

Dinamika Opini Publik Terkait *Quarter Life Crisis* Pada Media Sosial X Menggunakan Support Vector Machine

Talitha Dwi Septyorini ¹, Khothibul Umam ², Maya Rini Handayani ³
^{1,2,3} Universitas Islam Negeri Walisongo, Jl. Prof Hamka Ngaliyan Semarang, 50185, Indonesia

Info Artikel

Riwayat Artikel:

Received 2025-04-19

Revised 2025-06-25

Accepted 2025-06-28

Corresponding Author:

Khothibul Umam

Email:

Khothibul_umam@walisongo.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract – This study aims to analyze the dynamics of public opinion related to quarter life crisis on platform X through a sentiment analysis approach based on machine learning Support Vector Machine (SVM) algorithm is used to classify positive and negative sentiments from text data. A total of 6.312 tweets were collected with the keyword “quarter life crisis” from January 2024 to January 2025. The data was then processed through the stages of text cleaning, tokenization, stopword removal, stemming, and lexicon-based sentiment labeling. The classification process is carried out using SVM with a data division of 80% training and 20% test. The results showed an accuracy of 81.57% with a sentiment distribution of 59.3% negative and 40.7% positive. Implementation was done on Google Colab platform with evaluation using confusion matrix and classification report. The findings prove the effectiveness of SVM in analyzing psychosocial phenomena on social media and provide an empirical basis for the development of digital data-based mental health interventions. The machine learning pipeline optimized in this study can be used as a reference for other studies in analyzing psychological phenomena on social media.

Keywords: Quarter Life Crisis; Sentiment Analysis; SVM; Social Media; Text Processing

Abstrak – Penelitian ini bertujuan untuk menganalisis dinamika opini publik terkait quarter life crisis di platform X melalui pendekatan analisis sentimen berbasis machine learning. Algoritma Support Vector Machine (SVM) digunakan untuk mengklasifikasikan sentimen positif dan negatif dari data teks. Sebanyak 6.312 tweet dikumpulkan dengan kata kunci “quarter life crisis” dalam rentang waktu Januari 2024 hingga Januari 2025. Data kemudian diproses melalui tahap pembersihan teks, tokenisasi, penghapusan stopweord, stemming, dan pelabelan sentimen berbasis lexicon. Proses klasifikasi dilakukan menggunakan SVM dengan pembagian data 80% latih dan 20% uji. Hasil penelitian menunjukkan akurasi 81.57% dengan distribusi sentimen 59.3% negatif dan 40.7% positif. Implementasi dilakukan di platform Google Colab dengan evaluasi menggunakan confusion matrix dan classification report. Temuan ini membuktikan efektivitas SVM dalam menganalisis fenomena psikososial di media sosial serta memberikan dasar empiris untuk pengembangan intervensi kesehatan mental berbasis data digital. Pipeline machine learning yang dioptimasi dalam penelitian ini dapat dijadikan referensi bagi penelitian lain dalam menganalisis fenomena psikologis di media sosial.

Kata Kunci: Quarter Life Crisis, Analisis Sentimen, SVM, Media Sosial, Pemrosesan Teks

I. PENDAHULUAN

Setiap individu mengalami berbagai tahap-tahap pertumbuhan, dimulai dengan masa kanak-kanak sampai usia tua [1]. Pada fase dewasa awal, individu menghadapi transisi menuju kedewasaan yang kerap disertai keraguan, kebingungan, dan tekanan sosial. Ketidakmampuan individu dalam merespons tantangan lingkungan dapat memicu masalah psikologis, salah satunya dikenal sebagai *quarter life crisis* [1]. Fase ini umum dialami oleh individu berusia 20-30 tahun yang menghadapi berbagai perubahan, tuntutan sosial, dan ketidakpastian masa depan [2].

Dampak *quarter life crisis* beragam, mulai dari kecemasan, depresi ringan, hingga gangguan psikososial. Seiring berkembangnya media sosial, generasi muda cenderung mengekspresikan kegelisahan mereka secara terbuka di platform digital. Salah satu media sosial yang dominan digunakan adalah X (sebelumnya Twitter), yang memungkinkan pengguna berbagi opini secara real-time dan terbuka [3]. Media sosial X memungkinkan pengguna berbagi opini tanpa batas, menyebarkan berita, *retweet*, dan berkomentar. Hal ini menjadikannya media efektif untuk menyebarkan informasi cepat sekaligus mengidentifikasi sentimen publik, baik positif maupun negatif [4]. Platform X memiliki karakteristik unggul dalam menyajikan data teks pendek yang cepat tersebar dan mudah diakses, menjadikannya sumber data yang valid untuk analisis opini publik. Selain itu, fitur interaksi langsung antar pengguna dan fleksibilitas ekspresi informal menjadikan data dari X sangat representatif dalam menggambarkan sentimen masyarakat terhadap isu sosial tertentu.

Perkembangan teknologi *big data* dalam beberapa tahun terakhir mendorong kebutuhan analisis data teks berskala besar secara otomatis. Salah satu pendekatan yang banyak digunakan adalah *Natural Language Processing* (NLP), yakni cabang ilmu komputer yang berfokus pada pemrosesan dan pemahaman bahasa alami

dengan teknik komputasi. Implementasi penting NLP adalah *sentiment analysis*, yaitu proses klasifikasi teks untuk menentukan kecenderungan sentimen, baik positif, negatif, maupun netral [5]. *Sentiment analysis* telah banyak diterapkan dalam bidang bisnis, politik, dan kesehatan mental untuk memetakan opini publik berbasis data media sosial.

Salah satu algoritma populer dalam *sentiment analysis* berbasis NLP adalah *Support Vector Machine* (SVM). SVM dikenal mampu menangani data teks berdimensi tinggi dengan performa yang stabil dan hasil klasifikasi yang akurat [6]. Meskipun saat ini algoritma *deep learning* seperti LSTM dan Transformer-based models mulai banyak digunakan, penerapannya memerlukan dataset besar dan sumber daya komputasi tinggi. Dengan mempertimbangkan ukuran dataset dalam penelitian ini yang berjumlah 6.312 tweet dan keterbatasan komputasi, SVM dipilih karena lebih efisien, cepat dalam proses pelatihan, dan hasilnya lebih mudah diinterpretasikan.

Sejauh ini, penelitian tentang *quarter life crisis* lebih banyak menggunakan pendekatan kualitatif dan analisis manual. Misalnya studi [7] menemukan bahwa 4,1% kasus *quarter life crisis* pada dewasa awal dipengaruhi oleh intensitas penggunaan media sosial. Sementara itu, penelitian [8] menunjukkan bahwa media sosial dapat memengaruhi perasaan dan sikap individu, termasuk menimbulkan kekhawatiran akan kegagalan dalam karir dan pendidikan di masa depan.

Di sisi lain, metode *Support Vector Machine* (SVM) telah banyak dipublikasikan dalam analisis sentimen untuk topik lain, seperti depresi dan kecemasan [9], namun belum banyak digunakan untuk meneliti *quarter life crisis*. Contohnya, penelitian [9] menggunakan SVM untuk menganalisis sentimen terkait depresi dan kecemasan di Twitter, dengan hasil akurasi 82,5%. Studi lain membandingkan kinerja SVM dengan *Naïve Bayes* dan *Random Forest* dalam analisis sentimen kalimat depresi, di mana *Random Forest* menunjukkan performa terbaik dengan akurasi 83,33% [10]. Selain itu, penelitian [11] mendeteksi depresi dari postingan media menggunakan SVM, dengan hasil akurasi 96%.

Berdasarkan tinjauan literatur, belum ada penelitian yang secara khusus mengimplementasikan SVM untuk menganalisis sentimen publik terkait *quarter life crisis* di media sosial X. *Gap* penelitian ini semakin terlihat mengingat sebagian besar studi terdahulu masih mengandalkan pendekatan kualitatif, sementara potensi *machine learning* seperti SVM belum dieksplorasi secara optimal. Oleh karena itu, penelitian ini bertujuan mengisi kekosongan tersebut dengan menerapkan algoritma SVM untuk menganalisis dinamika opini publik tentang *quarter life crisis* di X, khususnya dalam mengklasifikasikan sentimen menjadi positif dan negatif.

II. METODE

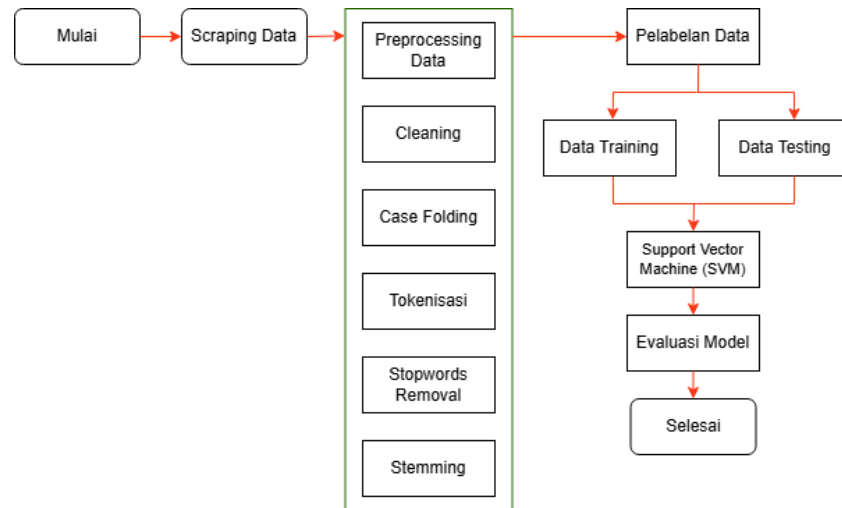
Penelitian ini bertujuan mengklasifikasikan sentimen terkait *quarter life crisis* di media sosial X menggunakan algoritma SVM. Metode yang digunakan meliputi pengumpulan data, *preprocessing*, pelabelan, pembagian dataset, pelatihan model, dan evaluasi dengan Google Colab sebagai platform utama. Penelitian ini menggunakan Google Colab sebagai platform berbasis komputasi awan (*cloud*) yang memungkinkan penulisan dan eksekusi kode Python langsung melalui *web browser*. Google Colab telah menjadi platform komputasi berbasis awan (*cloud*) yang banyak dimanfaatkan komunitas riset dan pengembangan, khususnya ilmuwan data, peneliti, dan pengembang, untuk menunjang beragam aktivitas komputasi seperti pra-pemrosesan data, konstruksi model AI, eksplorasi dataset, hingga pelatihan algoritma *machine learning* [12]. Adapun tahap penelitian yang dijalankan pada penelitian ini dengan Google Colab meliputi:

1. Pengumpulan Data: Mengumpulkan sebanyak 6.312 tweet data dari platform media sosial X dengan menggunakan kata kunci "*quarter life crisis*" dengan menggunakan library *tweet harvest* sebagai dataset.
2. *Preprocessing* Data: Meliputi pembersihan teks (*cleaning*), konversi ke huruf kecil (*case folding*), tokenisasi, penghilangan *stopwords* dan *stemming* untuk mempersiapkan data analisis.
3. Pelabelan Data: Menggunakan pendekatan *lexicon-based* dengan memanfaatkan kamus kata yang telah diberi label positif dan negatif. Setiap *tweet* dianalisis berdasarkan kemunculan kata-kata dalam lexicon tersebut, kemudian diberi skor sentimen yang akan menjadi dasar klasifikasi. Meskipun metode *lexicon-based* memiliki keunggulan dari segi kecepatan dan kemudahan implementasi, pendekatan ini memiliki keterbatasan dalam mendeteksi makna ambigu, istilah slang, atau konteks sarkasme yang tidak tercakup dalam kamus. Oleh karena itu, untuk penelitian selanjutnya dapat dipertimbangkan penerapan *manual labeling* pada sebagian data sebagai *ground truth* guna memvalidasi hasil pelabelan otomatis serta meningkatkan akurasi klasifikasi sentimen secara keseluruhan.
4. Pembagian Data: Dataset yang telah diproses dibagi menjadi data latih (*training*) dan data uji (*testing*) dengan rasio 80:20. Pembagian ini bertujuan untuk memvalidasi model, dimana data latih (*training*) berfungsi sebagai dasar pembentukan model, sementara data uji (*testing*) bertujuan mengukur akurasi dan efektivitas model tersebut.
5. Pelatihan model SVM: Algoritma *Support Vector Machine* (SVM) dipilih untuk membangun model klasifikasi sentimen karena kemampuannya menangani data teks berdimensi tinggi. Model dilatih

menggunakan data training yang telah diproses untuk mengenai pola sentimen dalam *tweet* terkait *quarter life crisis*.

6. Evaluasi Model: Evaluasi performa model dilakukan menggunakan *confusion matrix* untuk melihat jumlah prediksi benar dan salah, perhitungan *accuracy* untuk mengukur persentase prediksi benar secara keseluruhan, serta *classification report* yang menyajikan matriks *precision*, *recall*, dan *F1-score* untuk mengevaluasi kinerja model secara lebih mendetail. Hasil evaluasi ini menjadi dasar penilaian efektivitas model dalam mengklasifikasikan sentimen.

Berikut alur atau tahapan proses yang akan digunakan dalam penelitian ini ditunjukkan pada gambar 1.



Gambar 1. Diagram alur proses

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Metode pengumpulan data ada dua cara yaitu pengumpulan data dengan metode manual dan otomatis. Metode manual dilakukan dengan menyalin data secara langsung melalui *copy-paste* dari sumber tertentu, sedangkan metode otomatis memanfaatkan *coding*, aplikasi, atau ekstensi *browser* untuk mengumpulkan data secara efisien [13]. Dalam penelitian ini, menerapkan *web scraping* sebagai metode pengumpulan data otomatis. *Web scraping* adalah teknik ekstraksi dalam jumlah besar dari sumber online, seperti situs web atau platform media sosial, untuk keperluan analisis. Teknik *web scraping* telah banyak diaplikasikan dalam berbagai bidang penelitian, termasuk pengembangan platform web, analisis opini publik (sentimen), serta studi komparatif harga produk [14]. Data dikumpulkan dari platform media sosial X menggunakan library *Tweet Harvest* dengan kata kunci “*quarter life crisis*” selama periode Januari 2024 hingga Januari 2025. Data yang terkumpul pada periode tersebut berjumlah 6. 312 tweet dan seluruh data tersebut digunakan dalam penelitian ini untuk proses analisis sentimen. Proses *scraping* dijalankan di Google Colab untuk mengoptimalkan ekstraksi data publik. Dataset yang digunakan dalam pemodelan dari hasil *import tweet* menggunakan Google Colab ditunjukkan pada tabel 1.

TABEL 1
SAMPel DATASET AWAL

Waktu	Full Tweet
1/30//2024	BENERAN JADI DEWASA DAN MULAI NGALAMIN QUARTER LIFE CRISIS ITU ADALAH FASE YG GAENAK BGT.
3/26/2024	masih ga nyangka anak kecil ini ternyata memasuki fase quarter-life crisis yang kalo bengong dikit kepikiran hidup arahnya mau dibawa kemana ngebandingin pencapaian diri sendiri sama orang lain ngerasa stuck gaada perkembangan dan ngerasa bersalah tiap seneng dikit.
1/30/2025	Umur 20an dan sedang menghadapi quarter life crisis itu emang isinya nangis doang yh?

menghasilkan teks lebih bersih dan terstruktur, seperti terlihat pada tabel 2, sehingga siap untuk tahap *preprocessing* berikutnya.

TABEL 2
 SAMPEL HASIL TWEET CLEANING

Full Tweet	Tweet Cleaning
Udh ngerasa banget yg namanya quarter life crisis emng tumbuh dewasa tu ga seindah yg di bayangkan selama ini. Aku ga panik santai aja krna semua tu normal. Kebanyakan org jga ngalamin hal yg serupa kok emng udh fase kehidupan ga si wk	Udh ngerasa banget yg namanya quarter life crisis emng tumbuh dewasa tu ga seindah yg di bayangkan selama ini Aku ga panik santai aja krna semua tu normal Kebanyakan org jga ngalamin hal yg serupa kok emng udh fase kehidupan ga si wk
besok udah genap 20 tahun aja tapi hidup udah semorat marit ini wkwk kan kata orang2 nanti bakalan ngerasain fase quarter life crisis diumur 23 tahun keatas ya? tapi gua dari umur 17/18 udah ngerasain quarter life crisis itu	besok udah genap tahun aja tapi hidup udah semorat marit ini wkwk kan kata orang nanti bakalan ngerasain fase quarter life crisis diumur tahun keatas ya tapi gua dari umur udah ngerasain quarter life crisis itu
@Seeredith Berjuang melawan quarter life crisis.	Berjuang melawan quarter life crisis

2) *Case folding*

Case folding merupakan teknik yang digunakan untuk mengubah semua karakter teks menjadi huruf kecil (*lowercase*) guna menciptakan keseragaman dalam data [17], seperti terlihat pada tabel 3. Tahap ini penting untuk menghilangkan variasi akibat perbedaan kapitalisasi dan memastikan konsistensi data sebelum analisis lebih lanjut. Dengan konversi ke huruf kecil, kata seperti “*QUARTER*”, “*Quarter*”, dan “*quarter*” akan diperlakukan sama oleh sistem.

TABEL 3
 SAMPEL HASIL TWEET CASE FOLDING

Tweet Cleaning	Tweet Case Folding
Udh ngerasa banget yg namanya quarter life crisis emng tumbuh dewasa tu ga seindah yg di bayangkan selama ini Aku ga panik santai aja krna semua tu normal Kebanyakan org jga ngalamin hal yg serupa kok emng udh fase kehidupan ga si wk	udh ngerasa banget yg namanya quarter life crisis emng tumbuh dewasa tu ga seindah yg di bayangkan selama ini aku ga panik santai aja krna semua tu normal kebanyakan org jga ngalamin hal yg serupa kok emng udh fase kehidupan ga si wk
besok udah genap tahun aja tapi hidup udah semorat marit ini wkwk kan kata orang nanti bakalan ngerasain fase quarter life crisis diumur tahun keatas ya tapi gua dari umur udah ngerasain quarter life crisis itu	besok udah genap tahun aja tapi hidup udah semorat marit ini wkwk kan kata orang nanti bakalan ngerasain fase quarter life crisis diumur tahun keatas ya tapi gua dari umur udah ngerasain quarter life crisis itu
Berjuang melawan quarter life crisis	berjuang melawan quarter life crisis

3) *Tokenisasi*

Seperti ditunjukkan pada tabel 4, proses tokenisasi yaitu memecah teks menjadi unit-unit kata individual (token) [18]. Proses tokenisasi dilakukan menggunakan library NLTK dengan fungsi `word_tokenize` yang tersedia dalam modul `nltk.tokenize`. Tahap ini merupakan fondasi penting untuk analisis teks pada level kata, dimana kalimat utuh diurai menjadi komponen-komponen penyusunnya. Hasil tokenisasi kemudian menjadi input untuk tahap *preprocessing* berikutnya.

TABEL 4
 SAMPEL HASIL TWEET TOKENISASI

Tweet Case Folding	Tweet Tokenisasi
udh ngerasa banget yg namanya quarter life crisis emng tumbuh dewasa tu ga seindah yg di bayangkan selama ini aku ga panik santai aja krna semua tu normal kebanyakan org jga ngalamin hal yg serupa kok emng udh fase kehidupan ga si wk	[udh, ngerasa, banget, yg, namanya, quarter, life, crisis, emng, tumbuh, dewasa, tu, ga, seindah, yg, di, bayangkan, selama, ini, aku, ga, panik, santai, aja, krna, semua, tu, normal, kebanyakan, org, jga, ngalamin, hal, yg, serupa, kok, emng, udh, fase, kehidupan, ga, si, wk]
besok udah genap tahun aja tapi hidup udah semorat marit ini wkwk kan kata orang nanti bakalan ngerasain fase quarter life crisis diumur tahun keatas	[besok, udah, genap, tahun, aja, tapi, hidup, udah, semorat, marit, ini, wkwk, kan, kata, orang, nanti, bakalan, ngerasain, fase, quarter, life, crisis, diumur,

ya tapi gua dari umur udah ngerasain quarter life crisis itu	tahun, keatas, ya, tapi, gua, dari, umur, udah, ngerasain, quarter, life, crisis, itu]
berjuang melawan quarter life crisis	[berjuang, melawan, quarter, life, crisis]

4) *Stopword removal*

Proses penyaringan *stopwords*, sebagaimana diilustrasikan pada tabel 5, bertujuan menghilangkan kata penghubung yang tidak bermakna substantif [19]. Menggunakan daftar *stopwords* bahasa Indonesia yang telah dikurasi, tahap ini secara efektif mengurangi dimensi data sekaligus memfokuskan analisis pada kata kunci yang lebih relevan dengan topik penelitian.

TABEL 5
SAMPEL HASIL TWEET STOPWORD REMOVAL

Tweet Tokenisasi	Tweet Stopword Removal
[udh, ngerasa, banget, yg, namanya, quarter, life, crisis, emng, tumbuh, dewasa, tu, ga, seindah, yg, di, bayangkan, selama, ini, aku, ga, panik, santai, aja, krna, semua, tu, normal, kebanyakan, org, jga, ngalamin, hal, yg, serupa, kok, emng, udh, fase, kehidupan, ga, si, wk]	[udh, ngerasa, banget, yg, namanya, quarter, life, crisis, emng, tumbuh, dewasa, tu, ga, seindah, yg, bayangkan, ga, panik, santai, aja, krna, tu, normal, kebanyakan, org, jga, ngalamin, yg, emng, udh, fase, kehidupan, ga, si, wk]
[besok, udah, genap, tahun, aja, tapi, hidup, udah, semorot, marit, ini, wkwk, kan, kata, orang, nanti, bakalan, ngerasain, fase, quarter, life, crisis, diumur, tahun, keatas, ya, tapi, gua, dari, umur, udah, ngerasain, quarter, life, crisis, itu]	[besok, udah, genap, aja, hidup, udah, semorot, marit, wkwk, orang, ngerasain, fase, quarter, life, crisis, diumur, keatas, ya, gua, umur, udah, ngerasain, quarter, life, crisis]
[berjuang, melawan, quarter, life, crisis]	[berjuang, melawan, quarter, life, crisis]

5) *Stemming*

Seperti terlihat pada tabel 6, algoritma *stemming* berfungsi dalam mentransformasikan berbagai varian kata berimbuhan menjadi bentuk kanoniknya dengan cara menghilangkan awalan, akhiran, atau sisipan kata [20]. Proses *stemming* dilakukan menggunakan algoritma Sastrawi, sebuah library *stemming* Bahasa Indonesia berbasis metode Nazief & Adriani yang telah diimplementasikan secara luas dalam analisis teks berbahasa Indonesia. Untuk bahasa Indonesia yang memiliki karakteristik morfologis kompleks, proses ini sangat penting dalam mengelompokkan variasi kata dengan makna serupa. Hasil *stemming* secara signifikan meningkatkan akurasi analisis dengan mengurangi redundansi dalam dataset.

TABEL 6
SAMPEL HASIL TWEET STEMMING

Tweet Stopword Removal	Tweet Stemming
[udh, ngerasa, banget, yg, namanya, quarter, life, crisis, emng, tumbuh, dewasa, tu, ga, seindah, yg, bayangkan, ga, panik, santai, aja, krna, tu, normal, kebanyakan, org, jga, ngalamin, yg, emng, udh, fase, kehidupan, ga, si, wk]	[udh, ngerasa, banget, yg, nama, quarter, life, crisis, emng, tumbuh, dewasa, tu, ga, indah, yg, bayang, ga, panik, santai, aja, krna, tu, normal, banyak, org, jga, ngalamin, yg, emng, udh, fase, hidup, ga, si, wk]
[besok, udah, genap, aja, hidup, udah, semorot, marit, wkwk, orang, ngerasain, fase, quarter, life, crisis, diumur, keatas, ya, gua, umur, udah, ngerasain, quarter, life, crisis]	[besok, udah, genap, aja, hidup, udah, semorot, marit, wkwk, orang, ngerasain, fase, quarter, life, crisis, umur, atas, ya, gua, umur, udah, ngerasain, quarter, life, crisis]
[berjuang, melawan, quarter, life, crisis]	[juang, lawan, quarter, life, crisis]

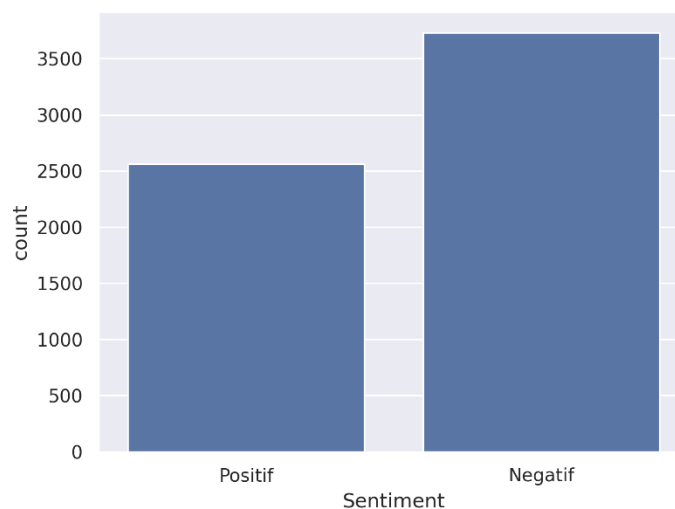
C. *Pelabelan Data*

Pelabelan data (*data labeling*) merupakan tahap kritis dalam mempersiapkan data mentah untuk pelatihan model *machine learning*. Dalam pelatihan model analisis sentimen, *tweet* perlu diberi label positif atau negatif sebagai data referensi [21]. Proses ini menganotasi data dengan label bermakna yang memberikan konteks dan kategorisasi, sehingga model dapat mempelajari pola secara efektif. Pada penelitian ini, pelabelan sentimen dilakukan dengan pendekatan *lexicon-based*, yaitu metode berbasis kamus kata yang telah diklasifikasikan berdasarkan polaritas positif negatif. Media sosial, khususnya platform X, telah menjadi ruang digital terbuka bagi individu untuk mengekspresikan berbagai bentuk emosi secara bebas, termasuk kecemasan, kemarahan, dan ketakutan terhadap masa depan. Seperti ditunjukkan pada gambar 4, hasil analisis menggunakan *lexicon InSet* (*Indonesian Sentiment Lexicon*) mengungkapkan bahwa mayoritas ulasan pengguna dalam dataset ini cenderung

negatif dibandingkan positif. Dominasi sentimen negatif ini sejalan dengan hasil penelitian [22] yang menunjukkan bahwa gangguan psikososial remaja seperti gangguan makan, kecemasan, hingga depresi merupakan dampak dari penggunaan media sosial dengan intensitas yang tinggi. Sebagai ilustrasi, tabel 7 menyajikan contoh hasil pelabelan sentimen terhadap beberapa tweet dalam dataset berdasarkan pendekatan *lexicon-based* menggunakan *lexicon InSet*.

TABEL 7
SAMPEL HASIL PELABELAN DATA

Full Tweet	Tweet Cleaned	Score	Sentimen
BENERAN JADI DEWASA DAN MULAI NGALAMIN QUARTER LIFE CRISIS ITU ADALAH FASE YG GAENAK BGT.	beneran dewasa ngalamin quarter life crisis fase yg gaenak bgt	0	Negatif
masih ga nyangka anak kecil ini ternyata memasuki fase quarter-life crisis yang kalo bengong dikit kepikiran hidup arahnya mau dibawa kemana ngebandingin pencapaian diri sendiri sama orang lain ngerasa stuck gaada perkembangan dan ngerasa bersalah tiap seneng dikit.	ga nyangka anak masuk fase quarterlife crisis kalo bengong dikit pikir hidup arah bawa mana ngebandingin capai orang ngerasa stuck gaada kembang ngerasa salah neng dikit	0	Negatif
Udh ngerasa banget yg namanya quarter life crisis emng tumbuh dewasa tu ga seindah yg di bayangkan selama ini. Aku ga panik santai aja krna semua tu normal. Kebanyakan org jga ngalamin hal yg serupa kok emng udh fase kehidupan ga si wk □	udh ngerasa banget yg nama quarter life crisis emng tumbuh dewasa tu ga indah yg bayang ga panik santai aja krna tu normal banyak org jga ngalamin yg emng udh fase hidup ga si wk	6	Positif
pikirannya sudah bercabang nggak bisa cuma mikir main doang kaya dulu welcome to quarter life crisis	pikir cabang nggak mikir main doang kaya welcome to quarter life crisis	6	Positif

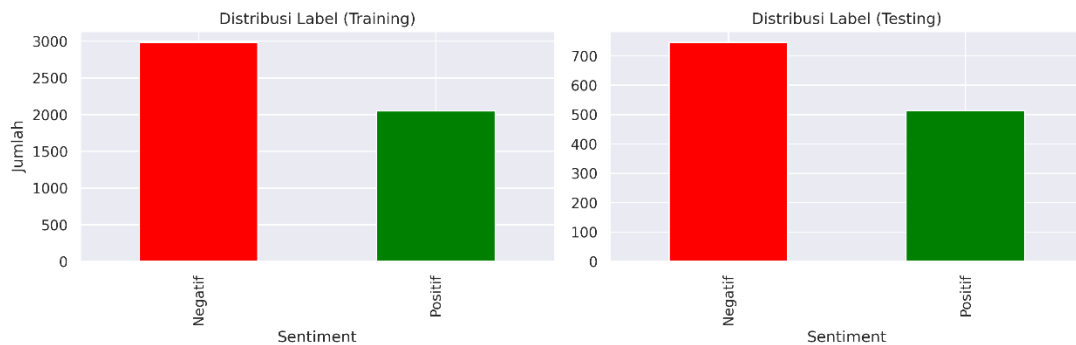


Gambar 4. Diagram analisis sentimen

D. Pembagian Data

Dalam penelitian ini, dataset dibagi dengan rasio 80:20 (*training:test*) menggunakan *stratified sampling*, menghasilkan 5.034 sampel *training* dan 1.259 sampel *testing*. Proporsi label negatif (59.3%) dan positif (40.7%) dipertahankan pada kedua subset seperti terlihat pada gambar 5. Proses pembagian dataset diimplementasikan melalui fungsi `train_test_split` library Scikit-learn dengan konfigurasi spesifik: proporsi data uji 20% (`test_size=0.2`), stratifikasi berdasarkan label kelas (`stratify=y`), dan pengacakan terkontrol (`random_state=42`) guna menjamin kestabilan kelas dan konsistensi hasil eksperimen. Data *training* berfungsi membangun model

dan menentukan parameter optimal, sementara data *testing* berperan mengukur performa model terhadap data baru sekaligus mencegah *overfitting* [13].



Gambar 5. Diagram jumlah data *training* dan data *testing*

E. Support Vector Machine (SVM)

Penelitian ini memanfaatkan SVM sebagai model klasifikasi yang mengadopsi prinsip *Structural Risk Minimization* [23] yang bertujuan menemukan *hyperlane* optimal untuk memisahkan dua kelas – dalam penelitian ini digunakan untuk klasifikasi sentimen positif dan negatif. Implementasi menggunakan kernel linear karena karakteristik data teks yang bersifat *high-dimensional sparse*, sehingga kernel linear lebih efisien dan sesuai untuk memisahkan kelas data teks tanpa memerlukan transformasi ke dimensi yang lebih tinggi. Kernel RBF dan non-linear lainnya telah dipertimbangkan, namun berdasarkan studi literatur dan pengujian awal, kernel linear menunjukkan performa akurasi serupa dengan waktu komputasi yang jauh lebih efisien untuk dataset berukuran menengah seperti penelitian ini.

Tahap ekstraksi fitur mengkombinasikan pembobotan TF-IDF untuk unigram dan bigram, diperkuat dengan seleksi fitur melalui uji chi-kuadrat guna meningkatkan relevansi fitur seperti pada gambar 6. Parameter `max_features=5000` dipilih berdasarkan hasil eksperimen awal yang menunjukkan nilai ini mampu menyeimbangkan antara kompleksitas model, waktu komputasi, dan akurasi prediksi. Nilai di atas 5000 tidak memberikan peningkatan akurasi signifikan, sementara nilai di bawahnya cenderung menurunkan performa model.

Proses tuning parameter mengimplementasikan grid search dengan 10-fold cross validation seperti ditunjukkan pada gambar 7, untuk mencapai parameter regularisasi terbaik. Rentang nilai `C = (0.1, 1, 3)` dipilih sebagai variasi nilai awal yang umum digunakan dalam pemodelan SVM, karena nilai `C` yang kecil cenderung menghasilkan model yang lebih sederhana, sedangkan nilai yang lebih besar memberikan toleransi lebih rendah terhadap kesalahan klasifikasi. Pemilihan nilai tersebut juga didasarkan pada hasil uji coba awal yang menunjukkan bahwa rentang ini memberikan hasil akurasi yang stabil tanda indikasi *overfitting*. Validasi dilakukan menggunakan *stratified cross-validation* untuk menjaga proporsi kelas seimbang pada setiap fold, sehingga evaluasi model lebih konsisten.

```
vectorizer = TfidfVectorizer(ngram_range=(1,2), stop_words=stopwords_combined, max_features=5000)
x_train_tfidf = vectorizer.fit_transform(x_train)
x_test_tfidf = vectorizer.transform(x_test)

selector = SelectKBest(chi2, k=min(5000, x_train_tfidf.shape[1]))
x_train_tfidf = selector.fit_transform(x_train_tfidf, y_train)
x_test_tfidf = selector.transform(x_test_tfidf)
```

Gambar 6. Tahap ekstraksi fitur

```
param_grid = {
    'C': [0.1, 1, 3],
    'kernel': ['linear'],
    'gamma': ['scale']
}
cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)
grid_search_svm = GridSearchCV(svm, param_grid, cv=cv, scoring='accuracy', n_jobs=-1, verbose=2)
grid_search_svm.fit(x_train_tfidf, y_train)
```

Gambar 7. Tahap optimasi model

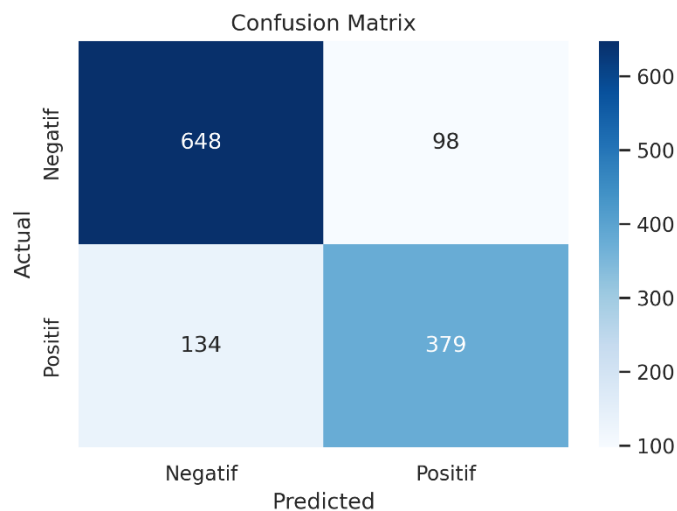
F. Evaluasi Model

Proses evaluasi model klasifikasi dilakukan dengan menganalisis nilai akurasi dan *confusion matrix* guna menilai kesesuaian model dengan tujuan yang diharapkan [24]. Evaluasi komprehensif ini menghasilkan

confusion matrix yang menggambarkan distribusi prediksi benar dan salah, serta *classification report* yang memuat nilai *precision* untuk mengukur ketepatan prediksi positif, *recall* yang menunjukkan kemampuan model dalam mengidentifikasi kasus positif aktual, dan *F1-score* sebagai rata-rata harmonik antara *precision* dan *recall*. Selain itu, evaluasi juga memperhitungkan *support* yang merepresentasikan jumlah sampel pada setiap kelas. Melalui analisis terhadap berbagai matriks ini, dapat diperoleh pemahaman menyeluruh mengenai performa model, termasuk keunggulan dalam menangani pola tertentu maupun keterbatasan yang perlu diperbaiki, sehingga memberikan dasar objektif untuk menilai kesiapan model dalam aplikasi nyata.

1) *Confusion Matrix*

Evaluasi performa SVM menggunakan *confusion matrix* yang membandingkan prediksi dengan label aktual melalui empat kuadran pengukuran: benar positif (TP), benar negatif (TN), salah positif (FP), dan salah negatif (FN) [25]. Analisis ini membantu mengidentifikasi pola kesalahan klasifikasi model secara detail. Hasil prediksi dari *confusion matrix* berbentuk diagram, seperti pada gambar 8. Nilai-nilai ini menjadi dasar perhitungan metrik evaluasi lain seperti *precision*, *recall*, dan *F1-score*.



Gambar 8. Diagram *confusion matrix*

2) *Accuracy*

Akurasi sebesar 81.57% menunjukkan bahwa model SVM mampu memprediksi dengan benar sekitar 81-82% dari total data uji. Nilai ini mengindikasikan tingkat kedekatan antara hasil prediksi model dengan nilai sebenarnya yang cukup baik. Nilai akurasi merupakan indikator yang menunjukkan sejauh mana model berhasil melakukan prediksi dengan benar [26].

Perhitungan akurasi dalam penelitian ini didasarkan pada rasio jumlah prediksi benar (baik positif maupun negatif) terhadap total prediksi, seperti ditunjukkan pada persamaan berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$= \frac{379+648}{379+648+98+134} = \frac{1027}{1259} \approx 81.57\%$$

Nilai akurasi 81.57% yang dicapai oleh model SVM dapat dikategorikan baik dalam konteks klasifikasi sentimen media sosial. Berdasarkan hasil uji coba internal, model Naive Bayes hanya mampu mencapai akurasi 74.90%, sedangkan Random Forest memperoleh akurasi 73.95%. Dengan demikian, capaian SVM lebih unggul sekitar 6-7% dibanding kedua model tersebut. Hal ini menunjukkan bahwa algoritma SVM lebih efektif dalam menangani karakteristik data media sosial yang cenderung tidak terstruktur, mengandung banyak bahasa informal, singkatan, dan konteks implisit. Performa ini juga mengindikasikan bahwa strategi *preprocessing* teks serta tuning parameter model, termasuk pemilihan kernel linear dan nilai C optimal, turut berkontribusi signifikan dalam peningkatan akurasi model. Dengan akurasi di atas 80%, model SVM dapat dianggap andal untuk diterapkan dalam analisis sentimen fenomena psikologis di media sosial.

3) Classification Report

Laporan klasifikasi merupakan dokumen evaluasi yang menyajikan matriks-matriks penting untuk menilai performa model klasifikasi, meliputi precision, recall, F1-score, dan support. Precision merepresentasikan akurasi prediksi positif dengan membandingkan prediksi benar positif terhadap total prediksi positif, dimana nilai tinggi menunjukkan sedikitnya kesalahan false positive [27]. Recall merupakan matriks yang menghitung jumlah prediksi dengan kelas yang sama secara [28]. Sedangkan F1-score merupakan rata-rata harmonik precision dan recall yang menunjukkan keseimbangan antara kedua matriks tersebut [29]. Dalam penelitian ini, hasil classification report model SVM yang mencakup nilai-nilai precision, recall, F1-score, dan support dapat dilihat pada gambar 9.

Laporan Klasifikasi:				
	precision	recall	f1-score	support
Negatif	0.83	0.87	0.85	746
Positif	0.79	0.74	0.77	513
accuracy			0.82	1259
macro avg	0.81	0.80	0.81	1259
weighted avg	0.81	0.82	0.81	1259

Gambar 9. Hasil klasifikasi report

IV. SIMPULAN

Penelitian ini berhasil menganalisis dinamika opini publik terkait quarter-life crisis di platform X menggunakan pendekatan machine learning berbasis SVM dengan akurasi klasifikasi mencapai 81.57%. Hasil analisis sentimen mengungkap bahwa 59.3% ekspresi pengguna bersifat negatif, didominasi oleh kekhawatiran akan masa depan, tekanan finansial, dan ketidakpastian karir, sementara 40.7% menunjukkan sikap positif berupa dukungan sosial dan strategi coping. Temuan ini memberikan bukti empiris bahwa media sosial dapat berfungsi sebagai cermin psikologis generasi muda, sekaligus menegaskan efektivitas pipeline NLP berbasis SVM dalam menangkap fenomena psikososial kompleks. Bagi praktisi kesehatan mental, hasil penelitian dapat menjadi baseline data untuk pengembangan program intervensi berbasis digital, sementara bagi akademisi, studi ini memperkaya metodologi analisis sentimen untuk isu-isu psikologis di ruang digital. Keberhasilan model dalam mempertahankan F1-score 0.82 pada kelas minoritas menunjukkan kelayakannya sebagai alat monitoring real-time untuk mendeteksi gejala *quarter-life crisis* dalam skala populasi. Potensi pengembangan selanjutnya mencakup penerapan model berbasis *deep learning* seperti LSTM atau BERT untuk menangkap konteks semantik yang lebih kompleks, penerapan manual labeling sebagian data sebagai *ground truth* guna meningkatkan validitas pelabelan *lexicon-based*, serta pengembangan sistem monitoring real-time berbasis API untuk deteksi dini fenomena psikososial di media sosial.

UCAPAN TERIMAKASIH

Peneliti menyampaikan rasa terima kasih kepada Fakultas Sains dan Teknologi Universitas Islam Negeri Walisongo Semarang atas bantuan dan fasilitas yang diberikan selama pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] P. R. Pamungkas and G. Hendrastomo, "Quarter life crisis di kalangan mahasiswa," *Saskara: Indonesian Journal of Society Studies*, vol. 4, no. 1, pp. 174–188, 2024.
- [2] Karpika I Puti and Segel Ni Wayan Widiyani, "Quarter Life Crisis Terhadap Mahasiswa Studi Kasus di Fakultas Keguruan dan Ilmu Pendidikan Universitas PGRI Mahadewa Indonesia," *Widyadari*, vol. 22, no. 2, pp. 513–527, 2021, doi: 10.5281/zenodo.5550458.
- [3] Kuku Wijayanti and Qoniah Nur Wijayani, "Peranan Aplikasi Twitter Atau X Dalam Interaksi Komunikasi Guna Membantu Penyeimbangan Kesehatan Mental Pada Remaja Saat Ini," *Journal Sains Student Research*, vol. 2, no. 1, pp. 07–15, 2023, doi: 10.61722/jssr.v2i1.469.
- [4] L. Nursingah, R. Ruuhwan, and T. Mufizar, "Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis Dengan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4336.
- [5] M. A. Amrustian, W. Widayat, and A. M. Wirawan, "Analisis Sentimen Evaluasi Terhadap Pengajaran Dosen di Perguruan Tinggi Menggunakan Metode LSTM," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 535, Jan. 2022, doi: 10.30865/mib.v6i1.3527.
- [6] H. I. Pratama and P. T. Prasetyaningrum, "Penerapan Support Vector Machine untuk Analisis Sentimen pada Google Review Hotel," vol. 6, no. 2, pp. 1244–1252, 2025, doi: 10.47065/josh.v6i2.6645.
- [7] H. Alkatiri and R. A. Aprianty, "Pengaruh Intensitas Penggunaan Media Sosial pada Quarter Life Crisis pada Dewasa Awal," *Jurnal Psikologi*, vol. 1, no. 2, p. 7, 2024, doi: 10.47134/pjp.v1i2.2269.

- [8] S. Amanda Permatasari, Mohammad Ammar Marsa, "Dampak Media Sosial dalam Quarter Life Crisis Gen Z di Indonesia," *Syntax Literate: Jurnal Ilmiah Indonesia*, vol. 7, no. Vol 7 No.6 (2022): Syntax Literate: Jurnal Ilmiah Indonesia, 2022.
- [9] F. Darmawan, M. Joe, Y. I. Kurniawan, and L. Afuan, "Analisis Sentimen Kemungkinan Depresi dan Kecemasan pada Twitter Menggunakan Support Vector Machine," *Jurnal Eksplora Informatika*, vol. 13, no. 1, pp. 24–36, 2023, doi: 10.30864/eksplora.v13i1.854.
- [10] M. Fachriza and Munawar, "Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest," *Jurnal Teknik Universitas Muhammadiyah Ponorogo*, pp. 49–58, 2023, [Online]. Available: <http://studentjournal.umpo.ac.id/index.php/komputek>
- [11] K. Putri *et al.*, "Implementasi Algoritma Support Vector Machine dalam Klasifikasi Deteksi Depresi dari Postingan pada Media Sosial," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 2, no. 1, pp. 193–202, 2023.
- [12] R. Andarsyah and A. Yanuar, "Sentimen Analisis Aplikasi Posaja Pada Google Playstore Untuk Peningkatan Pospay Superapp Menggunakan Support Vector Meachine," *Jurnal Teknik Informatika*, vol. 16, no. Vol. 16, No. 2, April 2024, 2024.
- [13] N. Aulia, S. N. Sari, and N. Wakhidah, "Jurnal Informatika : Jurnal pengembangan IT Analisis Sentimen Aplikasi Get Contact di APP Store Menggunakan Metode SVM (Support Vector Machine)," vol. 10, no. 1, pp. 139–148, 2025, doi: 10.30591/jpit.v9ix.xxx.
- [14] K. Dwicahyo and C. Indah Ratnasari, "Perbandingan Metode Web Scraping Dalam Pengambilan Data: Kajian Literatur," *Automata*, vol. 4, 2023.
- [15] N. A. Haqimi and T. A. Roshinta, "Analisis Komentar Spam Instagram menggunakan Support Vector Machine dengan Variasi Hyperparameter," vol. 9, no. 3, pp. 242–253, 2024, doi: 10.30591/jpit.v9i3.7834.
- [16] A. Fathiarahma, A. Youtama, T. Ridwan, and N. Heryana, "Analisis Text Mining Klasifikasi Kegiatan Keluarga menggunakan Orange dengan Metode Naive Bayes," *Jurnal Teknologi Terpadu*, vol. 9, no. 1, pp. 35–41, 2023, doi: 10.54914/jtt.v9i1.606.
- [17] I. Amelia, A. Mutiara, and I. Santoso, "Analisis Sentimen Opini Publik Terhadap Pengambil Alihan Tmii Oleh Pemerintah Dengan Algoritma Naive Bayes," *Jurnal IKRAITH-INFORMATIKA*, vol. 7, no. 2, pp. 142–148, 2023, [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/issue/archive>
- [18] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 2, no. 2, pp. 91–97, 2021, doi: 10.37859/coscitech.v2i2.2851.
- [19] A. Santosa, I. Purnamasari, and Mayasari Rini, "Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6, pp. 81–93, 2022.
- [20] N. Hadi and D. Sugiarto, "Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM , Logistic Regression dan Naive Bayes," vol. 10, no. 1, pp. 37–49, 2025, doi: 10.30591/jpit.v10i1.7106.
- [21] M. A. Saddam, E. Kurniawan D, and I. Indra, "Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan Support Vector Machine," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, no. 3, pp. 226–233, 2023, doi: 10.30591/jpit.v8i3.4884.
- [22] I. A. N. Gunawan, . S., and I. Shalahuddin, "Dampak Penggunaan Media Sosial Terhadap Gangguan Psikososial Pada Remaja: A Narrative Review," *Jurnal Kesehatan*, vol. 15, no. 1, pp. 78–92, 2022, doi: 10.23917/jk.v15i1.17426.
- [23] S. S. Maulina Putri, M. Arhami, and H. Hendrawaty, "Penerapan Metode SVM pada Klasifikasi Kualitas Air," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 3, no. 2, p. 93, 2023, doi: 10.30811/jaise.v3i2.4630.
- [24] W. D. Yuniarti, A. N. Faiz, and B. Setiawan, "Identifikasi Potensi Keberhasilan Studi Menggunakan Naive Bayes Classifier," *Walisono Journal of Information Technology*, vol. 2, no. 1, p. 1, 2020, doi: 10.21580/wjit.2020.2.1.5204.
- [25] I. Siti Aisah, B. Irawan, and T. Suprapti, "Algoritma Support Vector Machine (Svm) Untuk Analisis Sentimen Ulasan Aplikasi Al Qur'an Digital," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3759–3765, 2024, doi: 10.36040/jati.v7i6.8263.
- [26] N. S. Syam *et al.*, "Model Support Vector Machine untuk Prediksi pada Penggunaan Energi Listrik di Rumah Hemat Energi," *Jurnal Informatika*, vol. 1, no. 2, pp. 56–59, 2022, [Online]. Available: <https://jurnal.uniraya.ac.id/index.php/JI>
- [27] M. F. Naufal, J. Siswanto, and M. G. K. Wicaksono, "Klasifikasi Tulisan Tangan Pada Resep Obat Menggunakan Convolutional Neural Network," *Techno.Com*, vol. 22, no. 2, pp. 508–526, 2023, doi: 10.33633/tc.v22i2.8075.
- [28] K. Citra, P. Daun, and T. Padi, "Klasifikasi Citra Penyakit Daun Tanaman Padi Menggunakan CNN dengan Arsitektur VGG-19," *Jurnal Sains dan Informatika*, vol. 9, no. 1, pp. 37–45, 2023, doi: 10.22216/jsi.v9i1.2175.
- [29] Y. S. HARIYANI, S. HADIYOSO, and T. S. SIADARI, "Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 8, no. 2, p. 443, 2020, doi: 10.26760/elkomika.v8i2.443.