword removal,* dan *stemming* yang ditunjukkan pada Tabel 3.

Data yang telah melewati tahap *preprocessing* menghasilkan dataset sebanyak 5.237 data Pegawai Negeri, 3.983 data Pegawai Swasta, dan 4.281 data Wirausaha. Dataset tersebut dilabeli sentimen menggunakan pendekatan *lexicon-based* dengan memberikan bobot pada setiap kata dalam rentang –5 hingga 5. Skor bernilai negatif merepresentasikan sentimen negatif, sedangkan skor bernilai positif menunjukkan sentimen positif. Skor setiap tweet diperoleh dari penjumlahan seluruh bobot kata, kemudian digunakan untuk menentukan kelas sentimen, yaitu positif apabila

skor lebih dari 0, netral apabila skor sama dengan 0, dan negatif apabila skor kurang dari 0.

Teks yang telah memiliki label tersebut selanjutnya dikonversi menjadi fitur numerik menggunakan TF-IDF, sehingga setiap dokumen menunjukkan vektor bobot kata yang digunakan sebagai input utama model SVM dalam mengklasifikasikan sentimen. Dataset selanjutnya dibagi menjadi data *training* dan data *testing* melalui tiga perbandingan pembagian, yaitu 70:30, 80:20, dan 90:10.

Tabel 3. Hasil *Preprocessing*

Tahapan	Hasil *Preprocessing*
Data Awal	@kompascom Karyawan swasta mah sudah biasa naik transportasi umum. Pejabat noh.. suruh naik transportasi umum juga gak usah dikawal pake tetot-tetot masuk jalur busway dihormati poltas pula..
Data *Cleaning*	Karyawan swasta mah sudah biasa naik transportasi umum Pejabat noh suruh naik transportasi umum juga gak usah dikawal pake tetottetot masuk jalur busway dihormati poltas pula
Case Folding	karyawan swasta mah sudah biasa naik transportasi umum pejabat noh suruh naik transportasi umum juga gak usah dikawal pake tetottetot masuk jalur busway dihormati poltas pula
Normalisasi	karyawan swasta mah sudah biasa naik transportasi umum pejabat suruh naik transportasi umum juga tidak perlu dikawal pakai tetottetot masuk jalur busway dihormati poltas pula
Tokenisasi	['karyawan', 'swasta', 'mah', 'sudah', 'biasa', 'naik', 'transportasi', 'umum', 'pejabat', 'noh', 'suruh', 'naik', 'transportasi', 'umum', 'juga', 'tidak', 'perlu', 'dikawal', 'pakai', 'tetottetot', 'masuk', 'jalur', 'busway', 'dihormati', 'poltas', 'pula']
Stopword Removal	['karyawan', 'swasta', 'transportasi', 'pejabat', 'suruh', 'transportasi', 'dikawal', 'pakai', 'masuk', 'jalur', 'busway', 'dihormati', 'poltas']
Stemming	karyawan swasta transportasi jabat suruh transportasi kawal pakai masuk jalur busway hormat poltas

Berdasarkan Tabel 4, pembagian data 90:10 memberikan nilai *accuracy* tertinggi sebesar 75,95% sehingga dianggap paling optimal untuk klasifikasi sentimen. Pembagian data 70:30 menghasilkan *precision* 69,03%, *recall* 63,72%, dan *F1-score* 65,20% sehingga dinilai paling baik berdasarkan ketiga metrik tersebut.

Tabel 4. Hasil Dataset Pegawai Negeri

Evaluasi	90:10	80:20	70:30
Accuracy	**75,95%**	74,33%	75,63%
Precision	68,14%	66,21%	**69,03%**
Recall	63,58%	61,94%	**63,72%**
F1-Score	64,88%	63,10%	**65,20%**

Berdasarkan Tabel 5, pembagian data 90:10 menghasilkan kinerja terbaik pada dataset pegawai swasta dengan nilai accuracy sebesar 74,18%, recall 57,33% dan F1-Score 56,61%. Pembagian data 70:30 menghasilkan precision 62,01% yang paling optimal untuk proses klasifikasi sentimen pada dataset tersebut.

Tabel 5. Hasil Dataset Pegawai Swasta

Evaluasi	90:10	80:20	70:30
Accuracy	**74,18%**	73,90%	73,55%
Precision	60,09%	57,97%	**62,01%**
Recall	**57,33%**	56,83%	56,03%
F1-Score	**56,61%**	55,52%	55,86%

Berdasarkan Tabel 6, pembagian data 90:10 menghasilkan kinerja terbaik pada dataset wirausaha dengan nilai *accuracy* sebesar 73,65%, *precision* sebesar 67,08%, *recall* sebesar 59,65%, dan *F1-score* sebesar 61,13%, sehingga rasio ini dipilih sebagai pembagian data yang paling optimal untuk proses klasifikasi sentimen pada dataset tersebut.

Tabel 6. Hasil Dataset Wirausaha

Evaluasi	90:10	80:20	70:30
Accuracy	**73,65%**	72,34%	72,76%
Precision	**67,08%**	63,32%	63,51%
Recall	**59,65%**	58,43%	59,12%
F1-Score	**61,13%**	59,14%	59,35%

Hasil evaluasi menunjukkan adanya perbedaan pola performa pada setiap dataset. Pada dataset pegawai negeri, nilai *accuracy* paling tinggi diperoleh pada rasio 90:10, tetapi metrik penilaian yang lain seperti *precision, recall*, serta *F1- score* justru menghasilkan nilai yang lebih baik pada rasio 70:30. Temuan ini mengindikasikan bahwa peningkatan *accuracy* tidak selalu diikuti oleh peningkatan kemampuan model dalam mengidentifikasi kelas secara seimbang. Pada dataset pegawai swasta, skema pembagian data 90:10 mampu memperoleh nilai *accuracy, recall,* dan *F1-score* yang lebih tinggi dibandingkan rasio lainnya. Nilai *precision* terbaik diperoleh pada rasio 70:30. Pada dataset wirausaha seluruh metrik evaluasi mencapai hasil tertinggi pada pembagian data 90:10, sehingga rasio tersebut dapat dianggap sebagai pembagian data yang paling tepat karena memberikan kinerja klasifikasi yang paling konsisten secara keseluruhan.

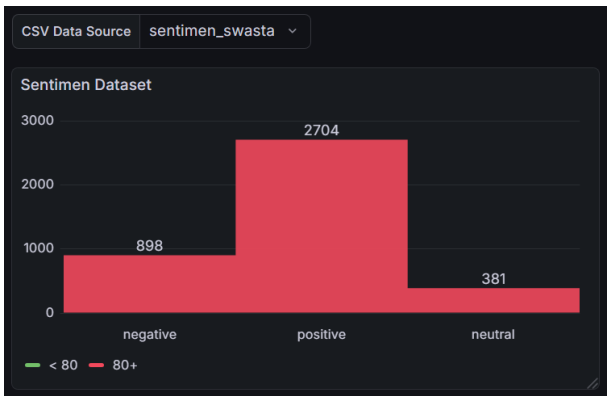
Visualisasi pada Grafana digunakan untuk menampilkan sebaran jumlah data pada setiap dataset sesuai dengan kategori sentimen positif, negatif, dan netral.

Distribusi dataset Pegawai Negeri yang ditampilkan pada Gambar 2 menunjukkan bahwa kategori positif berjumlah 3322 data, kategori negatif sebanyak 1236 data, dan kategori netral sebanyak 679 data.



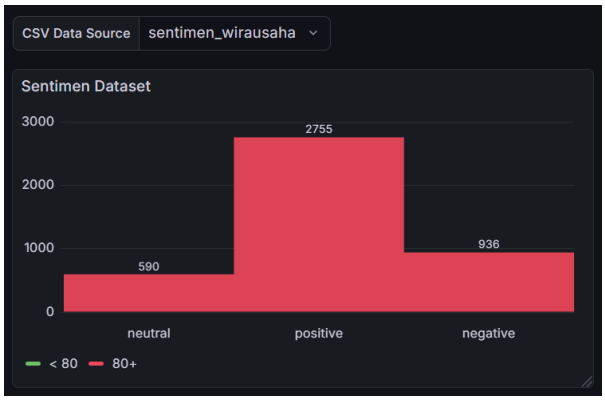
Gambar 1. Hasil Dataset Pegawai Negeri

Distribusi dataset Pegawai Swasta pada Gambar 3 memperlihatkan 2704 data positif, 898 data negatif, dan 381 data netral.



Gambar 2. Hasil Dataset Pegawai Swasta

Distribusi dataset Wirausaha pada Gambar 4 menunjukkan 2755 data pada kategori positif, 936 data pada kategori negatif, dan 590 data pada kategori netral.



Gambar 3. Hasil Dataset Wirausaha

Hasil analisis distribusi data memperlihatkan bahwa sebagian besar opini pada ketiga dataset cenderung bersifat positif, sementara sentimen negatif dan netral dalam proporsi yang lebih rendah. Perbedaan jumlah data pada setiap kategori sentimen tersebut memberikan gambaran awal mengenai karakteristik dan kecenderungan distribusi kelas sebelum proses klasifikasi dilakukan.

4. Kesimpulan

Hasil penelitian memperlihatkan bahwa metode SVM dapat mengelompokkan sentimen masyarakat terkait pilihan jalur karir di Platform X dengan performa yang cukup baik. Dari tiga skema pembagian data yang digunakan, rasio 90:10 secara konsisten menghasilkan nilai *accuracy* tertinggi pada ketiga dataset, yaitu Pegawai Negeri 75,95%, Pegawai Swasta 74,18%, dan Wirausaha 73,65%. Hasil ini menunjukkan bahwa jumlah data latih yang lebih besar berkontribusi terhadap peningkatan kemampuan model dalam mengenali pola sentimen secara lebih optimal.

Distribusi sentimen pada ketiga kategori jalur karir menunjukkan dominasi sentimen positif dibandingkan dengan sentimen negatif dan netral. Kondisi tersebut mengindikasikan bahwa masyarakat secara umum memiliki persepsi yang cenderung baik terhadap pilihan karir sebagai Pegawai Negeri, Pegawai Swasta, maupun Wirausaha. Proporsi sentimen pada setiap kategori juga menunjukkan adanya variasi pandangan masyarakat terhadap masing-masing jalur karir tersebut.

Potensi untuk pengembangan selanjutnya yang dapat dilakukan berdasarkan penelitian ini yaitu penggunaan metode lain seperti *deep learning* atau model berbasis *transformer* untuk membandingkan performa dengan SVM. Analisis tambahan berdasarkan periode waktu atau peristiwa ekonomi tertentu juga dapat memberikan gambaran yang lebih mendalam mengenai perubahan sentimen masyarakat terhadap pilihan jalur karir.

Referensi

[1]. F. Hulu, Y. Zega, M. H. Waruwu, M. Oktapiani, and P. F. Aziza, "Pengaruh Pengembangan Karir, Hubungan Kerja dan Motivasi Kerja terhadap Kepuasan Kerja Karyawan," *Edunomika*, vol. 8, no. 2, p. 2, 2024.

[2]. A. A. Rachman, A. Fauzi, D. Sudiantini, and T. Yunita, "Pengaruh Perencanaan Karir dan Engagement Karyawan Terhadap Pengembangan Karir Melalui Kepuasan Kerja Sebagai Variabel Mediating Pada PT. Inkote Indonesia," *Efisiensi : Jurnal Penelitian dan Pengembangan Ilmu Manajemen*, vol. 1, no. 2, p. 38, Feb. 2025, doi: 10.63217/efisiensi.v1i2.131.

[3]. N. Hidayati and M. Maloti, "Bimbingan Karir dalam Perencanaan Karir Remaja Akhir," *Jurnal Bintang Pendidikan Indonesia*, vol. 3, no. 1, p. 174, Dec. 2024, doi: 10.55606/jubpi.v3i1.3541.

[4]. A. Nuraisah *et al.*, "Pengaruh Persepsi Status Sosial Dan Pilihan Karier Terhadap Pengangguran Di Kalangan Sarjana," *Bursa: Jurnal Ekonomi dan Bisnis*, vol. 4, no. 2, p. 55, May 2025, doi: 10.59086/jeb.v4i2.714.

[5]. R. Fuji Oktaviani, R. Meidiyustiani, Qodariah, and A. Rachmat, "Peran Literasi Digital terhadap Wirausaha Muda: Studi Kasus pada Mahasiswa di Jakarta," *Jurnal Ekonomi Pembangunan STIE Muhammadiyah Palopo*, vol. 10, no. 1, p. 101, 2024, doi: 10.35906/jep.v10i1.1917.

[6]. "Dorong Pengembangan Kewirausahaan Nasional untuk Menopang Indonesia Menjadi Negara Maju," https://mpr.go.id/.

[7]. A. Maskhuri and T. Hardiani, "Analisis Sentimen Pengguna pada Aplikasi Tokopedia Menggunakan Algoritma Convolutional Neural Network," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 4, p. 2502, Mar. 2025, doi: 10.47065/bits.v6i4.6923.

[8]. F. Khairi Basri, M. Afdal, A. Angraini, and N. Evrilyan Rozanda, "Analisis Sentimen Masyarakat Terhadap Kebocoran Pusat Data Nasional Sementara Menggunakan Algoritma Random Forest dan Support Vector Machine," *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 2, p. 961, 2025, doi: 10.47065/bits.v7i2.7473.

[9]. D. Trezza, "To scrape or not to scrape, this is dilemma. The post-API scenario and implications on digital research," *Frontiers in Sociology*, p. 05, 2023.

[10]. A. Saepudin, R. Aryanti, E. Fitriani, Royadi, and D. Ardiansyah, "Analisis Sentimen Pemanfaatan Artificial Intelligence di Dunia Pendidikan Menggunakan SVM Berbasis Particle Swarm Optimization," *Computer Science (CO-SCIENCE)*, vol. 4, no. 1, p. 73, 2024, doi: 10.31294/coscience.v4i1.2921.

[11]. P. Sulistiyawati *et al.*, "Prediksi Kata Kasar Berbahasa Indonesia menggunakan Machine Learning Berbasis Mobile Infrastructure," *Transmisi*, vol. 24, no. 2, p. 56, May 2022, doi: 10.14710/transmisi.24.2.55-61.

[12]. S. V. Nur Afni and E. P. Silmina, "Sentiment Analysis of Public on The COVID-19," 2022. doi: 10.28989/senatik.v7i1.451.

[13]. M. Sulhan and Erizal, "Perbandingan Metode Naïve Bayes Dengan SVM Pada Analisis Sentimen Aplikasi Pemesanan Tiket Kapal Ferizy," *Technology and Science (BITS)*, vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6715.

[14]. D. A. Tarigan, Z. Situmorang, and R. Rosnelly, "Analisis Sentimen Aplikasi Playstore SIREKAP 2024 Pasca Pilpres dengan Perbandingan Metode Support Vector Machine (SVM), Naive Bayes Classifier dan Random Forest," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 11, no. 3, p. 668, 2025, doi: 10.25126/jtiik.2025129608.

[15]. L. H. Sarumpaet and R. R. Suryono, "Analisis Sentimen Publik Program PPPK di Media Sosial X menggunakan Naïve Bayes dan SVM," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 2, pp. 362–371, Aug. 2025, doi: 10.29408/edumatic.v9i2.30065.

[16]. Styawati, N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *Jurnal pengembangan IT (JPIT)*, vol. 6, no. 3, p. 155, 2021, doi: 10.30591/jpit.v6i3.2870.

[17]. T. D. Septyorini, K. Umam, and M. R. Handayani, "Dinamika Opini Publik Terkait Quarter Life Crisis Pada Media Sosial X Menggunakan Support Vector Machine," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 3, pp. 617–627, Jul. 2025, doi: 10.30591/jpit.v10i3.8648.

[18]. A. A. Munandar, Farikhin, and C. E. Widodo, "Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM," *Journal of Information Technology and Computer Science*, vol. 8, no. 2, pp. 84–84, 2023.

[19]. Z. Alfian, M. Afdal, R. Novita, and Z. Zarnelly, "Penerapan Support Vector Machine untuk Analisis Sentimen Pengguna X terhadap IndiHome, Biznet, dan Starlink," *Technology and Science (BITS)*, vol. 7, no. 2, 2025, doi: 10.47065/bits.v7i2.7429.

[20]. S. Juniarsih, E. F. Ripanti, and E. E. Pratama, "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 3, p. 239, Jul. 2020, doi: 10.26418/justin.v8i3.39118.

[21]. N. A. Rizky Putri, "Analisis Jaringan pada Media Sosial X dengan #Boikot Menggunakan Social Network Analysis," *Indonesian Journal of Information Technology*, vol. 2, no. 1, p. 2, 2024, doi: 10.25077/.

[22]. E. Fatra, T. Tri Funny Manguma, and Mastura, "Analisis Sentimen Media Sosial Menggunakan Metode Naïve Bayes," Makassar, 2025. doi: 10.31004/ctr.v2i2.166.

[23]. S. Khairunnisa, Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 408, Apr. 2021, doi: 10.30865/mib.v5i2.2835.

[24]. F. A. Ryandi, D. Pratiwi, and S. Sari, "Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM," *Jurnal Sains dan Teknologi*, vol. 7, no. 1, p. 2, 2025, doi: 10.55338/saintek.v7i1.4615.

[25]. F. M. Ayudewi and A. R. Firdaus, "Analisis Sentimen Ulasan Pengguna Platform Belajar Coursera menggunakan Model Distilbert dengan Augmentasi Data dan Penanganan Ketidakseimbangan Data," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 5, p. 7707, 2025, doi: 10.36040/jati.v9i5.14844.

[26]. A. N. Indraini and I. Ernawati, "Analisis Sentimen Terhadap Pembelajaran Daring Di Indonesia Menggunakan Support Vector Machine (SVM)," *Jurnal Ilmiah FIFO*, vol. 14, no. 1, p. 70, Jul. 2022, doi: 10.22441/fifo.2022.v14i1.007.

[27]. Herwinsyah and A. Witanti, "Analisis Sentimen Masyarakat terhadap Vaksinasi COVID-19 pada Media Sosial Twitter menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Sistem Informasi dan Informatika (Simika) P-ISSN*, vol. 5, no. 1, p. 61, 2022, doi: 10.47080/simika.v5i1.1411.

[28]. D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*, 1st ed. Depok: Gunadarma, 2023.

[29]. M. U. Albab, Y. Karuniawati P, and M. N. Fawaiq, "Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic," *Jurnal Transformatika*, vol. 20, no. 2, p. 3, Jan. 2023, doi: 10.26623/transformatika.v20i2.5374.

[30]. A. S. Rizkia, Wufron, and F. F. Roji, "Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, dan Lexicon-Based pada Performa IndoBERT," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, p. 1038, Jul. 2025, doi: 10.57152/malcom.v5i3.2151.

[31]. Y. Setiawan, D. Gunawan, and R. Efendi, "Feature Extraction TF-IDF to Perform Cyberbullying Text Classification: A Literature Review and Future Research Direction," in *2022 International Conference on Information Technology Systems and Innovation, ICITSI 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 283–288. doi: 10.1109/ICITSI56531.2022.9970942.

[32]. S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, p. 154, Oct. 2023, doi: 10.57152/malcom.v3i2.897.

[33]. A. G. Sooai, P. A. Nani, N. M. Rafu Mamulak, C. O. Sianturi, S. C. Sianturi, and A. H. Mondolang, "Klasifikasi Citra Daun Anggur Menggunakan SVM Kernel Linear," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 8, no. 1, p. 22, Mar. 2023, doi: 10.31328/jointecs.v8i1.4496.

[34]. Eskiyaturrofikoh and R. R. Suryono, "Analisis Sentimen Aplikasi X pada Google Play Store menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, p. 1414, Aug. 2024, doi: 10.29100/jipi.v9i3.5392.

[35]. R. T. Darmawan, "Implementasi Grafana untuk Auto-Scaling pada NGINX Web-Server di Lingkungan Docker Container," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 7, pp. 3165–3167, Jul. 2023, [Online]. Available: http://j-ptiik.ub.ac.id