

Perbandingan Kinerja Algoritma Random Forest, KNN, dan SVM dalam Analisis Sentimen Cryptocurrency

Rinaldi Sukma Andarujaya*, Ryan Randy Suryono

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}rinaldi_sukma_andarujaya@teknokrat.ac.id, ²ryan@teknokrat.ac.id

Email Penulis Korespondensi: rinaldi_sukma_andarujaya@teknokrat.ac.id

Submitted: 30/12/2024; Accepted: 26/02/2025; Published: 01/03/2025

Abstrak—Cryptocurrency merupakan uang digital berbasis teknologi *blockchain* yang menawarkan keamanan dan transparansi dalam transaksi, sehingga semakin menarik perhatian masyarakat, termasuk di Indonesia. Dengan jumlah investor yang telah melampaui 20 juta orang, *cryptocurrency* memunculkan beragam opini di media sosial. Beberapa pihak melihatnya sebagai peluang investasi modern yang menjanjikan, sementara yang lain menyoroti risiko fluktuasi harga, keamanan, dan regulasi yang kurang jelas. Untuk memahami sentimen masyarakat terhadap *cryptocurrency*, analisis sentimen berbasis pembelajaran mesin menjadi solusi yang relevan. Penelitian ini membandingkan kinerja tiga algoritma populer, yaitu *Random Forest*, *K-Nearest Neighbors (KNN)*, dan *Support Vector Machine (SVM)*, dalam analisis sentimen opini masyarakat. Ketiga algoritma ini memiliki keunggulan dan kelemahan yang berbeda, tergantung pada karakteristik data dan tujuan analisis. *Random Forest* dikenal stabil tetapi memerlukan komputasi tinggi, *KNN* mudah diterapkan namun kurang andal pada data berdimensi tinggi, dan *SVM* unggul dalam memisahkan data kompleks tetapi membutuhkan tuning parameter yang cermat. Penelitian sebelumnya menunjukkan perbedaan hasil akurasi ketiga algoritma ini pada berbagai dataset, sehingga perlu dilakukan evaluasi lebih lanjut untuk menentukan algoritma yang paling efektif. Hasil penelitian ini diharapkan memberikan panduan dalam memilih algoritma yang tepat untuk analisis sentimen, khususnya pada data opini terkait *cryptocurrency*, serta memperluas pemahaman tentang penerapan algoritma pada data yang dinamis dan kompleks.

Kata Kunci: Random Forest; KNN; SVM; Cryptocurrency; Analisis Sentimen

Abstract—Cryptocurrency is a digital money based on blockchain technology that offers security and transparency in transactions, so it has increasingly attracted the attention of the public, including in Indonesia. With the number of investors surpassing 20 million, cryptocurrencies have generated a variety of opinions on social media. Some see it as a promising modern investment opportunity, while others highlight the risks of price fluctuations, security, and unclear regulations. To understand public sentiment towards cryptocurrencies, machine learning-based sentiment analysis is a relevant solution. This research compares the performance of three popular algorithms, namely Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), in sentiment analysis of public opinion. These three algorithms have different advantages and disadvantages, depending on the characteristics of the data and the purpose of the analysis. Random Forest is known to be stable but requires high computation, KNN is easy to apply but less reliable on high-dimensional data, and SVM excels at separating complex data but requires careful parameter tuning. Previous research has shown differences in the accuracy of these three algorithms on various datasets, so further evaluation is needed to determine the most effective algorithm. The results of this study are expected to provide guidance in choosing the right algorithm for sentiment analysis, especially on cryptocurrency-related opinion data, as well as expand the understanding of the application of algorithms on dynamic and complex data.

Keywords: Random Forest; KNN; SVM; Cryptocurrency; Sentiment Analysis

1. PENDAHULUAN

Cryptocurrency adalah uang digital yang dibuat dengan teknologi *blockchain* guna menjamin keamanan serta transparansi dalam setiap transaksi.[1]. Teknologi *blockchain*, sebagai dasar dari *cryptocurrency*, memungkinkan transaksi dilakukan secara *online* tanpa memerlukan perantara pihak ketiga, seperti bank atau institusi keuangan lainnya[2]. Keamanan pada *cryptocurrency* dijaga melalui berbagai algoritma, enkripsi, dan teknik kriptografi yang rumit berbasis teknologi *blockchain*, menjadikannya lebih sulit untuk diretas atau dipalsukan [3].

Dalam beberapa tahun terakhir, *cryptocurrency* mengalami lonjakan popularitas yang sangat signifikan. Tidak hanya di negara-negara maju, namun juga di negara-negara berkembang seperti Indonesia. Berdasarkan data dari Kementerian Perdagangan, jumlah investor *cryptocurrency* di Indonesia telah melampaui angka 20 juta orang[4]. Tingginya angka ini menggambarkan besarnya minat masyarakat Indonesia terhadap aset digital. *Cryptocurrency* dianggap sebagai alternatif investasi modern yang menarik, terutama bagi generasi muda yang menginginkan investasi dengan potensi keuntungan tinggi. Ketersediaan berbagai platform *trading* yang mudah diakses juga menjadi faktor yang mempercepat pertumbuhan jumlah investor di Indonesia. Kemajuan teknologi dan informasi juga berperan penting dalam proses edukasi masyarakat mengenai potensi keuntungan dari investasi kripto[5]. Salah satu *cryptocurrency* yang paling populer dan menjadi acuan dalam dunia aset digital adalah *Bitcoin*. *Bitcoin* sering disebut sebagai pelopor dalam dunia *cryptocurrency* karena menjadi mata uang kripto pertama yang diperkenalkan secara luas ke publik. Diluncurkan pada tahun 2009 oleh sosok yang dikenal dengan nama samaran Satoshi Nakamoto, *Bitcoin* menawarkan konsep baru dalam sistem keuangan yang sepenuhnya terdesentralisasi dan tidak bergantung pada bank atau lembaga keuangan tradisional. Hingga saat ini, *bitcoin* tetap menjadi aset digital yang paling banyak diperdagangkan dan memiliki nilai yang tinggi di pasar kripto global. Popularitasnya yang tetap bertahan ini sebagian besar disebabkan oleh sistem keamanan *bitcoin* yang menggunakan teknik kriptografi canggih untuk menjamin integritas dan keamanan setiap transaksi[6][7].

Dengan semakin tingginya minat masyarakat terhadap *cryptocurrency*, berbagai jenis aset digital ini banyak dibahas di media sosial dan platform online lainnya. Media sosial seperti X (dulu dikenal sebagai *Twitter*) menjadi tempat di mana para pengguna berbagi pengalaman, pandangan, dan analisis mereka tentang *cryptocurrency* [2]. Beberapa orang melihat *cryptocurrency* sebagai peluang investasi yang menjanjikan, sementara yang lain melihatnya sebagai aset yang sangat fluktuatif dan berisiko tinggi. Di satu sisi, pengguna yang optimis melihat *cryptocurrency* sebagai cara untuk melindungi nilai aset di tengah ketidakpastian ekonomi global, inflasi, dan ketidakpastian pasar tradisional[8]. Di sisi lain, ada juga yang khawatir akan risiko tinggi dari fluktuasi harga yang ekstrem, risiko penipuan, serta kurangnya regulasi yang jelas di pasar kripto[9]. Dari diskusi-diskusi ini muncul berbagai permasalahan, seperti volatilitas harga yang ekstrem, keamanan investasi, hingga ketidakpastian regulasi.

Beragamnya opini tentang *cryptocurrency* menciptakan ruang untuk analisis sentimen di media sosial X, di mana pandangan masyarakat terhadap *cryptocurrency* dapat dikelompokkan menjadi dua kategori utama yakni opini positif dan opini negatif [10][11]. Pengguna yang memiliki pandangan positif biasanya menyoroti potensi keuntungan besar, inovasi teknologi *blockchain*, dan aksesibilitas investasi yang lebih luas. Di sisi lain, opini negatif sering kali muncul dari kekhawatiran terhadap fluktuasi harga yang tajam, risiko penipuan, dan kurangnya regulasi yang jelas. Analisis sentimen ini memberikan gambaran yang lebih lengkap tentang persepsi masyarakat terhadap *cryptocurrency*, yang mencerminkan bagaimana peluang dan risiko dari aset digital ini dipahami dan dibahas di ruang publik online.

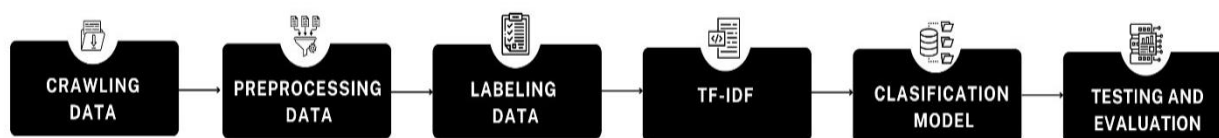
Analisis sentimen, yang juga dikenal sebagai penambangan opini, adalah metode yang digunakan untuk secara otomatis memahami, mengambil, dan mengolah data teks untuk mengidentifikasi sentimen yang terkandung dalam suatu pernyataan, apakah bersifat positif atau negatif[12]. Metode ini berguna untuk menemukan solusi atas masalah yang telah diidentifikasi sebelumnya. Perbandingan kinerja ketiga algoritma, yaitu *Random Forest*, *K-Nearest Neighbors (KNN)*, dan *Support Vector Machine (SVM)*, diperlukan karena masing-masing algoritma memiliki karakteristik, keunggulan, dan kelemahan yang berbeda dalam menangani analisis sentimen. *Random Forest*, meskipun cenderung memberikan hasil yang stabil dan akurat karena kemampuannya mengurangi *overfitting* melalui kombinasi pohon keputusan[13], terkadang memerlukan waktu komputasi yang lebih tinggi pada dataset besar. *KNN*, di sisi lain, menunjukkan hasil yang cukup baik dengan pendekatan yang sederhana, namun akurasi sangat bergantung pada pemilihan jumlah tetangga (*k*) yang optimal, sehingga kurang andal pada dataset yang tidak seimbang atau berdimensi tinggi[14]. Sementara itu, *SVM*, meskipun dikenal memiliki kemampuan untuk memisahkan data dalam dimensi tinggi dengan kernel yang fleksibel, sering kali membutuhkan tuning parameter yang cermat agar dapat memberikan hasil akurasi yang optimal, terutama pada dataset yang sangat kompleks dan dinamis seperti data opini *cryptocurrency*[15]. Oleh karena itu, penting untuk melakukan perbandingan kinerja ketiga algoritma tersebut dalam analisis sentimen agar dapat menentukan algoritma yang paling efektif dan sesuai dengan karakteristik data yang digunakan. Selain itu, hasil perbandingan ini juga dapat memberikan panduan bagi peneliti atau praktisi dalam memilih algoritma yang tepat untuk diterapkan pada analisis sentimen di masa depan.

Penelitian sebelumnya yang dilakukan oleh Catur Rahmawati dan Pristi Sukmasetya menunjukkan bahwa algoritma *Random Forest*, *KNN*, dan *SVM* memberikan hasil yang berbeda tergantung pada karakteristik data yang digunakan, dengan nilai akurasi yang relatif berdekatan dalam analisis pandangan masyarakat terhadap Kebijakan Kominfo Terkait Pemblokiran Situs non-PSE (85,8% untuk *Random Forest* dan *KNN*, serta 85,4% untuk *SVM*)[16]. Namun, pada penelitian Muhammad Yasir dkk dalam analisis sentimen terhadap Fatwa MUI Nomor 83 Tahun 2023, ditemukan adanya perbedaan signifikan, di mana *Random Forest* memberikan akurasi tertinggi (67%), diikuti oleh *SVM* (63%) dan *KNN* (53%)[17].

Perbedaan tersebut menunjukkan bahwa performa algoritma sangat dipengaruhi oleh karakteristik data dan tujuan analisis. Oleh sebab itu, penelitian ini bertujuan untuk mengeksplorasi lebih jauh keandalan masing-masing algoritma dalam menganalisis sentimen data opini terkait *cryptocurrency*. Dengan demikian, penelitian ini tidak hanya membantu dalam memahami performa algoritma pada topik *cryptocurrency*, tetapi juga memberikan wawasan tentang penerapan algoritma yang lebih luas pada berbagai jenis data sentimen yang dinamis dan kompleks.

2. METODOLOGI PENELITIAN

Penelitian ini melibatkan 6 tahapan sentimen analisis, yakni *crawling data* pada X, lalu *preprocessing data*, *labeling data*, pembobotan *TF-IDF*, klasifikasi terhadap algoritma *Random forest*, *KNN*, dan *SVM*, serta testing dan evaluasi data. Tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Crawling Data

Penelitian ini menggunakan metode *crawling* data untuk mendapatkan dataset berupa opini masyarakat tentang *cryptocurrency* di media sosial X. Proses *crawling* dilakukan untuk mengumpulkan berbagai bentuk data teks, seperti postingan, komentar, atau interaksi lainnya yang secara spesifik membahas *cryptocurrency*. Dalam penelitian ini, digunakan *library Harvest* sebagai alat utama untuk melakukan proses pengambilan data secara otomatis. *Library* ini memungkinkan peneliti untuk menyesuaikan parameter *crawling* sesuai kebutuhan, seperti mengatur kata kunci pencarian, rentang waktu postingan, serta tipe data yang akan dikumpulkan[18]. Rentang waktu dalam pengumpulan data pada penelitian ini dari 04 agustus 2024 s.d. 20 september 2024. Alur *crawling* data dapat dilihat pada Gambar 2.



Gambar 2. Alur *Crawling* Data

2.2 Preprocessing Data

Pada tahap *preprocessing* data, peneliti melakukan beberapa langkah penting untuk memastikan data yang dikumpulkan dari hasil *crawling* siap digunakan dalam analisis lebih lanjut. Tahapan *preprocessing* ini bertujuan untuk membersihkan dan menstrukturkan data agar lebih konsisten dan relevan[19]. Adapun langkah-langkah yang dilakukan pada tahap ini yakni *cleansing* data, *casefolding*, *tokenizing*, *stopword removal*, dan *stemming*.

2.3 Labeling Data

Tahap selanjutnya yakni labeling data, Tahap *labeling* adalah proses untuk memberikan label pada setiap *post* atau *tweet*. Dalam penelitian ini, digunakan dua kategori label, yaitu positif dan negatif. Dengan memberikan label tersebut, kita dapat menganalisis *post* yang telah dikelompokkan sesuai dengan kategorinya, sehingga memudahkan dalam memahami sentimen yang terkandung dalam setiap *post* atau *tweet*.

2.4 TF-IDF

TF-IDF (*Term Frequency - Inverse Document Frequency*) adalah sebuah metode yang digunakan untuk memberikan bobot pada fitur kata berdasarkan seberapa sering kata tersebut muncul dalam suatu dokumen [20]. Teknik ini menghitung frekuensi kemunculan kata (*Term Frequency*, *TF*) dan menggabungkannya dengan *Inverse Document Frequency* (*IDF*), yang mengurangi bobot kata-kata umum yang sering muncul di banyak dokumen.

a. *Term Frequency* (*TF*)

TF dilakukan untuk menghitung frekuensi kemunculan kata tertentu dalam sebuah dokumen. Berikut adalah rumus *TF*:

$$TF(t, d) = \frac{\text{jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{total kata dalam dokumen } d} \quad (1)$$

Pada rumus 1 digunakan untuk menghitung seberapa sering suatu kata *t* muncul dalam dokumen tertentu *d*. Nilai *TF* memberikan bobot berdasarkan jumlah kemunculan kata dalam dokumen tersebut, semakin sering kata muncul, semakin tinggi nilai *TF*-nya. Namun, *TF* hanya melihat frekuensi relatif dalam satu dokumen tanpa mempertimbangkan konteks dokumen lainnya.

b. *Inverse Document Frequency* (*IDF*)

IDF digunakan untuk mengukur seberapa jarang suatu kata muncul di seluruh dokumen dalam dokumen. Berikut adalah rumus *IDF*:

$$IDF(t, D) = \frac{\text{total dokumen dalam } D}{\text{jumlah dokumen yang mengandung kata } t} \quad (2)$$

Pada rumus 2 digunakan untuk mengukur seberapa jarang kata *t* muncul di seluruh kumpulan dokumen *D*. Semakin sedikit dokumen yang mengandung kata *t*, semakin besar nilai *IDF*-nya. Hal ini membantu membedakan kata-kata yang umum (seperti "dan" atau "adalah") dari kata-kata yang lebih spesifik dan relevan dalam dokumen tertentu.

c. *TF-IDF*

Setelah nilai *TF* dan *IDF* dihitung, nilai *TF-IDF* dari suatu kata diperoleh dengan mengalikan kedua nilai tersebut:

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

Pada rumus 3 menggabungkan kedua konsep sebelumnya, yaitu frekuensi kemunculan kata *t* dalam dokumen *d* (*TF*) dengan tingkat keunikan kata *t* dalam kumpulan dokumen *D* (*IDF*). Nilai *TF-IDF* digunakan untuk memberikan bobot pada kata-kata yang sering muncul dalam dokumen tertentu tetapi jarang ditemukan di

dokumen lainnya, sehingga membantu mengidentifikasi kata-kata yang lebih penting atau relevan dalam dokumen tersebut.

Setelah menghitung nilai *TF-IDF* untuk setiap kata dalam setiap dokumen, hasilnya biasanya disusun dalam bentuk matriks, di mana setiap baris mewakili dokumen, dan setiap kolom mewakili kata dalam kumpulan teks.

2.5 Classification Model

2.5.1 Random Forest

Random Forest adalah algoritma klasifikasi berbasis *ensemble* yang menggabungkan banyak *decision tree* untuk meningkatkan akurasi dan kestabilan prediksi. Teknik ini menggunakan pendekatan *bagging* (*Bootstrap Aggregating*), di mana beberapa *subset* dari data latih diambil secara acak dengan penggantian (*bootstrap sampling*), dan setiap *decision tree* dilatih pada *subset* tersebut. Setiap *tree* berkembang secara independen dan fitur yang dipilih untuk membagi *node* juga dipilih secara acak. Setelah semua *tree* terbentuk, hasil prediksi akhir ditentukan dengan cara *voting majority* (klasifikasi) atau rata-rata (regresi). Algoritma *Random Forest* mampu mengatasi masalah *overfitting* yang sering terjadi pada *decision tree* tunggal, menjadikannya algoritma yang lebih stabil dan akurat[21] berikut adalah rumus *voting majority* untuk klasifikasi.

$$\hat{y} = \text{mode}(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n) \quad (4)$$

Untuk regresi, prediksi akhir \hat{y} ditentukan dengan menghitung rata-rata dari semua prediksi *tree*:

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i \quad (5)$$

Pada rumus 4 digunakan untuk menentukan prediksi akhir pada *Random Forest* dalam kasus klasifikasi, di mana *mode* adalah nilai yang paling sering muncul sebagai hasil voting dari semua prediksi \hat{y}_i dari setiap *decision tree* di dalam hutan (*forest*). Pada rumus 5 digunakan dalam kasus regresi, di mana \hat{y} adalah rata-rata dari semua prediksi \hat{y}_i yang dihasilkan oleh setiap *decision tree* di dalam hutan (*forest*). Dalam hal ini, $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$ adalah hasil prediksi dari masing-masing *decision tree*, dengan n menunjukkan jumlah total *decision tree* yang membentuk *Random Forest*. Setiap \hat{y}_i mewakili prediksi dari *decision tree* ke- i . Proses ini dilakukan untuk menghasilkan prediksi akhir yang lebih stabil dan akurat dengan memanfaatkan kekuatan kolektif dari seluruh *decision tree* yang ada.

2.5.2 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (*KNN*) adalah algoritma klasifikasi yang menentukan kelas data baru berdasarkan kedekatannya dengan data latih yang sudah diberi label. Algoritma ini menghitung jarak antara data uji dan data latih, lalu memilih sejumlah tetangga terdekat (k) untuk menentukan kelas data baru berdasarkan mayoritas kelas tetangga tersebut. Nilai k memengaruhi hasil klasifikasi, di mana k yang lebih kecil fokus pada pola lokal, sementara k yang lebih besar menangkap pola yang lebih umum[22].

$$d(x, x_i) = \sqrt{\sum_{j=1}^n (x_j - x_{i,j})^2} \quad (6)$$

Pada rumus 6 digunakan untuk menghitung jarak Euclidean antara data uji (x) dan data latih (x_i). Jarak ini menunjukkan seberapa mirip atau dekat dua data berdasarkan nilai fitur-fiturnya. Dalam rumus ini, $d(x, x_i)$ adalah jarak antara data uji dan data latih, x_j adalah nilai fitur ke- j dari data uji, dan $x_{i,j}$ adalah nilai fitur ke- j dari data latih. Sementara itu, n adalah jumlah fitur yang digunakan untuk perhitungan jarak. Pada bagian $\sum_{j=1}^n (x_j - x_{i,j})^2$ menghitung jumlah kuadrat selisih antara nilai fitur data uji dan data latih untuk setiap fitur, sedangkan akar kuadrat ($\sqrt{\dots}$) digunakan untuk mendapatkan jarak geometris langsung dalam ruang fitur n -dimensi. Rumus ini sering diterapkan dalam algoritma seperti *K-Nearest Neighbors* (*KNN*) untuk menentukan data latih yang paling dekat dengan data uji berdasarkan jarak terkecil.

2.5.3 Support Vector Machine (SVM)

Algoritma *Support Vector Machine* (*SVM*) adalah metode yang umum digunakan untuk menyelesaikan masalah klasifikasi dan regresi. *SVM* dikembangkan sebagai peningkatan dari *Support Vector Classifier* dan dirancang untuk memperbaiki kelemahan klasifikasi margin maksimal yang hanya efektif pada data yang dapat dipisahkan secara linier. Dalam *SVM*, data dipetakan ke ruang fitur berdimensi tinggi, di mana algoritma mencari *hyperplane* yang optimal untuk memisahkan dua kelas secara linear [23]. *Hyperplane* ini dipilih dengan cara memaksimalkan *margin*, yaitu jarak antara titik data terdekat dari kedua kelas. Meskipun demikian, kesalahan klasifikasi dapat terjadi jika data tidak dapat dipisahkan dengan jelas menggunakan pemisahan linier.

2.6 Testing and Evaluation

Proses testing dan evaluasi bertujuan untuk menilai performa model yang telah dibangun. Setelah model dilatih menggunakan data latih, data uji digunakan untuk menguji model dalam memprediksi label data baru. Pengujian

dilakukan untuk melihat seberapa baik model dapat menggeneralisasi terhadap data yang tidak pernah dilihat sebelumnya. Beberapa metrik evaluasi yang umum digunakan dalam klasifikasi meliputi:

a. Akurasi Proporsi prediksi yang benar dari seluruh prediksi yang dibuat.

$$Accuracy = \frac{Prediksi\ Benar}{Total\ Prediksi} \tag{7}$$

b. Precision: Proporsi prediksi positif yang benar.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Positive\ (FP)} \tag{8}$$

c. Recall (Sensitivity): Proporsi kasus positif yang benar-benar terdeteksi oleh model.

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)} \tag{9}$$

d. F1-Score: Rata-rata harmonis antara precision dan recall.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

3. HASIL DAN PEMBAHASAN

3.1 Crawling Data

Jumlah data yang diperoleh melalui metode *crawling* adalah 8.130, yang semuanya berasal dari media sosial X dan mencakup berbagai bentuk interaksi masyarakat yang berkaitan dengan topik *cryptocurrency*. Data ini terdiri dari postingan, komentar, serta bentuk interaksi lainnya, seperti retweet atau balasan yang secara eksplisit membahas atau menanggapi isu-isu terkait *cryptocurrency*.

Data dikumpulkan dalam rentang waktu mulai dari 04 Agustus 2024 hingga 20 September 2024, yang memungkinkan untuk melihat dinamika opini masyarakat dalam periode tertentu. Periode ini juga sangat relevan mengingat adanya perkembangan signifikan dalam dunia *cryptocurrency*, seperti perubahan harga, kebijakan regulasi baru, atau peristiwa penting lainnya yang dapat mempengaruhi persepsi masyarakat. Dengan demikian, data yang terkumpul tidak hanya memberikan gambaran tentang bagaimana masyarakat berinteraksi dengan topik *cryptocurrency*, tetapi juga mencerminkan reaksi mereka terhadap berbagai peristiwa atau kebijakan yang terjadi dalam rentang waktu tersebut. Data hasil *crawling* yang ditampilkan pada gambar masih berupa data mentah atau data kotor yang membutuhkan proses pembersihan sebelum dapat digunakan untuk analisis lebih lanjut. Hal ini terlihat dari adanya elemen-elemen yang tidak relevan atau mengganggu, seperti tag pengguna (contoh: @duta_crypto), URL (contoh: https://t.co/CSS7J5uwE1), karakter khusus atau encoding yang tidak terbaca (contoh: â€œâ€dan), serta potongan teks yang tidak memiliki struktur jelas. Selain itu, data juga dapat mengandung redundansi, noise, atau informasi yang tidak berkontribusi secara signifikan pada tujuan analisis. Hasil *crawling* dapat dilihat pada Gambar 3.

	id_str	full_text	username
0	1.834162e+18	@duta_crypto Kalau saya ke bittime	FSombowadile
1	1.834162e+18	@Stefanny_Crypto @Brainstems_ Selesai	HariiniBigWin01
2	1.834162e+18	Indodax telah mengeluarkan pernyataan resmi ya...	TradersFii
3	1.834161e+18	@gmaazs ati-ati bro nanti ditanya mclaren nya ...	crypto_legend88
4	1.834161e+18	@MPRAVEENREDDY13 Rice bag ki alvatu padi schoo...	GKcrypto369
...
8125	1.831611e+18	5. Risiko sebelumnya disclaimer ya karena ini ...	0xLoonar
8126	1.831611e+18	1. Mekanisme Campaign mekanisme campaign ini s...	0xLoonar
8127	1.831604e+18	@shahh Told ya https://t.co/CSS7J5uwE1	Crypto_Inside_
8128	1.831602e+18	Kemitraan antara @KAVA_CHAIN â€œâ€dan @BitGo ...	1ONE_Crypto
8129	1.831586e+18	Terungkap! Penambang Bitcoin Rusia Diam-diam K...	duniafintechID

8130 rows x 3 columns

Gambar 3. Hasil Crawling Data

3.2 Preprocessing Data



Gambar 4. Tahapan Preprocessing Data



Pada tahap *preprocessing*, data yang diperoleh melalui *crawling* diolah melalui serangkaian langkah untuk memastikan kualitasnya sebelum analisis lebih lanjut. Proses dimulai dengan *cleansing*, yaitu menghilangkan karakter atau simbol yang tidak diperlukan, seperti tanda baca atau emotikon yang tidak relevan, untuk membersihkan teks dari elemen-elemen yang dapat mengganggu analisis. Setelah itu, dilakukan *casefolding* untuk menyamakan format huruf, mengubah semua teks menjadi huruf kecil agar variasi penulisan seperti "Bitcoin" dan "bitcoin" dianggap sebagai kata yang sama. Langkah berikutnya adalah *tokenizing*, di mana data dipecah menjadi kata-kata atau token, memudahkan analisis lebih lanjut. Kemudian, *stopword removal* diterapkan untuk menghapus kata-kata umum yang tidak memberikan kontribusi signifikan terhadap analisis, seperti "dan", "atau", "yang", sehingga fokus hanya pada kata-kata yang lebih relevan[24]. Setelah itu, *stemming* dilakukan untuk mengembalikan kata-kata ke bentuk dasarnya, misalnya mengubah "berinvestasi" menjadi "investasi", agar variasi kata yang memiliki makna serupa dapat dianalisis sebagai entitas yang sama. Terakhir, *drop data duplicate* diterapkan untuk menghilangkan duplikasi data, memastikan bahwa analisis tidak terdistorsi oleh entri yang sama yang muncul lebih dari sekali. Hasil data yang telah melalui tahap *preprocessing* ini akan lebih terstruktur, bersih, dan siap untuk dianalisis lebih lanjut, dan contoh hasil yang telah diproses dapat dilihat pada Tabel 1.

Tabel 1. Hasil *Preprocessing*

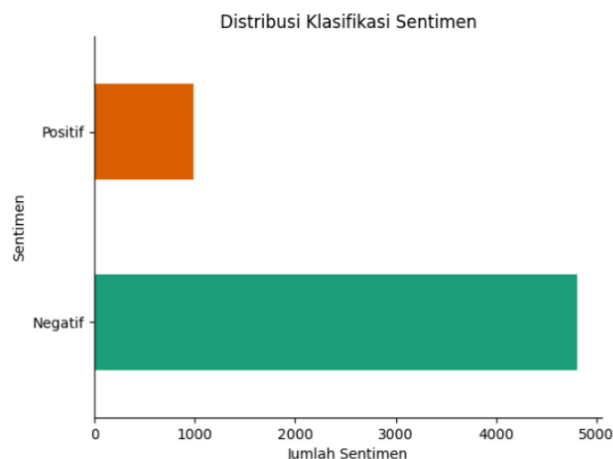
Tahapan	Hasil <i>Preprocessing</i>
<i>Data awal</i>	Katanya Crypto ga ada gunanya? Gunanya Kripto adalah membunuh KPR KKB KTA KARTU KREDIT KIRIM UANG VIA WESTERN UNION YG BIAYANYA 300RB dan kredit kredit lintah darat yg bikin miskin. https://t.co/nVg56dhdb
<i>Cleansing</i>	Katanya Crypto ga ada gunanya Gunanya Kripto adalah membunuh KPR KKB KTA KARTU KREDIT KIRIM UANG VIA WESTERN UNION YG BIAYANYA RB dan kredit kredit lintah darat yg bikin miskin
<i>Casefolding</i>	katanya crypto ga ada gunanya gunanya kripto adalah membunuh kpr kkb kta kartu kredit kirim uang via western union yg biayanya rb dan kredit kredit lintah darat yg bikin miskin
<i>Tokenizing</i>	['katanya', 'crypto', 'ga', 'ada', 'gunanya', 'gunanya', 'kripto', 'adalah', 'membunuh', 'kpr', 'kbb', 'kta', 'kartu', 'kredit', 'kirim', 'uang', 'via', 'western', 'union', 'yg', 'biayanya', '300rb', 'dan', 'kredit', 'kredit', 'lintah', 'darat', 'yg', 'bikin', 'miskin']
<i>Stopword removal</i>	['katanya', 'crypto', 'gunanya', 'gunanya', 'kripto', 'membunuh', 'kpr', 'kbb', 'kta', 'kartu', 'kredit', 'kirim', 'uang', 'western', 'union', 'biayanya', '300rb', 'kredit', 'kredit', 'lintah', 'darat', 'bikin', 'miskin']
<i>Stemming</i>	kata crypto guna guna kripto bunuh kpr kkb kta kartu kredit kirim uang western union biaya 300rb kredit kredit lintah darat bikin miskin

Tabel 1 menunjukkan tahapan preprocessing yang dilakukan pada data teks sebelum dianalisis lebih lanjut. Preprocessing bertujuan untuk membersihkan dan menyederhanakan data agar lebih mudah diolah. Tahapan-tahapan yang dilakukan adalah sebagai berikut:

- Data awal:** Merupakan teks asli yang diperoleh dari proses *crawling*. Data ini masih mengandung karakter khusus, tanda baca, tautan, dan elemen lain yang tidak relevan untuk analisis.
- Cleansing:** Pada tahap ini, karakter khusus, tanda baca, angka, dan tautan dihapus sehingga teks menjadi lebih bersih. Sebagai contoh, tautan dan angka "300RB" dihilangkan.
- Casefolding:** Seluruh huruf dalam teks diubah menjadi huruf kecil untuk menyamakan format dan menghindari kesalahan pengolahan akibat perbedaan huruf besar dan kecil.
- Tokenizing:** Teks dibagi menjadi daftar kata-kata terpisah (token). Tahap ini mempermudah analisis kata per kata, seperti pada daftar ['katanya', 'crypto', 'ga', ...].
- Stopword removal:** Kata-kata umum yang tidak memiliki makna signifikan untuk analisis, seperti "dan", "yg", dan "ada", dihapus. Ini bertujuan untuk mengurangi gangguan dan fokus pada kata-kata yang lebih bermakna.
- Stemming:** Kata-kata dikembalikan ke bentuk dasarnya, seperti "membunuh" menjadi "bunuh". Ini membantu dalam menyamakan kata yang memiliki makna serupa.

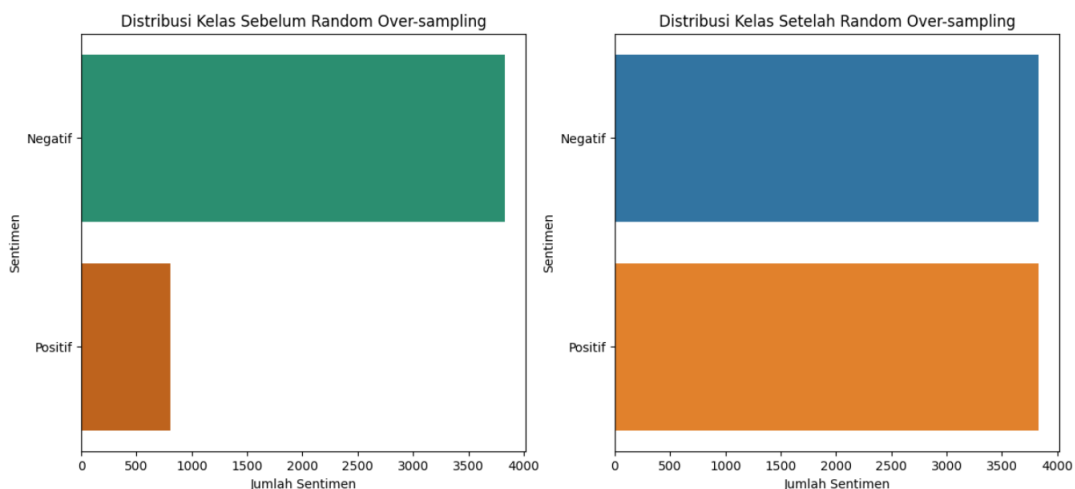
3.3 Labeling Data

Setelah dilakukan tahap *preprocessing*, tahap selanjutnya adalah labeling data. Penelitian ini menggunakan dua kelas untuk melabeli setiap post atau tweet yang telah diproses, yaitu positif dan negatif. *Labeling* dilakukan secara otomatis menggunakan library *VADER (Valence Aware Dictionary and Sentiment Reasoner)*, yang merupakan metode *lexicon-based* untuk analisis sentimen. *VADER* cocok untuk analisis teks pendek seperti tweet, karena mampu mengukur sentimen positif, negatif, atau netral dengan memperhitungkan bahasa informal, tanda baca, huruf kapital, dan emotikon yang sering digunakan di media sosial [25]. Berikut adalah hasil distribusi kelas sentimen yang telah dianalisis.



Gambar 5. Distribusi Kelas Sentimen

Berdasarkan Gambar 5, terlihat adanya ketidakseimbangan jumlah kelas antara sentimen negatif dan positif, di mana sentimen negatif memiliki jumlah data yang jauh lebih banyak. Untuk mengatasi hal ini, peneliti menerapkan teknik *random over-sampling* guna menambah jumlah data pada sentimen Positif hingga jumlahnya setara dengan sentimen negatif.



Gambar 6. Distribusi Kelas Sentimen sebelum dan sesudah *Random Over-Sampling*

Pada Gambar 6 menunjukkan distribusi kelas sentimen sebelum dan sesudah dilakukan proses *Random Over-sampling*. Pada grafik kiri (sebelum *over-sampling*), terlihat bahwa dataset tidak seimbang, di mana jumlah data sentimen negatif jauh lebih banyak dibandingkan sentimen positif. Ketidakseimbangan ini dapat menyebabkan bias pada model *machine learning*. Setelah dilakukan proses *Random Over-sampling* (grafik kanan), jumlah data pada kedua kelas menjadi seimbang dengan cara menduplikasi data kelas positif (kelas minoritas). Proses ini bertujuan untuk mengatasi ketidakseimbangan dataset sehingga model dapat belajar secara adil dan menghasilkan prediksi yang lebih akurat untuk kedua kelas.

3.4 Hasil Pengujian

Pada tahap ini dilakukan pengujian data menggunakan tiga algoritma untuk membandingkan hasil akurasi dan kinerja masing-masing algoritma. Algoritma yang digunakan adalah *Random Forest*, *K-Nearest Neighbors (KNN)*, dan *Support Vector Machine (SVM)*. Setiap algoritma dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk memastikan performa terbaik dalam klasifikasi sentimen. Hasil dari pengujian ini akan menentukan algoritma mana yang paling cocok digunakan untuk analisis sentimen pada dataset yang digunakan. Berikut adalah tabel hasil *accuracy* pengujian pada ketiga algoritma.

Tabel 2. Hasil *Accuracy*

Algoritma	Accuracy
<i>Random Forest</i>	0,97
<i>KNN</i>	0,87
<i>SVM</i>	0,98

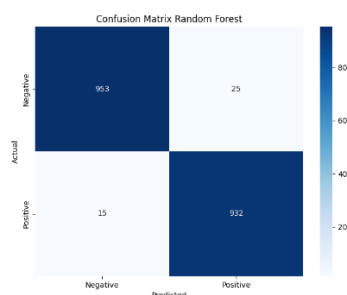
Berdasarkan Tabel 2, hasil akurasi dapat dilihat bahwa algoritma *SVM* memiliki nilai akurasi tertinggi, yaitu 0,98 atau 98% dibandingkan dengan algoritma *Random Forest* yang memiliki akurasi sebesar 0,97 atau 97% dan algoritma *KNN* dengan akurasi sebesar 0,87 atau 87%. Perbedaan ini menunjukkan bahwa algoritma *SVM* lebih unggul dalam mengklasifikasikan data secara tepat. Selain akurasi, penelitian ini juga menggunakan metrik lain seperti *precision*, *recall*, dan *F1-score* sebagai bahan pertimbangan untuk mendapatkan gambaran yang lebih komprehensif mengenai performa masing-masing algoritma, terutama dalam menangani ketidakseimbangan data dan kesalahan klasifikasi. Tabel hasil *precision*, *recall*, dan *F1-score* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Precision, Recall, dan F1-score.

	Precision		Recall		F1-Score	
	Positive	Negative	Positive	Negative	Positive	Negative
Random Forest	0,97	0,98	0,98	0,97	0,98	0,98
KNN	0,83	0,92	0,93	0,82	0,88	0,87
SVM	0,99	0,98	0,98	0,99	0,98	0,98

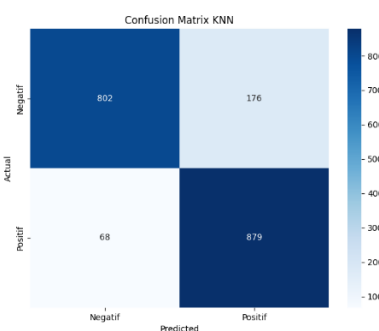
Berdasarkan Tabel 3, hasil *precision*, *recall*, dan *f1-score* di atas, algoritma *SVM* memiliki hasil tertinggi, yang membuktikan bahwa *SVM* mampu memberikan performa terbaik dalam klasifikasi sentimen dibandingkan dengan algoritma lainnya. Dengan *precision* sebesar 0,99 atau 99% untuk kelas positif dan *recall* sebesar 0,99 atau 99% untuk kelas negatif, *SVM* menunjukkan kemampuannya dalam meminimalkan kesalahan klasifikasi, baik dalam mendeteksi kelas positif maupun negatif. Hasil tersebut membuat *SVM* unggul dalam situasi di mana ketepatan klasifikasi dan keseimbangan antara *precision* dan *recall* sangat penting. Algoritma *Random Forest* juga menunjukkan performa yang baik, sementara *KNN* sedikit tertinggal dalam hal akurasi dan *f1-score*. Secara keseluruhan, *SVM* adalah algoritma yang paling direkomendasikan untuk dataset ini, terutama karena kemampuannya menangani data dengan ketidakseimbangan kelas.

Dengan mempertimbangkan hasil *precision*, *recall*, dan *f1-score* yang telah dipaparkan, terlihat bahwa *SVM* memiliki performa unggul dalam mengklasifikasikan data sentimen secara akurat dibandingkan algoritma lainnya. *Random Forest* juga menunjukkan hasil yang baik, dengan skor *precision* dan *recall* yang mendekati *SVM*, sementara *KNN* memiliki performa yang sedikit lebih rendah, terutama dalam hal konsistensi prediksi. Untuk memperoleh pemahaman yang lebih mendalam mengenai kualitas klasifikasi tiap algoritma dalam menangani data sentimen ini, analisis lebih lanjut terhadap *confusion matrix* masing-masing algoritma akan dilakukan. Melalui *confusion matrix* dapat dilihat distribusi prediksi benar dan salah pada setiap kelas secara lebih spesifik, yang membantu mengidentifikasi pola kesalahan dan ketepatan dalam prediksi pada setiap model, berikut hasil *confusion matrix* ketiga algoritma.



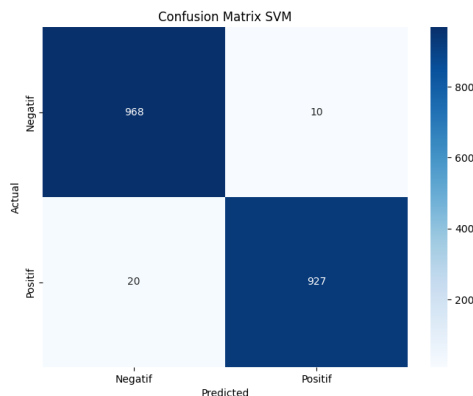
Gambar 7. Confusion Matrix algoritma Random Forest

Berdasarkan Gambar 7, hasil *confusion matrix* dari algoritma *Random Forest* menunjukkan bahwa model ini memiliki performa yang cukup baik dalam memprediksi kelas *negative* dan *positive*. Dengan 953 *True Negative (TN)* dan 932 *True Positive (TP)*, model ini mampu mengklasifikasikan sebagian besar sampel dengan benar. Namun, masih terdapat 25 *False Positive (FP)* dan 15 *False Negative (FN)*, yang menunjukkan beberapa kesalahan dalam prediksi.



Gambar 8. Confusion Matrix algoritma KNN

Berdasarkan Gambar 8, hasil *confusion matrix* dari algoritma *KNN* membirikan performa yang cukup baik dengan memprediksi kelas *negative* dan *positive* dengan 968 *True Negative (TN)*, dan 927 *True Positive (TP)*, model ini mampu mengklasifikasikan sebagian besar sampel dengan benar. Namun masih terdapat 10 *False Positive (FP)* yang berarti ada 10 sampel yang diprediksi sebagai positif padahal sebenarnya negatif. Selain itu, terdapat 15 *False Negative (FN)*, yang menunjukkan ada 15 sampel yang diprediksi sebagai negatif padahal sebenarnya positif.



Gambar 9. *Confusion Matrix* algoritma SVM

Berdasarkan Gambar 9, hasil *confusion matrix* dari algoritma SVM menunjukkan bahwa model berhasil memprediksi 968 *True Negative (TN)* dan 927 *True Positive (TP)*, yang berarti model ini mampu mengklasifikasikan sebagian besar sampel dengan benar. Namun, terdapat 10 *False Positive (FP)*, yaitu kasus di mana model memprediksi data sebagai positif padahal sebenarnya negatif, serta 20 *False Negative (FN)*, di mana model memprediksi data sebagai negatif padahal sebenarnya positif.

Berdasarkan hasil *confusion matrix* dari tiga algoritma (*Random Forest, KNN, dan SVM*), dapat disimpulkan bahwa ketiga model memiliki performa yang cukup baik dalam memprediksi kelas *negative* dan *positive*, dengan sebagian besar sampel berhasil diklasifikasikan dengan benar. Model *Random Forest* menghasilkan 953 *True Negative (TN)* dan 932 *True Positive (TP)* dengan 25 *False Positive (FP)* dan 15 *False Negative (FN)*. Model *KNN* menunjukkan performa yang lebih baik dalam mengurangi jumlah *False Positive* dengan 10 *FP*, tetapi masih terdapat 15 *FN*, dengan total 968 *TN* dan 927 *TP*. Model SVM juga menunjukkan hasil serupa dengan *KNN*, yaitu 968 *TN*, 927 *TP*, 10 *FP*, dan 20 *FN*. Dari ketiga model ini, *KNN* dan *SVM* memiliki jumlah prediksi benar yang sama, namun *KNN* lebih baik dalam mengurangi kesalahan *False Negative* dibandingkan *SVM*. Secara keseluruhan, algoritma *KNN* memberikan hasil yang sedikit lebih baik dibandingkan dua algoritma lainnya dalam hal keseimbangan antara *True Negative* dan *True Positive* serta jumlah kesalahan prediksi.

3.5 Visualisasi Wordcloud

Visualisasi data dengan menggunakan wordcloud menampilkan kata-kata yang paling sering muncul dalam opini masyarakat tentang *cryptocurrency* di media sosial X. Kata-kata dengan frekuensi kemunculan tinggi ditampilkan lebih besar, sementara kata-kata yang jarang muncul terlihat lebih kecil. *Wordcloud* ini memudahkan identifikasi topik atau isu utama yang dibahas pengguna, seperti "Bitcoin," "Cryptocurrency," "Ethereum" dan "investasi" yang mencerminkan fokus perbincangan seputar peluang investasi dan tantangan terkait *cryptocurrency*. Visualisasi ini memberikan gambaran cepat tentang sentimen dominan dan tema populer di diskusi online. Hasil *wordcloud* dapat dilihat pada Gambar 10.



Gambar 10. Hasil *Wordcloud*

yang hampir setara dengan SVM, SVM tetap unggul dalam hal pengurangan kesalahan klasifikasi, baik pada kelas positif maupun negatif. Algoritma KNN, meskipun memiliki akurasi yang lebih rendah (87%) dibandingkan SVM dan *Random Forest*, tetap memberikan hasil yang baik terutama dalam mengurangi jumlah *False Positive*, namun sedikit tertinggal dalam hal *recall* dan *f1-score*. Melalui analisis *confusion matrix*, dapat dilihat bahwa ketiga algoritma memiliki kecenderungan yang serupa dalam hal distribusi *True Negative (TN)* dan *True Positive (TP)*, namun SVM dan KNN memiliki keunggulan dalam mengurangi jumlah *False Positive* dan *False Negative*. SVM, meskipun sedikit lebih banyak mengalami *False Negative* dibandingkan KNN, tetap menunjukkan performa yang lebih baik secara keseluruhan. Oleh karena itu, SVM direkomendasikan sebagai algoritma terbaik untuk analisis sentimen pada dataset ini.

REFERENCES

- [1] Kementerian Keuangan RI, "Yuk, Berkenalan dengan Kripto!," <https://www.djkn.kemenkeu.go.id/kpknl-tangerang1/baca-artikel/16059/Yuk-Berkenalan-dengan-Kripto.html>.
- [2] Adis Syahrul, Ade Irma Purnamasari, and Irfan Ali, "Analisis Sentimen Twitter Terhadap Cryptocurrency Menggunakan Algoritma Naive Bayes Dan Decision Tree," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2213–2220, Apr. 2024, doi: <https://doi.org/10.36040/jati.v8i2.8381>.
- [3] T. Wira and E. Suryawijaya, "Memperkuat Keamanan Data melalui Teknologi Blockchain: Mengeksplorasi Implementasi Sukses dalam Transformasi Digital di Indonesia Strengthening Data Security through Blockchain Technology: Exploring Successful Implementations in Digital Transformation in Indonesia," *JSKP (Jurnal Studi Kebijakan Publik)*, vol. 2, no. 1, pp. 55–67, 2023, doi: [10.21787/jskp.2.2023.55-67](https://doi.org/10.21787/jskp.2.2023.55-67).
- [4] Kementerian Perdagangan RI, "Transaksi Kripto Indonesia Sentuh Rp 211 Triliun hingga April 2024," <https://www.kemendag.go.id/berita/pojok-media/transaksi-kripto-indonesia-sentuh-rp-211-triliun-hingga-april-2024>.
- [5] T. A. Pratama, "Eksplorasi Naratif Media: Analisis Framing CNN Indonesia Terhadap Pelanggan Aset Kripto," *JKOMDIS : Jurnal Ilmu Komunikasi Dan Media Sosial*, vol. 4, no. 2, pp. 350–355, May 2024, doi: [10.47233/jkomdis.v4i2.1618](https://doi.org/10.47233/jkomdis.v4i2.1618).
- [6] Frida Nur Amalina Wijaya, "Bitcoin Sebagai Digital Aset Pada Transaksi Elektronik Di Indonesia (Studi Pada PT. Indodax Nasional Indonesia)," *Jurnal Hukum Bisnis Bonum Commune*, vol. 2, no. 2, pp. 126–136, Jul. 2019, doi: <https://doi.org/10.30996/jhbhc.v2i2.2388>.
- [7] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, A. Mulyani, and J. I. Komputer, "Comparison Of Data Mining Algorithm For Forecasting Bitcoin Crypto Currency Trends," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 2, pp. 245–248, 2022, doi: [10.20884/1.jutif.2022.3.2.194](https://doi.org/10.20884/1.jutif.2022.3.2.194).
- [8] N. Huda and R. Hambali, "Risiko dan Tingkat Keuntungan Investasi Cryptocurrency," *Jurnal Manajemen dan Bisnis*, vol. 17, no. 1, Mar. 2020, doi: [10.29313/performa.v17i1.7236](https://doi.org/10.29313/performa.v17i1.7236).
- [9] M. Albirr, I. Yazidillah, and B. S. Barus, "Studi Tinjauan Pustaka Analisis Risiko Cryptocurrency Sebagai Alat Untuk Berinvestasi," *SOSTECH Jurnal Sosial dan Teknologi*, vol. 3, no. 12, pp. 989–995, Dec. 2023, doi: <https://doi.org/10.59188/jumalsostech.v3i12.1016>.
- [10] J. E. Savero, V. H. Pranatawijaya, and E. Christian, "Analisis Sentimen Pengguna Media Sosial X terhadap Perubahan Harga Bitcoin: Pendekatan Machine Learning," *KONSTELASI (Konvergensi Teknologi dan Sistem Informasi)*, vol. 4, no. 1, pp. 196–208, Jun. 2024, doi: <https://doi.org/10.24002/konstelasi.v4i1.9043>.
- [11] N. Merlina, A. Chandra, and N. A. Mayangky, "Penerapan Pso Untuk Sentimen Analisis Pada Review Mata Uang Kripto Menggunakan Metode Naïve Bayes," *INTI Nusa Mandiri*, vol. 18, no. 2, pp. 115–121, Feb. 2024, doi: [10.33480/inti.v18i2.4982](https://doi.org/10.33480/inti.v18i2.4982).
- [12] I. Hulu and H. Budiati, "Implementasi Text Mining Pada Pengukuran Sentimen Opini Masyarakat Terhadap Universitas Kristen Immanuel," *JURNAL SAINS DAN KOMPUTER*, vol. 7, no. 02, pp. 1–7, Jul. 2023, doi: [10.61179/jurnalinfact.v7i02.445](https://doi.org/10.61179/jurnalinfact.v7i02.445).
- [13] D. Prawita Sari, Z. Halim, and B. Waseso, "Implementasi Machine Learning untuk Deteksi Intrusi pada Jaringan Komputer," *Jurnal Minfo Polgan*, vol. 13, no. 2, 2024, doi: [10.33395/jmp.v13i2.14074](https://doi.org/10.33395/jmp.v13i2.14074).
- [14] A. Sugesti, M. Abdul Mukid, and Tarno Tarno, "Perbandingan Kinerja Mutual K-Nearest Neighbor (Mknn) Dan K-Nearest Neighbor (Knn) Dalam Analisis Klasifikasi Kelayakan Kredit," *Jurnal Gaussian*, vol. 8, no. 3, pp. 366–376, 2019, doi: <https://doi.org/10.14710/j.gauss.8.3.366-376>.
- [15] S. W. Iriananda, R. W. Budiawan, A. Y. Rahman, and I. Istiadi, "Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm, Grid Search Dan Kombinasi N-Gram," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 743–752, Aug. 2024, doi: [10.25126/jtiik.1148244](https://doi.org/10.25126/jtiik.1148244).
- [16] C. Rahmawati and P. Sukmasetya, "Sentimen Analisis Opini Masyarakat Terhadap Kebijakan Kominfo atas Pemblokiran Situs non-PSE pada Media Sosial Twitter," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, p. 1393, Oct. 2022, doi: [10.30865/jurikom.v9i5.4950](https://doi.org/10.30865/jurikom.v9i5.4950).
- [17] M. Yasir, M. Grace Haque, R. Suraji, and C. Author, "Analisis Sentimen Terhadap Kontroversi Fatwa MUI Nomor 83 Tahun 2023 Tentang Pemboikotan Produk yang Terafiliasi Israel," *JURNAL EKONOMI MANAJEMEN SISTEM INFORMASI (JEMSI)*, vol. 5, Mar. 2024, doi: [10.31933/jemsi.v5i4](https://doi.org/10.31933/jemsi.v5i4).
- [18] S. A. Putra and Wijaya Andri, "Analisis Sentimen Artificial Intelligence (AI) Pada Media Sosial Twitter Menggunakan Metode Lexicon Based," *JuSiTik (Jurnal Sistem Informasi Komunikasi)*, vol. 7, no. 1, pp. 21–28, Dec. 2023, doi: <https://doi.org/10.32524/jusitik.v7i1.1042>.
- [19] F. Fiddin, M. Yusuf Syahbarna, and M. Ridwan, "Penggunaan Supervised Learning untuk Prediksi Validitas Ulasan Negatif Aplikasi Tokopedia Berdasarkan Pengalaman Pengguna Ahli," *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 23, pp. 409–417, Aug. 2024, doi: <https://doi.org/10.53513/jis.v23i2.10030>.
- [20] Nurul Istiqamah and Muhammad Rijal, "Klasifikasi Ulasan Konsumen Menggunakan Random Forest dan SMOTE," *Journal of System and Computer Engineering (JSCE)*, vol. 5, Jan. 2024, doi: <https://doi.org/10.61628/jsce.v5i1.1061>.



- [21] M. AUFAR, R. ANDRESWARI, and D. PRAMESTI, “Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study,” in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, 2020, pp. 1–7. doi: 10.1109/ICoDSA50139.2020.9213078.
- [22] S. ULYA, M. ARIEF SOELEMEN, F. BUDIMAN, and M. TEKNIK INFORMATIKA, “Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa Optimization of K Parameters in the K-NN Algorithm for Priority Classification of Village Development Assistance,” Feb. 2021. doi: <https://doi.org/10.33633/tc.v20i1.4215>.
- [23] R. OBIEDAT *et al.*, “Sentiment Analysis of Customers’ Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution,” *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.
- [24] D. RIFALDI, ABDUL FADLIL, and HERMAN, “Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet ‘Mental Health,’” *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 161–171, Apr. 2023, doi: 10.51454/decode.v3i2.131.
- [25] A. MUHAMMAD ADITYA, S. ALAM, and M. ANDAYANI KOMARA, “Analisis Sentimen Pengguna Twitter Terhadap Boikot Produk Israel Pada Merek Dagang Unilever Indonesia Menggunakan Algoritma Long Short Term Memory,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 5, pp. 9883–9890, Oct. 2024, doi: <https://doi.org/10.36040/jati.v8i5.10823>.