

Penerapan Metode Naïve Bayes dengan PSO pada Analisis Sentimen Pengguna Aplikasi X terhadap Bitcoin

Sentiment Analysis of X Application Users on Bitcoin Using the Naïve Bayes Method Optimized with Particle Swarm Optimization (PSO)

¹Raja Allifin Muhammad, ²Elin Haerani*, ³Fitri Wulandari, ⁴Lola Oktavia

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

^{1,2,3,4}Jl. H.R. Soebrantas KM. 15 No. 155 Tuah Madani Kec. Tuah Madani, Pekanbaru

*e-mail: 12050110520@students.uin-suska.ac.id, elin.haerani@uin-suska.ac.id, fitri.wulandari@uin-suska.ac.id, lola.oktavia@uin-suska.ac.id

(received: 4 June 2025, revised: 11 June 2025, accepted: 12 June 2025)

Abstrak

Kemajuan teknologi dan media sosial telah mendorong perubahan signifikan dalam cara individu mengekspresikan opini, salah satunya terhadap mata uang digital terdesentralisasi yang menggunakan teknologi *blockchain* untuk memungkinkan transaksi tanpa perantara yaitu Bitcoin. Studi ini bertujuan untuk mengevaluasi sentimen pengguna aplikasi X (Twitter) terhadap Bitcoin dengan mengimplementasikan metode *Naïve Bayes* yang dioptimalkan menggunakan *Particle Swarm Optimization* (PSO). Data penelitian ini diperoleh melalui proses *web scraping* pada unggahan pengguna yang mengandung kata kunci “Bitcoin”. Tahapan pra-pemrosesan teks dilakukan untuk meningkatkan kualitas data, dilanjutkan dengan ekstraksi fitur menggunakan pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF) guna mengubah teks menjadi representasi numerik. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil awal menunjukkan bahwa *Naïve Bayes* memberikan performa yang baik dalam mengklasifikasikan sentimen. Dengan integrasi PSO sebagai metode optimasi, meningkatkan performa klasifikasi dari 66.14% menjadi 69.14%. Penelitian ini memberikan kontribusi terhadap pemahaman opini publik terkait Bitcoin dan menunjukkan efektivitas penerapan kombinasi *Naïve Bayes* dan PSO dalam analisis sentimen berbasis teks.

Kata kunci: analisis sentimen, bitcoin, *naïve bayes*, TF-IDF, PSO

Abstract

Advancements in technology and social media have significantly transformed the way individuals express their opinions—one of which is toward decentralized digital currencies that utilize blockchain technology to enable peer-to-peer transactions, such as Bitcoin. This study aims to evaluate user sentiment toward Bitcoin by implementing the Naïve Bayes method optimized with Particle Swarm Optimization (PSO), using data gathered from the X application (formerly Twitter). The data were collected through web scraping of user posts containing the keyword “Bitcoin.” Text preprocessing was performed to enhance data quality, followed by feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) approach to convert textual data into numerical representations. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. Initial results show that the Naïve Bayes classifier performs well in sentiment classification. The integration of PSO as an optimization method improved classification performance from 66.14% to 69.14%. This study contributes to a deeper understanding of public opinion on Bitcoin and demonstrates the effectiveness of combining Naïve Bayes and PSO in text-based sentiment analysis.

Keywords: sentiment analysis, bitcoin, *naïve bayes*, TF-IDF, PSO

1 Pendahuluan

Kemajuan teknologi komunikasi, terutama dengan hadirnya internet, telah mendorong perubahan besar dalam cara masyarakat berinteraksi, berperilaku, dan mengambil keputusan [1], [2]. Salah satu media sosial yang banyak diminati, khususnya di kalangan anak muda, adalah aplikasi Twitter yang sekarang berganti nama menjadi aplikasi X. Platform ini memungkinkan pengguna untuk berbagi pesan, gambar, dan video secara luas ke seluruh dunia [3]. Selain sebagai alat komunikasi, media sosial kini juga berfungsi sebagai sarana promosi dan perdagangan.

Dalam ranah transaksi digital, jual beli masih sangat bergantung pada lembaga keuangan sebagai pihak ketiga untuk memproses pembayaran elektronik, meskipun pendekatan ini memiliki keterbatasan dalam hal kepercayaan [4]. Kemunculan Bitcoin pada tahun 2009 sebagai bentuk awal *cryptocurrency* memperkenalkan sistem transaksi langsung antar pengguna tanpa perantara [5], yang kemudian melahirkan berbagai reaksi, baik positif maupun negatif, dari masyarakat.

Analisis sentimen menjadi metode yang digunakan untuk mengkaji opini publik terkait berbagai isu [6], [7]. Dalam penelitian ini, analisis sentimen digunakan untuk menggali pandangan pengguna aplikasi X terhadap Bitcoin, dengan menerapkan metode *Naïve Bayes* yang dikenal efektif dalam klasifikasi teks [8], [9].

Naïve Bayes merupakan metode klasifikasi yang sederhana namun akurat dalam mengidentifikasi sentimen dengan memperhitungkan frekuensi nilai-nilai dalam data [10], di optimasi menggunakan *Particle Swarm Optimization* (PSO) untuk mencari dan menentukan solusi terbaik dalam menyelesaikan suatu permasalahan optimasi dari sekumpulan partikel [11], sehingga menjadikannya metode optimasi yang layak untuk digunakan dalam penelitian analisis data berbasis teks. Oleh karena itu, penelitian ini memilih untuk menerapkan metode *Naïve Bayes* dengan optimasi dari PSO dalam menganalisis sentimen pengguna aplikasi X terhadap Bitcoin.

Penelitian ini dibatasi hanya pada analisis komentar pengguna aplikasi X tentang Bitcoin dengan pendekatan *Naïve Bayes* dan PSO, tanpa mengulas aspek teknis dari algoritma yang digunakan. Tujuan penelitian ini adalah untuk memperdalam pemahaman mengenai dampak media sosial terhadap penerimaan Bitcoin, memetakan sentimen pengguna, dan menyediakan informasi yang dapat dijadikan acuan bagi akademisi, praktisi, serta pembuat kebijakan di bidang sosial, ekonomi, dan teknologi.

2 Tinjauan Literatur

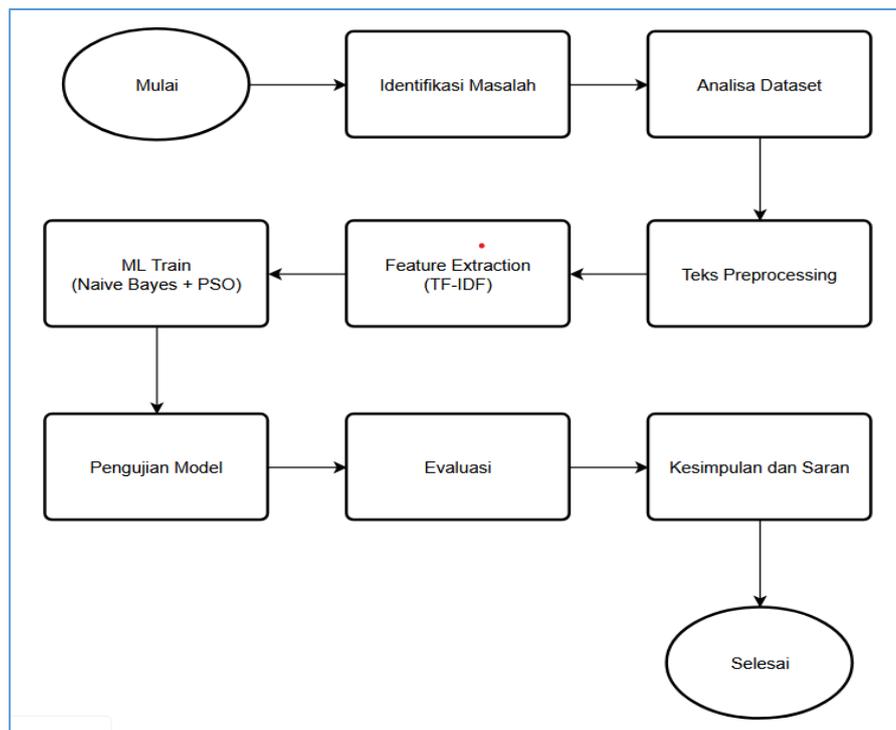
Perkembangan teknologi informasi, khususnya internet, telah mengubah pola komunikasi masyarakat secara signifikan. Media sosial menjadi inovasi utama yang memungkinkan pengguna menyebarkan informasi, berinteraksi, dan membentuk opini publik secara luas [12]. Aplikasi X merupakan salah satu platform yang kerap digunakan untuk menanggapi isu-isu aktual, termasuk topik ekonomi digital seperti Bitcoin. Selain fungsi komunikatif, media sosial juga berperan sebagai sarana promosi dan perdagangan, yang berkontribusi dalam membentuk persepsi publik terhadap aset digital.

Bitcoin, yang diperkenalkan pada tahun 2008, menawarkan sistem keuangan alternatif berbasis *blockchain* yang memungkinkan transaksi tanpa perantara [13]. Inovasi ini menimbulkan beragam reaksi masyarakat, mulai dari dukungan hingga penolakan. Untuk mengkaji sentimen publik terhadap Bitcoin, peneliti menggunakan salah satu metode analisis sentimen yang berbasis teks dengan bantuan teknik pengolahan bahasa alami atau *Natural Language Processing*. Sebuah studi yang dilakukan untuk mendeteksi berita palsu di ranah politik Indonesia, dengan menggunakan metode *Naïve Bayes* yang dioptimasi oleh PSO mencapai tingkat akurasi sebesar 90.25%, lebih tinggi 11.71% dari metode CNN yang dioptimasi oleh PSO dengan akurasi 79.68% [14], penelitian ini menunjukkan bahwa metode *Naïve Bayes* memiliki efisiensi dalam menangani data teks dengan jumlah fitur yang besar dan bersifat *sparse*.

Penggabungan antara *Naïve Bayes* dan PSO diyakini dapat meningkatkan performa analisis sentimen terhadap Bitcoin di media sosial, serta memberikan pemahaman yang lebih mendalam mengenai persepsi publik terhadap aset digital tersebut [15].

3 Metode Penelitian (or Research Method)

Penelitian ini dilakukan dengan tahapan yang ada pada Gambar 1:



Gambar 1. Tahapan penelitian

3.1 Identifikasi Masalah

Penelitian ini mencakup permasalahan tentang data sentimen dari pengguna aplikasi X terhadap Bitcoin, khususnya menggunakan metode *Naïve Bayes* dengan tambahan PSO, yang dapat meningkatkan pemahaman tentang opini dan sikap pengguna aplikasi X terhadap Bitcoin, dengan berfokus pada sentimen yang terkait terhadap Bitcoin tersebut pada platform aplikasi X [16].

3.2 Analisa Dataset

Pemahaman terhadap data merupakan langkah awal yang penting dalam membangun model yang efektif. Oleh karena itu, analisis dilakukan untuk meninjau kualitas, struktur, dan karakteristik utama dari dataset yang digunakan. Rangkaian tahapan analisis dijabarkan menjadi dua tahap sebagai berikut:

a. Pengumpulan Data

Data untuk penelitian ini dikumpulkan dari unggahan pengguna aplikasi X dalam bentuk komentar dan juga tanggapan menggunakan kata kunci “Bitcoin”. Proses pengumpulan data menggunakan metode “*web scraping*”, *web scraping* tersebut merupakan teknik pengambilan data dari halaman *web* secara otomatis dengan menggunakan program atau skrip, sehingga mendapatkan sentimen (komentar dan tanggapan) dari para pengguna aplikasi X dengan format .xlsx sebagai data mentah untuk diteliti [17]. Tabel 1 merupakan data yang didapat setelah melakukan webscraping.

Tabel 1. Data hasil web scraping

full_text	id_str	created_at	username	...	source
Berita Bitcoin: 3 Alasan Kenapa BTC Masih Akan Tetap Turun di Bawah US\$16.000 #bitcoin #kripto https://t.co/HU0DIFiV0a	1.60946E+18	2023-01-01 08:09:13+00:00	cintabitcoin	...	cintabitcoin
@cryptondo Kapan bitcoin turun ke bawah lagi testing ke	1.60974E+18	2023-01-02 02:49:08+00:00	miswar_riki	...	miswar_riki

<http://sistemasi.ftik.unisi.ac.id>

harga 10		0			
@DU09BTC #Bitcoin kalau turun ke support 14k bagus sekali... dan semoga rebound dari support ini	1.60989E+18	2023-01-02 12:25:10+00:00	TraderCoinCrypt	...	TraderCoinCrypt
\$ETH TF 2H!! Untuk kenaikan saat ini buyer sudah ketemu area supplynya kembali di level 1212-1225 dan belum ada konfirmasi bagus untuk memastikan apakah lanjut breakout atau turun lagi. Wns dlu ya #ETH #Ethereum #BTC #Bitcoin #Binance #BNB #blockchain #cryptocurrency #NFTs https://t.co/IPN4VXE9A9	1.60989E+18	2023-01-02 12:32:29+00:00	Tousnesol_	...	Tousnesol_
...
1. #BTC Cut Loss - Minggu ini Bitcoin mengalami volatilitas tinggi yang mengharuskan kita untuk memotong kerugian lebih awal demi menjaga modal.	1.8043E+18	2024-06-22 00:00:02+00:00	essentscars	...	essentscars

b. Pemberian Label pada Data

Data yang awalnya telah didapat melalui *web scraping* akan diberikan label terlebih dahulu, serta menghapus beberapa variable yang tidak diperlukan dalam penelitian ini, sehingga tabel data yang dihasilkan hanya memiliki 2 variable yaitu tweet (sentimen pengguna) dan label (positif dan negatif).

3.3 Pra-pemrosesan Teks

Tahapan awal yang dilakukan dalam proses analisis data berbasis teks, khususnya dalam ranah *Natural Language Processing* (NLP). Tujuan utama dari tahap ini adalah untuk mereduksi kompleksitas data mentah serta menyaring elemen-elemen yang tidak relevan guna meningkatkan kualitas input bagi pemrosesan data yang akan digunakan [18]. Berikut merupakan Langkah teks pre-processing yang digunakan :

a. *Case Folding*

Proses ini mengubah seluruh karakter teks menjadi huruf kecil untuk memastikan keseragaman penulisan.

b. Hapus URL

Membuang setiap tautan URL yang terdapat dalam teks menggunakan pola ekspresi reguler.

c. Hapus Karakter Khusus, Mention, dan Hashtag

Menghilangkan elemen-elemen seperti tanda baca, sebutan akun (@username), tagar (#hashtag), serta tautan web (http://...) dari teks.

d. Hapus Angka

Menghilangkan seluruh angka yang ada dalam teks menggunakan pola ekspresi reguler.

e. *Tokenization*

Memecah teks menjadi kata-kata (token) dengan memanfaatkan fungsi *word tokenize* dari library *NLTK*.

f. *Remove Stopwords*

Menghilangkan kata-kata yang sering muncul dan tidak memberikan kontribusi informasi, seperti "dan", "di", "untuk", dan lainnya.

3.4 Feature Extraction (TF-IDF)

TF-IDF (*Term Frequency - Inverse Document Frequency*) adalah teknik untuk mengubah teks menjadi fitur numerik yang dapat digunakan oleh model pembelajaran mesin, termasuk dalam analisis sentimen dengan *Naïve Bayes*. Metode ini menilai relevansi suatu kata dalam dokumen dengan mempertimbangkan frekuensi kemunculannya dalam dokumen dan seberapa jarang kata tersebut ditemukan di seluruh kumpulan dokumen [19].

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

$$TF(t, d) = \frac{f_{t,d}}{\sum_{k=1}^n f_{k,d}} \quad (2)$$

$$IDF(t) = \log\left(\frac{N}{1 + n_t}\right) \quad (3)$$

Keterangan;

- t = term (kata) yang dihitung frekuensinya
- d = dokumen tertentu
- $f_{t,d}$ = frekuensi kemunculan term t dalam dokumen d
- V = jumlah total term unik dalam dokumen d atau dalam korpus (vocabulary)
- $\sum_{k=1}^V f_{k,d}$ = total semua kemunculan kata dalam dokumen d (jumlah kata dalam dokumen)

Rumus (1) (2) dan (3) ini berfungsi untuk menentukan nilai bobot suatu kata dalam dokumen, dengan mempertimbangkan frekuensi kemunculannya serta tingkat keumumannya di seluruh kumpulan dokumen.

3.5 Machine Learning Train (*Naïve Bayes dan Particle Swarm Optimization*)

- a. *Naïve Bayes* adalah model klasifikasi probabilistik yang menganggap fitur-fitur independen satu sama lain ketika kelas diketahui [20].
- b. *Multinomial Naïve Bayes* secara luas dimanfaatkan dalam bidang pemrosesan bahasa alami, khususnya dalam menyelesaikan permasalahan klasifikasi teks seperti analisis sentimen, deteksi pesan spam, serta pengelompokan dokumen. Algoritma ini menunjukkan kinerja yang baik pada data teks yang direpresentasikan dalam bentuk frekuensi kemunculan kata, seperti pada model representasi *bag-of-words* maupun *term frequency-inverse document frequency* (TF-IDF).

$$\hat{y} = \arg \max_{C_k} \left[\log P(C_k) + \sum_{i=1}^n x_i \log P(x_i | C_k) \right] \quad (4)$$

Keterangan:

- \hat{y} = Hasil prediksi kelas
- $\arg \max_{C_k}$ = Argumen maksimum
- $P(C_k)$ = Probabilitas apriori dari kelas C_k
- x_i = Fitur ke- i dari data input
- $P(x_i | C_k)$ = Probabilitas fitur x_i muncul **dengan syarat** bahwa kelasnya adalah C_k
- $\sum_{i=1}^n$ = Penjumlahan terhadap semua fitur (dari 1 sampai n)

Rumus (4) ini digunakan untuk mengidentifikasi kelas atau kategori yang paling sesuai untuk suatu dokumen dengan mempertimbangkan nilai probabilitas dalam kerangka kerja *Multinomial Naïve Bayes*.

- c. *Particle Swarm Optimization* (PSO) merupakan metode optimasi populasi yang terinspirasi dari perilaku kolektif organisme sosial, seperti burung atau ikan. Setiap solusi kandidat disebut partikel dan bergerak dalam ruang pencarian dengan mempertimbangkan pengalaman terbaik individu maupun kelompok, untuk memperoleh solusi optimal terhadap suatu fungsi objektif [21]. Dalam penerapan ini, algoritma PSO dimanfaatkan untuk mengoptimalkan nilai parameter α terbaik di antara 0.01 hingga 1.0 pada model *Multinomial Naïve Bayes*. Proses optimasi dilakukan dengan melibatkan 30 partikel yang secara simultan menelusuri ruang solusi selama 50 iterasi. Setiap solusi yang diusulkan dievaluasi menggunakan fungsi objektif yang mengacu pada nilai *F1-score*, sehingga parameter α terbaik diperoleh berdasarkan kinerja klasifikasi tertinggi pada data validasi.

3.6 Evaluasi

Setelah proses pelatihan model *Naïve Bayes* yang dikombinasikan dengan optimasi melalui algoritma PSO diselesaikan, tahap evaluasi dilakukan untuk mengukur performa model dan menguji kemampuannya dalam menangani data yang belum pernah diproses sebelumnya. Evaluasi ini menggunakan sejumlah indikator, seperti akurasi, *precision*, *recall*, dan *F1-score* guna memberikan gambaran menyeluruh terhadap efektivitas model yang dikembangkan, berikut merupakan matrix evaluasi yang digunakan dalam penelitian ini antara lain:

- a. Akurasi: Mengukur sejauh mana model dapat mengklasifikasikan data dengan benar, dihitung dengan membagi jumlah prediksi yang benar dengan total prediksi yang dilakukan (persamaan 5).

$$\text{Akurasi} = \frac{\text{True Positif} + \text{True Negatif}}{\text{Seluruh Data}} \quad (5)$$

Keterangan :

True Positif = Jumlah data yang positif dan diprediksi dengan benar sebagai positif.

True Negatif = Jumlah data yang negatif dan diprediksi dengan benar sebagai negatif.

Rumus ini digunakan untuk menggambarkan frekuensi model dalam menghasilkan prediksi yang tepat, mencakup baik prediksi terhadap kelas positif maupun negatif.

- b. *Precision*: *Precision* merupakan ukuran yang menunjukkan proporsi prediksi positif yang benar-benar tepat dibandingkan dengan seluruh prediksi positif yang dibuat oleh model (persamaan 6).

$$\text{Precision} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Positif}} \quad (6)$$

Keterangan :

True Positif = Jumlah data yang positif dan diprediksi dengan benar sebagai positif.

False Positif = Jumlah data yang salah diprediksi sebagai positif padahal sebenarnya negatif.

Rumus ini berfungsi untuk menilai tingkat ketepatan model dalam memprediksi kelas positif, dengan cara membandingkan jumlah prediksi positif yang benar dengan total seluruh prediksi yang dikategorikan sebagai positif oleh model.

- c. *Recall*: *Matrix* ini digunakan dalam evaluasi analisis sentimen untuk mengukur kemampuan model dalam mendeteksi seluruh data yang seharusnya termasuk ke dalam suatu kelas tertentu (persamaan 7).

$$\text{Recall} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Negatif}} \quad (7)$$

Keterangan :

True Positif = Jumlah data yang positif dan diprediksi dengan benar sebagai positif.

False Negatif = Merupakan jumlah data yang secara keliru diklasifikasikan sebagai negatif, padahal sebenarnya data tersebut termasuk dalam kelas positif.

Rumus ini digunakan untuk menilai sejauh mana model mampu mengenali semua *instance* yang benar-benar termasuk dalam kelas positif, dengan cara membandingkan jumlah prediksi positif yang tepat terhadap keseluruhan data yang secara aktual tergolong positif.

d. *F1-Score*: Ukuran yang memadukan *precision* dan *recall* guna mengevaluasi keseimbangan antara akurasi dan cakupan prediksi model (persamaan 8).

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

Rumus ini digunakan untuk menentukan nilai *F1-Score*, yaitu sebuah metrik evaluasi yang menggabungkan *precision* dan *recall* menjadi satu ukuran terpadu guna menunjukkan keseimbangan antara keduanya. *F1-Score* sangat bermanfaat ketika model dihadapkan pada data yang tidak seimbang, karena memberikan penilaian yang adil terhadap kinerja model dalam mengenali data positif dengan benar (*recall*) sekaligus memastikan ketepatan prediksi positif yang dihasilkan (*precision*).

4 Hasil dan Pembahasan

4.1 Dataset

Dataset yang digunakan dalam penelitian ini berjumlah 1750 komentar dari pengguna aplikasi X tentang bitcoin, yang didapat melalui *web scrapping* pada aplikasi X, dilanjutkan kepada tahap pemberian label pada data tersebut, agar bisa digunakan pada penelitian ini. Data yang telah diberi label tersebut divalidasi pemberian label-nya oleh 3 anator yang merupakan guru Bahasa Indonesia agar sesuai dengan KBBI. Tabel 2 merupakan tabel data yang telah diberi label:

Tabel 2. Data setelah diberi label

tweet	label
Berita Bitcoin: 3 Alasan Kenapa BTC Masih Akan Tetap Turun di Bawah US\$16.000 #bitcoin #kripto https://t.co/HU0DIFiV0a	negatif
@cryptondo Kapan bitcoin turun ke bawah lagi testing ke harga 10	negatif
@DU09BTC #Bitcoin kalau turun ke support 14k bagus sekali... dan semoga rebound dari support ini	positif
\$ETH TF 2H!! □ Untuk kenaikan saat ini buyer sudah ketemu area supplynya kembali di level 1212-1225 dan belum ada konfirmasi bagus untuk memastikan apakah lanjut breakout atau turun lagi. Wns dlu ya #ETH #Ethereum #BTC #Bitcoin #Binance #BNB #blockchain #cryptocurrency #NFTs https://t.co/IPN4VXE9A9	positif
...	...
1. #BTC Cut Loss - Minggu ini Bitcoin mengalami volatilitas tinggi yang mengharuskan kita untuk memotong kerugian lebih awal demi menjaga modal.	negatif

4.2 Pra-pemrosesan Teks

Dalam tahap ini, dilakukan beberapa teknik pra-pemrosesan, antara lain pembersihan teks (*cleaning*), penyeragaman huruf menjadi format huruf kecil (*case folding*), pemisahan kata berdasarkan satuan terkecil (*tokenization*), serta penghapusan kata-kata umum yang tidak memiliki nilai informasi tinggi (*stopword removal*). Seperti yang dapat dilihat pada Tabel 3.

Tabel 3. Data setelah pra-pemrosesan teks

Teks Sebelum Pra-pemrosesan	Teks Setelah Pra-pemrosesan
Berita Bitcoin: 3 Alasan Kenapa BTC Masih Akan Tetap Turun di Bawah US\$16.000 #bitcoin #kripto https://t.co/HU0DIFiV0a	berita bitcoin alasan btc turun us bitcoin kripto
@cryptondo Kapan bitcoin turun ke bawah lagi testing ke harga 10	bitcoin turun testing harga
@DU09BTC #Bitcoin kalau turun ke support 14k bagus sekali... dan semoga rebound dari support ini	bitcoin turun support k bagus semoga rebound support
\$ETH TF 2H!! <input type="checkbox"/> Untuk kenaikan saat ini buyer sudah ketemu area supplynya kembali di level 1212-1225 dan belum ada konfirmasi bagus untuk memastikan apakah lanjut breakout atau turun lagi. Wns dlu ya #ETH #Ethereum #BTC #Bitcoin #Binance #BNB #blockchain #cryptocurrency #NFTs https://t.co/IPN4VXE9A9	eth tf h kenaikan buyer ketemu area supplynya level konfirmasi bagus breakout turun wns dlu ya eth ethereum btc bitcoin binance bnb blockchain cryptocurrency nfts
...	...
1. #BTC Cut Loss - Minggu ini Bitcoin mengalami volatilitas tinggi yang mengharuskan kita untuk memotong kerugian lebih awal demi menjaga modal.	btc cut loss minggu bitcoin mengalami volatilitas mengharuskan memotong kerugian menjaga modal

4.3 TF-IDF

Pada tahap ini, teks diubah menjadi representasi numerik berupa bobot kata yang mencerminkan relevansinya dalam dokumen. Bobot-bobot ini kemudian digunakan sebagai fitur input untuk model. Hasil vektorisasi TF-IDF ditampilkan pada Gambar 2:

kata	tfidf
altcoin	0.509698
altseason	0.490913
mendominasi	0.323289
alternatif	0.306429
musim	0.306429
serentak	0.306429
season	0.260747
coin	0.208000
bitcoin	0.041796

Gambar 2. Hasil vektorisasi TF-IDF

TF-IDF digunakan untuk mengidentifikasi kata-kata yang memiliki pengaruh besar dalam kumpulan data pelatihan. Pendekatan ini berguna untuk menemukan istilah-istilah yang tidak hanya sering muncul, tetapi juga memiliki bobot penting karena kemunculannya tidak merata di seluruh dokumen. Kata-kata dengan skor TF-IDF yang tinggi umumnya menggambarkan tema atau isu utama dalam data. Tabel 4 merupakan daftar sembilan kata teratas berdasarkan total skor TF-IDF dalam data pelatihan yang dianalisis:

Tabel 4. Kata teratas dari total skor TF-IDF disemua dokumen train

Kata	Total TF-IDF
bitcoin	109.87
turun	82.28
harga	61.53
btc	44.11
kripto	36.93
us	30.14
yg	28.35
pasar	26.35
beli	22.54

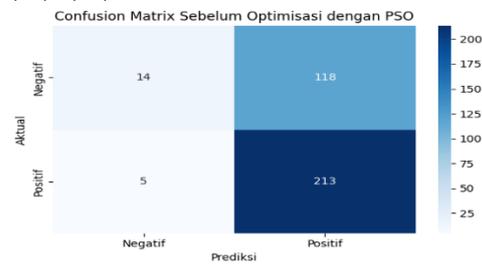
4.4 NAÏVE BAYES – PSO

a. Naïve Bayes

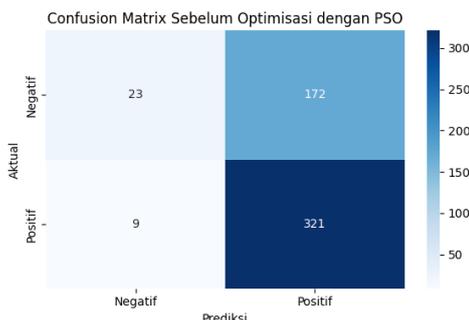
Pada tahap ini peneliti mengevaluasi kinerja awal model dengan *Naïve Bayes*, dilakukan pengujian menggunakan sejumlah metrik evaluasi seperti *precision*, *recall*, *f1-score* dan *support*, dengan beberapa perbandingan (90:10, 80:20, 70:30, 60:40). Evaluasi ini bertujuan untuk memberikan gambaran awal terkait tingkat akurasi serta ketimpangan dalam kemampuan model mendeteksi kelas positif maupun negatif. Adapun hasil prediksi dari *Confusion Matrix* pada 4 perbandingan tersebut dapat dilihat pada Gambar 3, 4, 5, 6;



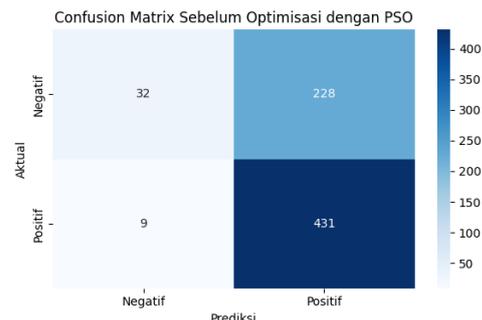
Gambar 3. Confusion matrix NB (90:10)



Gambar 4. Confusion matrix NB (80:20)



Gambar 5. Confusion matrix NB (70:30)



Gambar 6. Confusion matrix NB (60:40)

Adapun hasil evaluasi *naïve bayes* sebelum proses optimisasi dapat dilihat pada Tabel 5:

Tabel 5. Hasil evaluasi naïve bayes

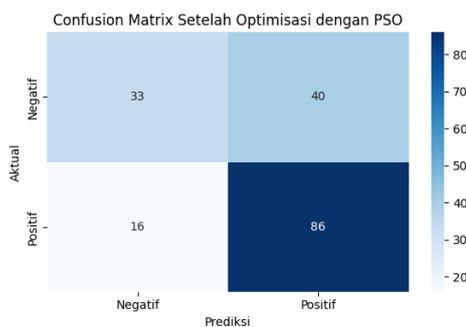
Perbandingan		Precision	Recall	F1-score	support
90 : 10	negatif	0.92	0.15	0.26	73
	positif	0.62	0.99	0.76	102
	accuracy			0.64	175
	macro avg	0.77	0.57	0.51	175
	weighted avg	0.74	0.64	0.55	175
80 : 20	negatif	0.74	0.11	0.19	132
	positif	0.64	0.98	0.78	218
	accuracy			0.65	350

	macro avg	0.69	0.54	0.48	350
	weighted avg	0.68	0.65	0.55	350
70 : 30	negatif	0.72	0.12	0.20	195
	positif	0.65	0.97	0.78	330
	accuracy			0.66	525
	macro avg	0.68	0.55	0.49	525
	weighted avg	0.68	0.66	0.57	525
60 : 40	negatif	0.78	0.12	0.21	260
	positif	0.65	0.98	0.78	440
	accuracy			0.66	700
	macro avg	0.72	0.55	0.50	700
	weighted avg	0.70	0.66	0.57	700

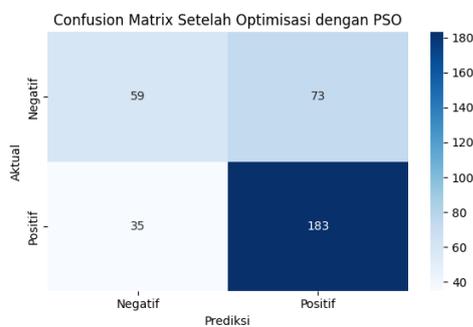
Hasil evaluasi model *Naïve Bayes* pada berbagai rasio pembagian data *training* dan *validation* menunjukkan variasi kinerja yang cukup konsisten. Akurasi model berada dalam rentang 64.00% hingga 66.14%, dengan akurasi tertinggi pada rasio 60:40 sebesar 66.14% diperoleh pada rasio 60:40, sementara akurasi terendah 64.00% tercatat pada rasio 90:10. Dari segi performa kelas, *precision* untuk kelas negatif cenderung tinggi, terutama pada rasio 90:10 mencapai 0.92, namun *recall* untuk kelas ini sangat rendah, hanya berkisar antara 0.11 hingga 0.15 di seluruh rasio, hal ini terjadi karena adanya ketidakseimbangan jumlah data antara kelas positif dan negatif, serta kurang efektifnya fitur-fitur dalam menggambarkan ciri khas dari komentar yang bersentimen negatif. sehingga *F1-score* kelas negatif juga rendah, yakni antara 0.19 hingga 0.26. Sebaliknya, untuk kelas positif, *recall* sangat tinggi, hampir selalu di atas 0.97, dengan *precision* sekitar 0.62 hingga 0.65, menghasilkan *F1-score* yang lebih baik, yaitu sekitar 0.76 hingga 0.78 di semua rasio. Hal ini menunjukkan bahwa model lebih efektif dalam mengidentifikasi sentimen positif dibandingkan sentimen negatif. Secara keseluruhan, nilai *macro* dan *weighted average F1-score* berada pada kisaran 0.48 hingga 0.57.

b. *Naïve Bayes* dengan PSO

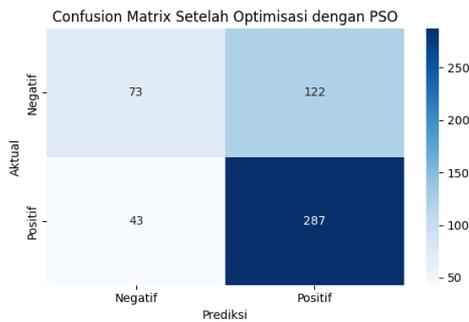
Pada tahap selanjutnya, peneliti mengevaluasi kinerja *naïve bayes* setelah menggunakan optimasi dari PSO untuk mengatasi permasalahan ketidakseimbangan jumlah data antara kelas positif dan negatif dengan cara mencari kombinasi parameter model *Naïve Bayes* yang optimal, sehingga meningkatkan sensitivitas model terhadap pola-pola dari kelas negatif, menggunakan perbandingan 90:10, 80:20, 70:30, 60:40, seperti sebelum menerapkan PSO. Pada evaluasi kali ini merupakan gambaran akhir untuk mendapatkan hasil dari performa *Naïve Bayes* yang dioptimasi PSO. Berikut ini merupakan hasil prediksi *Confusion Matrix* menggunakan *Naïve Bayes* dengan PSO yang terlihat pada Gambar 7, 8, 9, 10:



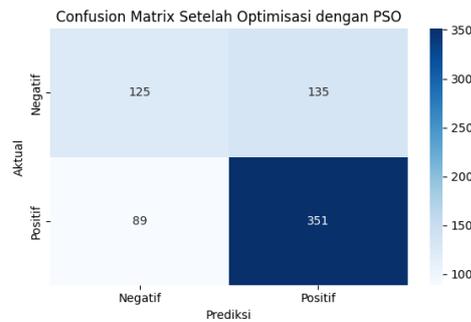
Gambar 7. *Confusion matrix* NB dengan PSO (90:10)



Gambar 8. *Confusion matrix* NB dengan PSO (80:20)



Gambar 9. Confusion matrix NB dengan PSO (70:30)



Gambar 10. Confusion matrix NB dengan PSO (60:40)

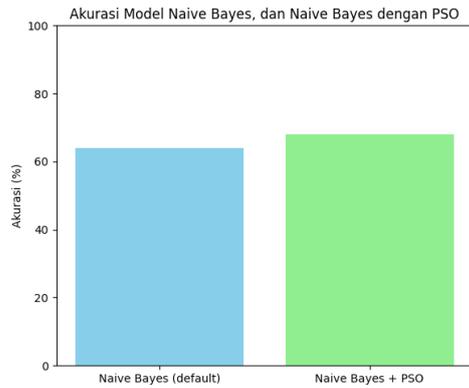
Adapun hasil evaluasi *Naïve Bayes* dengan menggunakan proses optimisasi PSO dapat dilihat pada Tabel 6:

Tabel 6. Hasil evaluasi *naïve bayes* dengan PSO

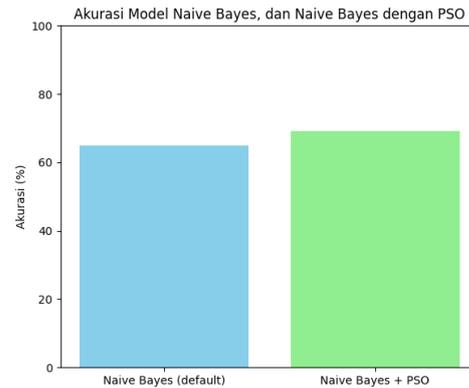
Perbandingan		Precision	Recall	F1-score	support
90 : 10	negatif	0.67	0.45	0.54	73
	positif	0.68	0.84	0.75	102
	accuracy			0.68	175
	macro avg	0.68	0.65	0.65	175
	weighted avg	0.68	0.68	0.67	175
80 : 20	negatif	0.63	0.45	0.52	132
	positif	0.71	0.84	0.77	218
	accuracy			0.69	350
	macro avg	0.67	0.64	0.65	350
	weighted avg	0.68	0.69	0.68	350
70 : 30	negatif	0.63	0.37	0.47	195
	positif	0.70	0.87	0.78	330
	accuracy			0.69	525
	macro avg	0.67	0.62	0.62	525
	weighted avg	0.67	0.69	0.66	525
60 : 40	negatif	0.58	0.48	0.53	260
	positif	0.72	0.80	0.76	440
	accuracy			0.68	700
	macro avg	0.65	0.64	0.64	700
	weighted avg	0.67	0.68	0.670	700

Hasil evaluasi model *Naïve Bayes* dengan optimasi PSO menggunakan berbagai rasio pembagian data *training* dan *validation* menunjukkan kinerja yang cukup stabil. Akurasi tertinggi dicapai pada rasio 80:20, yaitu sebesar 69.14%, sedangkan akurasi terendah sebesar 68.00% terdapat pada rasio 90:10 dan 60:40. Dari sisi performa kelas, *precision* untuk kelas negatif berada di kisaran 0.58 hingga 0.67, dengan *recall* berkisar antara 0.37 hingga 0.48, menghasilkan *F1-score* antara 0.47 hingga 0.54. Sementara itu, kelas positif memiliki *precision* yang sedikit lebih tinggi, antara 0.68 hingga 0.72, dengan *recall* yang juga tinggi, sekitar 0.80 hingga 0.87, sehingga *F1-score* kelas positif lebih baik, yakni antara 0.75 hingga 0.78. Secara keseluruhan, nilai *macro average F1-score* berkisar antara 0.62 hingga 0.65, sedangkan *weighted average F1-score* berada pada rentang 0.66 hingga 0.67, menunjukkan bahwa model memberikan performa yang relatif seimbang antara kelas positif dan negatif, meskipun terdapat sedikit penurunan pada *recall* kelas negatif pada rasio 70:30.

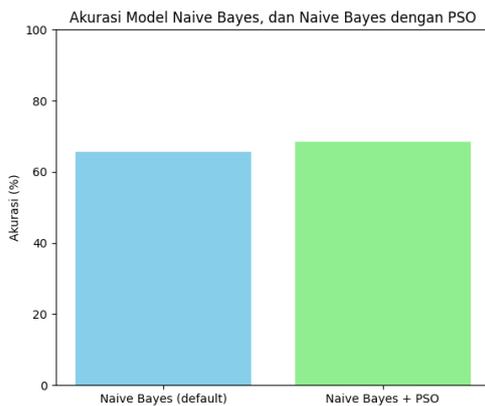
Visualisasi berbentuk bar chart digunakan untuk memperjelas perbandingan performa model yang menampilkan akurasi dari model *Naïve Bayes* sebelum dan sesudah proses optimasi menggunakan PSO dengan empat perbandingan rasio yang ditampilkan pada Gambar 11, 12, 13, 14.



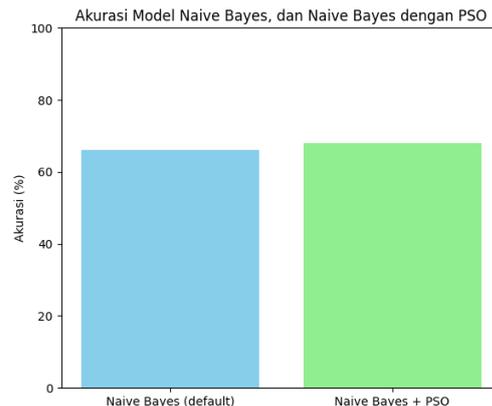
Gambar 11. Visualisasi bar chart NB sebelum dan sesudah optimasi (90:10)



Gambar 12. Visualisasi bar chart NB sebelum dan sesudah optimasi (80:20)



Gambar 13. Visualisasi bar chart NB sebelum dan sesudah optimasi (70:30)



Gambar 14. Visualisasi bar chart NB sebelum dan sesudah optimasi (60:40)

Keempat diagram *bar* tersebut menampilkan perbandingan akurasi antara model *Naive Bayes* dalam kondisi *default* (tanpa proses optimisasi) dan model yang telah disempurnakan menggunakan metode PSO. Warna biru pada *bar* mewakili akurasi sebelum optimisasi, sedangkan warna hijau menunjukkan hasil akurasi setelah proses PSO diterapkan. Berdasarkan keempat grafik ini, terlihat adanya peningkatan akurasi setelah optimisasi, yang ditunjukkan oleh bertambahnya tinggi bar pada sumbu vertikal. Temuan ini menunjukkan bahwa penggunaan PSO berkontribusi secara positif terhadap peningkatan performa model.

Kategori Akurasi Secara Umum

Tabel 7. Rentang kategori akurasi

Rentang Akurasi	Kategori
< 60%	Rendah
60–75%	Sedang
> 75%	Tinggi

Tabel 7 menampilkan hasil evaluasi model *Naive Bayes* yang dioptimasi menggunakan PSO menunjukkan akurasi tertinggi pada rasio 80:20 sebesar 69,14%. Berdasarkan rentang akurasi yang digunakan sebagai acuan, yaitu kurang dari 60% dikategorikan rendah, 60–75% sedang, dan di atas 75% tinggi, maka performa model ini termasuk dalam kategori sedang. Hal ini menunjukkan bahwa meskipun model telah mencapai tingkat ketepatan yang cukup baik dalam klasifikasi data, masih terdapat ruang untuk peningkatan agar dapat mencapai hasil yang lebih optimal dan akurat.

5 Kesimpulan

Penelitian ini menganalisis sentimen pengguna aplikasi X terhadap Bitcoin menggunakan algoritma *Naïve Bayes* yang ditingkatkan dengan PSO. Melalui tahapan komprehensif mulai dari pengumpulan data via *web scraping*, pra-pemrosesan teks, ekstraksi fitur dengan TF-IDF, hingga pelatihan model klasifikasi, diperoleh peningkatan performa signifikan setelah penerapan PSO. Sebelum optimasi, model cenderung lemah dalam mengenali sentimen negatif, terutama ditunjukkan oleh rendahnya nilai *recall*. Namun, setelah dioptimasi, performa meningkat pada semua metrik evaluasi, termasuk *recall*, *F1-score*, dan akurasi, dengan kenaikan akurasi rata-rata dari 64.00–66.14% menjadi 68.00–69.14%. Peningkatan ini menunjukkan bahwa metode optimasi PSO efektif dalam meningkatkan kualitas klasifikasi teks, khususnya pada domain analisis sentimen. Selain itu, hasil penelitian mengungkapkan bahwa sentimen positif terhadap Bitcoin lebih dominan di media sosial aplikasi X, sebagaimana tercermin dari distribusi data dan hasil klasifikasi yang cenderung lebih akurat pada kelas positif. Penelitian ini memiliki potensi untuk dikembangkan lebih lanjut melalui penerapan model *deep learning* seperti LSTM maupun BERT, yang mampu merepresentasikan konteks semantik secara lebih mendalam dan kompleks. Selain itu, peningkatan cakupan dataset serta eksplorasi terhadap pendekatan klasifikasi multi-kelas dapat memperluas pemahaman mengenai keragaman sentimen publik terhadap aset kripto dalam berbagai kondisi dan dinamika pasar.

Referensi

- [1] Z. Martha, “Penggunaan Fitur Media Sosial *Instagram Stories* sebagai Media Komunikasi,” *Jurnal Komunikasi Nusantara*, 2021.
- [2] R. Armanto, and Muji Gunarto, “Analisis Dampak mEdia Sosial terhadap Penjualan Perumahan: Studi Empiris penggunaan Iklan *Facebook* dan *Instagram*,” *Jurnal Bisnis, Manajemen dan Ekonomi*, 2022.
- [3] R. W. Agustin and M. Ayu, “*The Impact of using Instagram for Increasing Vocabulary and Listening Skill*,” *Journal of English Language Teaching and Learning*, 2021.
- [4] S. Nakamoto, “Bitcoin: sebuah Sistem Uang Tunai Elektronik *Peer-to-Peer*,” terjemahan dari *Bitcoin: A Peer-to-Peer Electronic Cash System*.
- [5] M. A. Fauzi, N. Paiman, and Z. Othman, “*Bitcoin and Cryptocurrency: Challenges, Opportunities and Future Works*,” *Journal of Asian Finance, Economics and Business*, Vol. 7, No. 8, pp. 695–704, Aug. 2020, doi: 10.13106/JAFEB.2020.Vol.7.No.8.695.
- [6] F. F. Rachman and S. Pramana, “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial *Twitter*,” *Indonesian of Health Information Management Journal*, 2020.
- [7] A. Dwiki, A. Putra, and S. Juanita, “Analisis Sentimen pada Ulasan Pengguna Aplikasi Bibit dan Bareksa dengan Algoritma KNN,” *Jurnal Teknik Informatika dan Sistem Informasi*, Vol. 8, No. 2, 2021.
- [8] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis sentimen Aplikasi Ruang Guru di *Twitter* menggunakan Algoritma Klasifikasi,” *Jurnal Teknoinfo*, Vol. 14, No. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [9] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, “*Sentiment Analysis of COVID-19 Vaccine in Indonesia using Naïve Bayes algorithm*,” *IOP Conf Ser Mater Sci Eng*, Vol. 1088, No. 1, p. 012045, Feb. 2021, doi: 10.1088/1757-899x/1088/1/012045.
- [10] M. R. Handoko, “Sistem Pakar Diagnosa Penyakit selama Kehamilan menggunakan Metode *Naive Bayes* berbasis Web,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, Vol. 2, No. 1, pp. 50–58, 2021.
- [11] T. M. Shami, A. A. El-Saleh, M. Alswaiti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, “*Particle Swarm Optimization: A Comprehensive Survey*,” *IEEE Access*, Vol. 10, pp. 10031–10061, 2022, doi: 10.1109/ACCESS.2022.3142859.
- [12] F. D. Ananda and Y. Pristyanto, “Analisis Sentimen Pengguna *Twitter* terhadap Layanan Internet Provider menggunakan Algoritma *Support Vector Machine*,” *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, Vol. 20, No. 2, pp. 407–416, May 2021, doi: 10.30812/matrik.v20i2.1130.

- [13] H. Jahanshahloo, F. Irresberger, and A. Urquhart, “Bitcoin under the Microscope,” *British Accounting Review*, 2023, doi: 10.1016/j.bar.2023.101237.
- [14] Y. Kurnia, E. D. Kusuma, L. W. Kusuma, Suwitno, and W. Apridius, “Perbandingan *Naïve Bayes* dan CNN yang dioptimasi PSO pada Identifikasi Berita Hoax Politik Indonesia,” *bit-Tech*, Vol. 6, No. 3, pp. 340–352, Apr. 2024, doi: 10.32877/bt.v6i3.1225.
- [15] H. B. Jatmiko, N. T. Kurniadi, and D. Maulana, “Optimasi *Naïve Bayes* dengan *Particle Swarm Optimization* untuk Analisis Sentimen Formula E-Jakarta,” *Journal Automation Computer Information System*, 2022.
- [16] A. S. Talita, O. S. Nataza, and Z. Rustam, “*Naïve Bayes classifier and Particle Swarm Optimization feature selection method for classifying intrusion detection system dataset*,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1752/1/012021.
- [17] M. Dogucu and M. Çetinkaya-Rundel, “Web Scraping in the Statistics and Data Science Curriculum: Challenges and Opportunities,” *Journal of Statistics and Data Science Education*, Vol. 29, No. S1, pp. S112–S122, 2021, doi: 10.1080/10691898.2020.1787116.
- [18] B. Hakim, “Analisa Sentimen Data *Text Preprocessing* pada Data Mining dengan menggunakan *Machine Learning*,” *JBASE - Journal of Business and Audit Information Systems*, Vol. 4, No. 2, Aug. 2021, doi: 10.30813/jbase.v4i2.3000.
- [19] R. A. Supono and M. A. Suprayogi, “Perbandingan Metode TF-ABS dan TF-IDF pada Klasifikasi Teks *Helpdesk* menggunakan *K-Nearest Neighbor*,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, Vol. 5, No. 5, pp. 911–918, Oct. 2021, doi: 10.29207/resti.v5i5.3403.
- [20] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara menggunakan Algoritma *Naïve Bayes* dan KNN,” *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [21] M. Zhao, H. Zhao, and M. Zhao, “*Particle Swarm Optimization Algorithm with Adaptive Two-Population Strategy*,” *IEEE Access*, Vol. 11, pp. 62242–62260, 2023, doi: 10.1109/ACCESS.2023.3287859.