

Analisis Sentimen Berita terhadap Bitcoin dengan Metode Klasifikasi K-Nearest Neighbor

Hari Susanto*¹, Arief Setyanto², Alva Hendi Muhammad³

^{1,2,3}Program Studi Magister Informatika, Universitas Amikom Yogyakarta

e-mail: *¹harisusanto@students.amikom.ac.id, ²arief_s@amikom.ac.id, ³alva@amikom.ac.id

Abstract – Globalization has affected various aspects of life, especially in the advancement of information technology and cryptocurrency as an innovation in financial technology. Cryptocurrencies like Bitcoin serve as a medium of exchange, and store of value, although they are not yet recognized as legal tender. The cryptocurrency market is growing rapidly, with more than 10,000 crypto assets in circulation worldwide. The number of users increased significantly from 18 million in 2017 to 516 million in 2023. Bitcoin dominates with a market share of 60.14%, confirming its position as a pioneer and reflecting high interest from investors as well as the public. This research also examines Bitcoin price movements through sentiment analysis using the k-nearest neighbor (KNN) classification method. The results of this study provide in-depth insight into the dynamics of the cryptocurrency market. The KNN method achieved an average accuracy of 74.40%, demonstrating the effectiveness of classifying using this method.

Keywords – sentiment analysis, knn, cryptocurrency, bitcoin

Abstrak – Globalisasi telah mempengaruhi berbagai aspek kehidupan, terutama dalam kemajuan teknologi informasi dan cryptocurrency sebagai inovasi dalam teknologi finansial. Cryptocurrency seperti Bitcoin berfungsi sebagai media pertukaran dan penyimpanan nilai, meski belum diakui sebagai alat pembayaran sah. Pasar cryptocurrency berkembang pesat, dengan lebih dari 10.000 aset crypto beredar di seluruh dunia. Jumlah pengguna meningkat signifikan dari 18 juta pada 2017 menjadi 516 juta pada 2023. Bitcoin mendominasi dengan pangsa pasar 60,14%, menegaskan posisinya sebagai pionir dan mencerminkan minat tinggi dari investor serta masyarakat. Penelitian ini juga mengkaji pergerakan harga Bitcoin melalui analisis sentimen menggunakan metode klasifikasi k-nearest neighbor (KNN). Hasil penelitian ini memberikan wawasan mendalam mengenai dinamika pasar mata uang kripto. Metode KNN mencapai rata-rata akurasi 74,40%, menunjukkan efektivitas pengklasifikasian menggunakan metode ini.

Kata Kunci – analisis sentiment, knn, cryptocurrency, bitcoin

I. PENDAHULUAN

Globalisasi saat ini berdampak signifikan pada berbagai aspek kehidupan, termasuk kemajuan teknologi informasi yang memfasilitasi berbagai aktivitas manusia [1]. Salah satu inovasi dalam teknologi finansial adalah cryptocurrency, atau mata uang digital, yang memperkenalkan paradigma baru dalam sistem pembayaran global [2]. Cryptocurrency berfungsi sebagai media pertukaran, unit akun, dan penyimpanan nilai, meskipun tidak memiliki status tender legal. Pasar cryptocurrency telah tumbuh pesat, dengan lebih dari 10.000 aset crypto yang beredar di seluruh dunia, dan jumlah pengguna meningkat dari 18 juta pada tahun 2017 menjadi 516 juta pada tahun 2023 [3], [4].

Bitcoin, sebagai mata uang kripto paling dominan, memegang pangsa pasar sebesar 60,14% dan menjadi subjek perhatian luas di kalangan investor dan masyarakat umum [5]. Dengan kapitalisasi pasar yang melampaui \$964,71 miliar, Bitcoin memainkan peran utama dalam ekosistem cryptocurrency [6]. Pemahaman yang lebih dalam tentang Bitcoin dan peran strategisnya dalam transformasi keuangan global menjadi semakin penting, termasuk kebutuhan pengembangan kerangka regulasi yang sesuai untuk melindungi pengguna dan memastikan stabilitas pasar [7].

Bitcoin mengalami fluktuasi harga yang signifikan, dipengaruhi oleh faktor ekonomi dan psikologis pasar [8]. Sentimen pasar yang tercermin dari berita dan opini di media massa, platform daring, serta jejaring sosial memiliki

pengaruh besar terhadap pergerakan harga Bitcoin [9]. Analisis sentimen terhadap berita Bitcoin, menggunakan metode klasifikasi k-nearest neighbor, dapat membantu memahami perilaku pasar dan meramalkan tren harga [10].

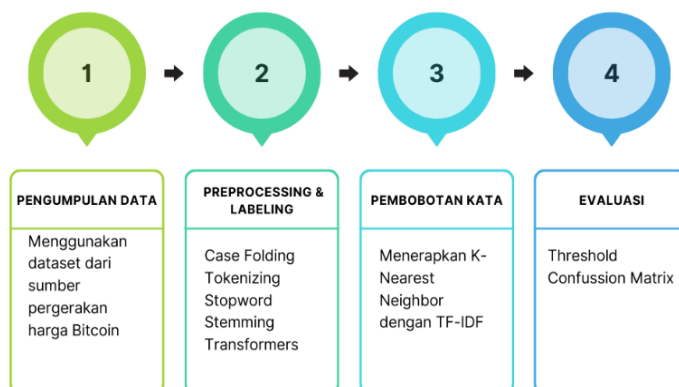
Penelitian ini memberikan kontribusi baru dalam analisis pasar cryptocurrency, khususnya Bitcoin, dengan memfokuskan pada analisis sentimen berita menggunakan metode klasifikasi k-nearest neighbor (KNN) untuk memprediksi pergerakan harga Bitcoin. Pendekatan ini mengintegrasikan data sentimen dari berita dan media sosial, yang belum banyak diterapkan dalam penelitian sebelumnya. Selain mengidentifikasi faktor ekonomi, penelitian ini juga memeriksa dampak psikologis pasar dari sentimen berita, memperluas pemahaman tentang pengaruh berita dan opini publik terhadap fluktuasi harga Bitcoin secara real-time. Penelitian ini bertujuan mengembangkan strategi perdagangan yang lebih adaptif terhadap perubahan sentimen pasar, memberikan wawasan aplikatif bagi trader dan investor. Penelitian ini bertujuan untuk menyelidiki berita terhadap pergerakan harga Bitcoin melalui analisis sentimen dengan metode klasifikasi k-nearest neighbor. Hasil penelitian ini memberikan wawasan mendalam tentang dinamika pasar mata uang kripto dan membantu dalam pengembangan strategi perdagangan yang adaptif terhadap perubahan sentimen pasar. Penelitian ini juga bertujuan untuk memberikan landasan bagi pengambilan keputusan yang lebih informatif bagi para pemangku kepentingan dalam ekosistem Bitcoin.

II. PENELITIAN YANG TERKAIT

Penelitian ini bertujuan untuk menganalisis sentimen dan pengaruh berita terhadap nilai Bitcoin menggunakan metode klasifikasi K-nearest neighbor (KNN). Deteksi analisis sentimen pada berita memiliki signifikansi penting dalam pengembangan teknologi yang memberikan informasi kepada masyarakat umum serta pemangku kepentingan ekonomi. Penelitian ini difokuskan pada teknik K-nearest neighbor, mengingat berbagai teknik analisis sentimen telah dikembangkan dalam beberapa tahun terakhir. Salah satu studi sebelumnya menggunakan metode regresi linier untuk memprediksi nilai tukar Rupiah terhadap Dolar AS, yang terbukti efektif dengan tingkat akurasi mencapai 95% dari hasil uji coba pada 100 data uji [11]. Berbeda dengan metode regresi linier, metode K-nearest neighbor sangat membantu dalam analisis sentimen terhadap cryptocurrency, seperti yang ditunjukkan oleh [12], yang menemukan bahwa penggunaan KNN efektif dalam klasifikasi sentimen terkait berita cryptocurrency. Investasi dalam cryptocurrency, khususnya Bitcoin, semakin populer karena potensi keuntungan finansial yang signifikan [9]. Penilaian kinerja investasi sering kali menggunakan variabel seperti return, risiko dan Jensen's alpha. Penelitian oleh [13] menunjukkan bahwa Bitcoin memiliki kinerja tertinggi dalam hal return, risiko, Sharpe ratio, dan Treynor ratio dibandingkan dengan indeks saham LQ45 dan emas, meskipun indeks saham LQ45 memiliki kinerja terbaik dalam pengukuran Jensen's alpha. Sentimen merupakan hasil dari interaksi antara individu yang mengungkapkan opini atau pesan dengan penerima pesan [14]. Faktor-faktor yang mempengaruhi sentimen, seperti dalam konteks pelecehan cyber dan penguntitan cyber di kalangan mahasiswa, diselidiki secara mendalam. Sentimen negatif semakin meluas, terutama terhadap tokoh terkenal dengan banyak pengikut di media sosial [15], serta para selebgram yang sering menjadi sasaran komentar negatif [16]. Pentingnya pesan yang kuat dalam mendukung sistem deteksi sentimen yang efektif diperhatikan, dengan penggunaan kata-kata yang dapat memperkuat sentimen [17]. Sentimen juga bervariasi dalam tingkat intensitasnya [18], dan kemiripan kata atau teks yang ada dalam bukti turut mempengaruhi bobot sentimen positif dan negatif [19].

III. METODE PENELITIAN

Penelitian ini menggunakan data sentimen berita terhadap pergerakan harga Bitcoin. Proses penelitian meliputi tahapan pengumpulan data, pra-pemrosesan data, pemberian bobot kata, proses klasifikasi, dan akhirnya evaluasi. Langkah-langkah detail penelitian tersaji pada Gambar 1.



Gambar. 1 Alur Penelitian

A. Pengumpulan Data

Dalam penelitian ini, dataset menggunakan data primer yang bersumber dari :

- 1) <https://www.investing.com>
- 2) <https://www.idx.co.id>
- 3) <https://en.wikipedia.org/wiki/2023#References>

Dataset ini terdiri dari 5 atribut dan 1 atribut class yang relevan dengan penelitian kami, dan informasinya ditampilkan pada tabel I.

TABEL 1
ATRIBUT DATA BERITA TERHADAP PREDIKSI HARGA BITCOIN

1	Berita
2	Tanggal
3	Negara
4	Sumber
5	Label

B. Preprocessing & Labeling

Tahap preprocessing dan labeling dalam penelitian bertujuan menghilangkan noise pada data dan memberikan label positif atau negatif, sehingga dapat mempengaruhi hasil klasifikasi yang lebih akurat. Tahap ini sangat penting agar proses klasifikasi data dapat dilakukan dengan baik dan efisien. Menurut [20], tahapan preprocessing terdiri dari beberapa langkah, yaitu:

- 1) Case Folding: Ini adalah langkah untuk melakukan perubahan semua kalimat menjadi huruf kecil, sehingga tidak ada perbedaan dalam penulisan kata yang bisa mempengaruhi proses klasifikasi [21].
- 2) Tokenizing: Langkah ini mengubah bentuk kalimat menjadi beberapa kata dalam kalimat dan menghilangkan tanda baca dalam kalimat. Hal ini dilakukan untuk mempersiapkan data menjadi unit-unit yang lebih kecil dan relevan dalam proses klasifikasi [22].
- 3) Stopword Removal: Proses ini melibatkan penghilangan kata-kata esensial dalam sebuah kalimat yang biasanya tidak memiliki makna khusus dan sering muncul dalam bahasa. Kata-kata ini terdaftar dalam kamus stopword list dan dianggap tidak relevan untuk proses klasifikasi [23].
- 4) Stemming: Proses stemming menghilangkan afiks atau akhiran dari kata sehingga hanya menyisakan kata dasar. Hal ini membantu dalam mengurangi dimensi data dan mempercepat proses klasifikasi [24].
- 5) Transformers adalah arsitektur pembelajaran mesin yang menggunakan mekanisme perhatian untuk memahami konteks dan hubungan antar kata dalam teks. Model ini sangat efektif untuk berbagai tugas NLP, seperti terjemahan bahasa, klasifikasi teks, dan analisis sentimen. Dengan memanfaatkan mekanisme perhatian yang memungkinkan model mempertimbangkan seluruh konteks kalimat secara simultan, transformers mampu menghasilkan performa yang tinggi dalam berbagai aplikasi NLP. Meskipun memiliki keunggulan seperti efisiensi dan akurasi yang tinggi (kategori positif), model ini juga memiliki kekurangan seperti kebutuhan sumber daya yang besar dan kompleksitas dalam implementasi (kategori negatif), yang memerlukan banyak sumber daya untuk pelatihan dan implementasi.

Preprocessing memiliki beberapa langkah dalam analisis teks. Namun, ada juga langkah lain yang dikenal sebagai lemmatization. Namun, dalam penelitian ini, digunakan metode stemming sebagai alternatif. Hal ini dikarenakan dataset yang digunakan berbentuk bahasa Inggris, dan untuk melakukan stemming, digunakan library bernama Natural Language Processing (NLP). Adapun lemmatization sulit diimplementasikan karena kesulitan dalam menemukan library yang sesuai dengan bahasa Inggris. Oleh karena itu, keputusan untuk menggunakan stemming sebagai metode preprocessing dalam penelitian ini dianggap sebagai pilihan yang tepat.

C. Pembobotan Kata

Output dari tahap preprocessing menghasilkan kumpulan term atau kata. Langkah berikutnya adalah melakukan proses term weighting, di mana setiap term atau kata diberikan bobot atau nilai untuk menunjukkan pentingnya term tersebut dalam dokumen [25]. Perhitungan Idf menggunakan rumus (1).

$$Idf = \log \left(\frac{\text{jumlah seluruh dokumen dalam koleksi}}{\text{jumlah dokumen yang mengandung istilah}} \right) \quad (1)$$

D. K-Nearest Neighbors (KNN)

Algoritma K-NN digunakan untuk melakukan pengukuran antar tetangga adalah dengan menggunakan metode Euclidean Distance. Euclidean Distance digunakan untuk mengevaluasi seberapa dekat atau jauhnya dua objek [26], dengan rumus sebagai berikut :

$$\sqrt{\sum_{i=1}^m (X_{ik} - X_{jk})^2} \quad (2)$$

Di mana,

- 1) X_{ik} adalah nilai X pada data latihan,
- 2) X_{jk} adalah nilai Y pada data tes,
- 3) m adalah jumlah atribut.

Nilai yang dihasilkan dari rumus tersebut menunjukkan tingkat kesamaan antara kedua objek. Dalam konteks ini, objek merujuk pada data latihan dan data tes. Pemilihan nilai k yang optimal bergantung pada ukuran dataset. Sebuah nilai k yang besar tidak selalu menjadi yang terbaik, demikian juga sebaliknya.

E. Evaluasi

Evaluasi merupakan tahap penting dalam penelitian ini untuk menilai efektivitas dan akurasi kinerja dari model yang telah dibuat. Evaluasi dilakukan berdasarkan tiga metrik penting, yaitu Accuracy, Precision, dan Recall. Nilai-nilai ini dapat dinyatakan dalam bentuk persentase (1-100%) atau dalam rentang bilangan 0 hingga 1.

Rumus (3) adalah persamaan untuk menghitung nilai akurasi, sementara rumus (4) adalah persamaan untuk menghitung nilai presisi, dan persamaan untuk nilai recall tersedia pada rumus (5) dari pengelompokan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

Penjelasan: True Positive (TP) adalah ketika prediksi kategori dan kategori sistem cocok dengan tepat, True Negative (TN) adalah ketika prediksi kategori dan kategori sistem tidak cocok dengan tepat, False Positive (FP) adalah ketika prediksi kategori dan kategori sistem terlihat cocok tapi sebenarnya tidak, dan False Negative (FN) adalah ketika prediksi kategori dan kategori sistem tidak cocok dan benar-benar tidak cocok.

F. Simulasi Kasus

Pada Gambar 2, disajikan sebuah simulasi kasus yang menggunakan berita untuk memprediksi pergerakan harga Bitcoin dengan menggunakan metode KNN dalam melakukan klasifikasi sentimen berita. Dalam simulasi ini, diasumsikan bahwa berita telah diperoleh dan kemudian dianalisis, setelah itu sentimen tersebut diselidiki melalui investigasi siber untuk memberikan informasi kepada masyarakat umum dan pelaku ekonomi khususnya.



Gambar. 2 Simulasi Kasus

Pada simulasi kasus di Gambar 2, dilakukan analisis sentimen berita. Proses investigasi cyber akan dimulai setelah berita diperoleh dan dianalisis. Tahap awal adalah mengidentifikasi berita yang dianggap berpengaruh terhadap pergerakan harga Bitcoin. Selanjutnya, akan dikumpulkan berita dari berbagai

sumber berita global, dengan mencatat informasi penting seperti berita, tanggal penerbitan, negara, dan sumber berita. Semua berita yang diperoleh akan disimpan dengan teliti, kemudian dilakukan proses analisis yang mendalam. Laporan analisis ini akan menjadi bukti yang kuat yang dapat digunakan sebagai informasi bagi masyarakat umum dan pelaku ekonomi. Hasil dari analisis ini juga dapat memberikan wawasan kepada para pelaku Bitcoin dan berkontribusi dalam mencegah kerugian di masa depan.

IV. HASIL DAN PEMBAHASAN

A. Data Set

Dalam penelitian ini, dataset yang digunakan berasal dari berita yang dianggap berpengaruh terhadap pergerakan harga Bitcoin. Data ini dikumpulkan dari oktober 2022 hingga april 2024 dan telah dilabeli dengan menggunakan python dengan library yaitu huggingface Dataset tersebut terdiri dari 350 record dengan 3511 kata, dan setiap record sudah dilabeli atau kategori. Terdapat 5 atribut dalam dataset, namun hanya atribut berita, tanggal dan negara yang relevan untuk kepentingan penelitian ini. Tabel 2 berisi metadata dari dataset yang digunakan dalam penelitian ini.

TABEL 2
METADATA BERITA

Id	Berita	Negara
1	Human error to blame for deadly train crash, says Greek PM	Yunani
2	Demonstrators, police clash as Greece train crash protests resume	Yunani
3	Greek election: Centre-right Mitsotakis hails big win but wants majority	Yunani
4	Senior judge to lead Greek caretaker government until fresh June election	Yunani
5	Death toll in Greece refugee boat tragedy climbs to 59	Yunani
6	Greek elections: Mitsotakis hails conservative win as mandate for reform	Yunani
up to
350	U.S. surgeons perform world's first whole eye transplant	Amerika Serikat

B. Preprocessing & Labeling

Pada penelitian ini, dilakukan perhitungan terhadap 350 record yang mengandung 3511 kata. berita yang terkumpul kemudian menjalani tahap preprocessing dan labeling untuk mempercepat dan mempermudah klasifikasi data. Terdapat beberapa tahapan labeling dan preprocessing yang dilakukan, yaitu:

1) Casefolding

Hasil perubahan pada tahap casefolding telah direpresentasikan dalam Tabel 3. Pada Tabel tersebut, Kolom pertama berisi id, kolom kedua berisi berita, dan kolom ketiga berisi negara. Perbedaan antara Tabel 2 dan Tabel 3 adalah pada perubahan huruf kapital di awal kata yang berubah menjadi huruf kecil. Sebagai contoh, kata "Human" berubah menjadi "human". Hasil dari proses case folding ini digunakan untuk mengubah berita menjadi bentuk kata yang kecil sehingga memudahkan dalam penyusunan teks.

TABEL 3
CASEFOLDING

Id	Berita	Negara
1	human error to blame for deadly train crash says greek pm	Yunani
2	demonstrators police clash as greece train crash protests resume	Yunani
3	greek election centreright mitsotakis hails big win but wants majority	Yunani

4	senior judge to lead greek caretaker government until fresh june election	Yunani
5	death toll in greece refugee boat tragedy climbs to	Yunani
6	greek elections mitsotakis hails conservative win as mandate for reform	Yunani
up to
350	us surgeons perform worlds first whole eye transplant	Amerika Serikat

2) Tokenizing

Hasil pada tahap tokenizing telah direpresentasikan dalam Tabel 4. Pada Tabel tersebut, terdapat kolom yang memuat berita, perbedaan yang terdapat antara Tabel 3 dan Tabel 4 adalah pada perubahan tanda koma (",") yang digunakan sebagai pemisah kata dalam setiap kalimat. Proses ini akan mempermudah dalam penghilangan kata-kata yang tidak esensial yang terdapat dalam beirta tersebut.

TABEL 4
TOKENIZING

Id	Berita	Negara
1	'human', 'error', 'to', 'blame', 'for', 'deadly', 'train', 'crash', 'says', 'greek', 'pm'	Yunani
2	'demonstrators', 'police', 'clash', 'as', 'greece', 'train', 'crash', 'protests', 'resume'	Yunani
3	'greek', 'election', 'centreright', 'mitsotakis', 'hails', 'big', 'win', 'but', 'wants', 'majority'	Yunani
4	'senior', 'judge', 'to', 'lead', 'greek', 'caretaker', 'government', 'until', 'fresh', 'june', 'election'	Yunani
5	'death', 'toll', 'in', 'greece', 'refugee', 'boat', 'tragedy', 'climbs', 'to'	Yunani
6	'greek', 'elections', 'mitsotakis', 'hails', 'conservative', 'win', 'as', 'mandate', 'for', 'reform'	Yunani
up to
350	'us', 'surgeons', 'perform', 'worlds', 'first', 'whole', 'eye', 'transplant'	Amerika Serikat

3) Stopword

Hasil perubahan yang dilakukan pada tahap stopword telah direpresentasikan dalam Tabel 5. Perbedaan antara Tabel 4 dan Tabel 5 adalah pada adanya kata-kata seperti “to”, ”for”, ”but”, ”until”, ”in”, ” as”, “for”, ”and”, ”out” , dan seterusnya. Kata-kata tersebut telah dihapus dari kalimat berita.. Dengan demikian, hanya kata-kata yang dianggap penting dan relevan yang akan digunakan dalam tahap selanjutnya.

TABEL 5
STOPWORD

Id	Berita	Negara
1	human error blame deadly train crash says greek pm	Yunani
2	demonstrators police clash greece train crash protests resume	Yunani
3	greek election centreright mitsotakis hails big win wants majority	Yunani
4	senior judge lead greek caretaker government fresh june election	Yunani
5	death toll greece refugee boat tragedy climbs	Yunani
6	greek elections mitsotakis hails conservative win mandate reform	Yunani
up to
350	us surgeons perform worlds first whole eye transplant	Amerika Serikat

4) Stemming

Tahap akhir dalam preprocessing adalah stemming, yang menggunakan library NLTK Porter Stemmer, seperti yang terlihat pada Tabel 6. Perbedaan antara Tabel 5 dan Tabel 6 adalah pada kata-kata seperti kata "deadly" menjadi "deadli", "says" menjadi "say", "demonstrators" menjadi "demonstr", "police" menjadi "polic", "greece" menjadi "greec", "protests" menjadi "protest", "resume" menjadi "resum", dan seterusnya. Kata-kata tersebut telah diubah dalam kalimat berita. Proses stemming bertujuan untuk menghilangkan imbuhan dan membuang akhiran kata sehingga diperoleh bentuk kata dasar yang sesuai. Dengan demikian, hasil stemming membuat semua kata yang memiliki akar kata yang sama menjadi seragam dan memudahkan dalam proses klasifikasi dan analisis data karena menggunakan bentuk kata dasar yang konsisten.

TABEL 6
STEMMING

Id	Berita	Negara
1	human error blame deadli train crash say greek pm	Yunani
2	demonstr polic clash greec train crash protest resum	Yunani
3	greek elect centreright mitsotaki hail big win but want major	Yunani
4	senior judg lead greek caretak govern fresh june elect	Yunani
5	death toll greec refuge boat tragedi climb	Yunani
6	greek elect mitsotaki hail conserv win mandat reform	Yunani
up to
350	us surgeon perform world first whole eye transplant	Amerika Serikat

5) Transformers

Hasil dari transformers sendiri yaitu berupa kategori positif ataupun kategori negatif. Proses ini menggunakan library <https://huggingface.co/distilbert/distilbert-base-uncased-finetuned-sst-2-english>. berikut merupakan metadata hasil dari pelabelan pada tabel 7.

TABEL 7
TRANFORMERS

No	Berita	Label
1	human error blame deadli train crash say greek pm	Negative
2	demonstr polic clash greec train crash protest resum	Negative
3	greek elect centreright mitsotaki hail big win but want major	Positive
4	senior judg lead greek caretak govern fresh june elect	Positive
5	death toll greec refuge boat tragedi climb	Negative
6	greek elect mitsotaki hail conserv win mandat reform	Positive
up to
350	us surgeon perform world first whole eye transplant	Positive

Tahap preprocessing dan labeling telah berhasil menghilangkan noise dan mengubah kata-kata dalam dokumen menjadi bentuk yang lebih standar dan konsisten dan memberikan label positif dan negatif pada berita. Sebagai hasilnya, dokumen yang semula berisi lebih banyak kata dapat dikompres yang lebih terstruktur dan siap untuk dilakukan analisis lebih lanjut menggunakan metode klasifikasi yang diusulkan.

Dengan data yang telah melewati tahap preprocessing dan labeling, proses klasifikasi dan evaluasi dapat dilakukan untuk mengidentifikasi adanya sentimen berita terhadap pergerakan harga Bitcoin dengan lebih efisien dan akurat.

C. Pembobotan Kata

Setelah melalui tahapan preprocessing untuk 350 data, term diubah menjadi data vektor melalui perkalian Tf*Idf, menghasilkan total 3511 term atau kata. Proses ini bertujuan untuk memberikan bobot atau nilai yang menggambarkan pentingnya setiap term atau kata terhadap dokumen dalam dataset yang digunakan. Dengan demikian, akan mempermudah dalam proses klasifikasi dan identifikasi sentimen berita terhadap pergerakan harga Bitcoin.

TABEL 8
HASIL TERM FREQUENCY DAN INVERSE DOCUMENT FREQUENCY

Id	acti s	admit	after	dead	s / d	zealand
1	0	0	0	0	.	0
2	0	0	0	0	.	0
3	0	0	0.271	0	.	0
4	0.330	0	0	0.286	.	0
5	0	0	0	0.193	.	0
6	0	0	0	0	.	0
7	0	0	0	0	.	0
8	0	0	0	0	.	0
9	0	0	0	0	.	0
10	0	0	0	0	.	0.398
s/d	0
368	0	0	0	0	.	0

D. K-Nearest Neighbors (KNN)

Setelah tahapan preprocessing dan labeling selesai serta perkalian Tf*Idf dilakukan untuk mengubah term menjadi vektor data, langkah selanjutnya yaitu pengujian dengan algoritma KNN pada analisis sentimen Berita dalam penelitian ini. Penelitian ini dilakukan sepuluh kali pengujian dengan nilai K yang berbeda, yaitu 1-10. Tujuannya untuk mendalami model klasifikasi mengklasifikasikan sentimen berita sebagai berpengaruh atau tidak terhadap pergerakan harga Bitcoin berdasarkan nilai K yang berbeda.

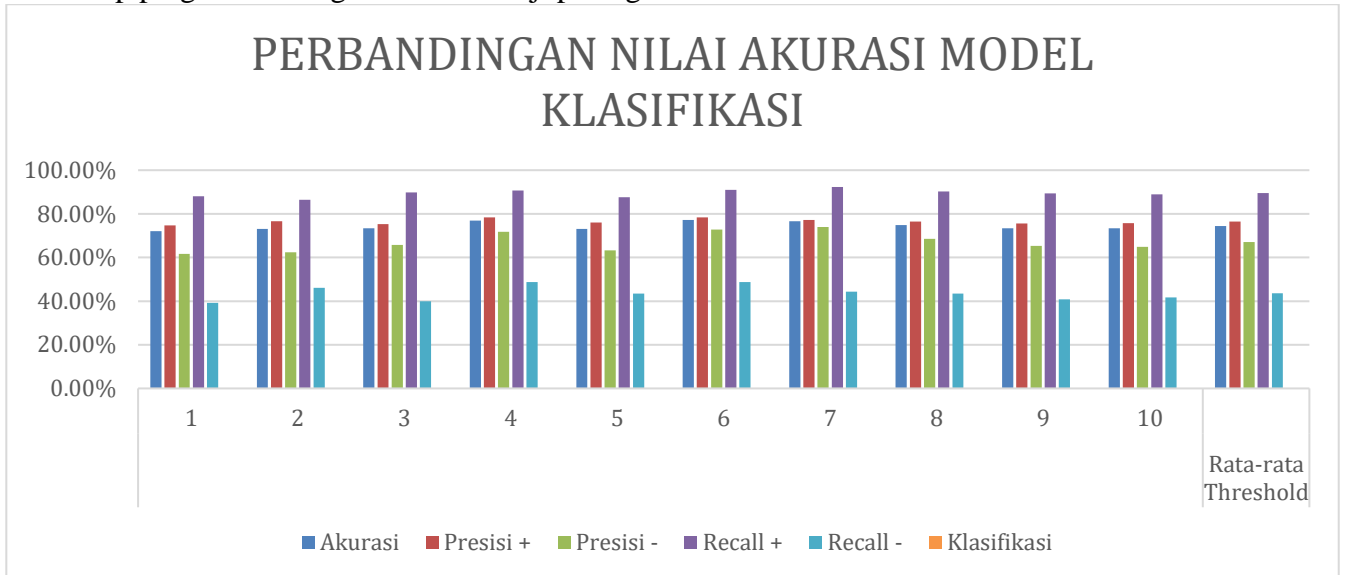
Hasil pengujian menggunakan K-Nearest Neighbors untuk klasifikasi sentimen Berita dengan nilai K pada pengujian pertama yaitu 1 menunjukkan akurasi sebesar 72.00%. Tingkat ketepatan presisi untuk opini positif adalah 74,73%, sementara untuk opini negatif adalah 61,64%. Selain itu, tingkat recallnya yaitu 88,09% untuk opini positif dan 39,13% untuk opini negatif. Evaluasi ini penting untuk mengidentifikasi potensi perbaikan model dan meningkatkan efektivitas klasifikasi sentimen Berita terhadap pergerakan harga Bitcoin.

	true POSITIVE	true NEGATIV E	class precision
pred. POSITIVE	207	70	74.73%
pred. NEGATIV E	28	45	61.64%
class recall	88.09%	39.13%	

Gambar. 3 Pengujian model k-nearest neighbors, dengan pengujian pertama

E. Evaluasi

Hasil pengujian menunjukkan variasi akurasi pada tiap model klasifikasi tergantung pada nilai threshold yang digunakan, dari rentang 0.1 hingga 1.0. Tabel 11 membandingkan hasil dari 10 kali pengujian dengan nilai akurasi pada metode K-Nearest Neighbor untuk setiap nilai threshold. Dari tabel tersebut, dapat diamati bahwa akurasi dari kedua metode bervariasi, dan penggunaan nilai threshold yang berbeda dapat mempengaruhi kinerja model klasifikasi. Evaluasi ini memberikan wawasan tentang performa dan efektivitas kedua metode dalam mengklasifikasikan sentimen berita sebagai berpengaruh atau tidak terhadap pergerakan harga Bitcoin tersaji pada gambar 4.



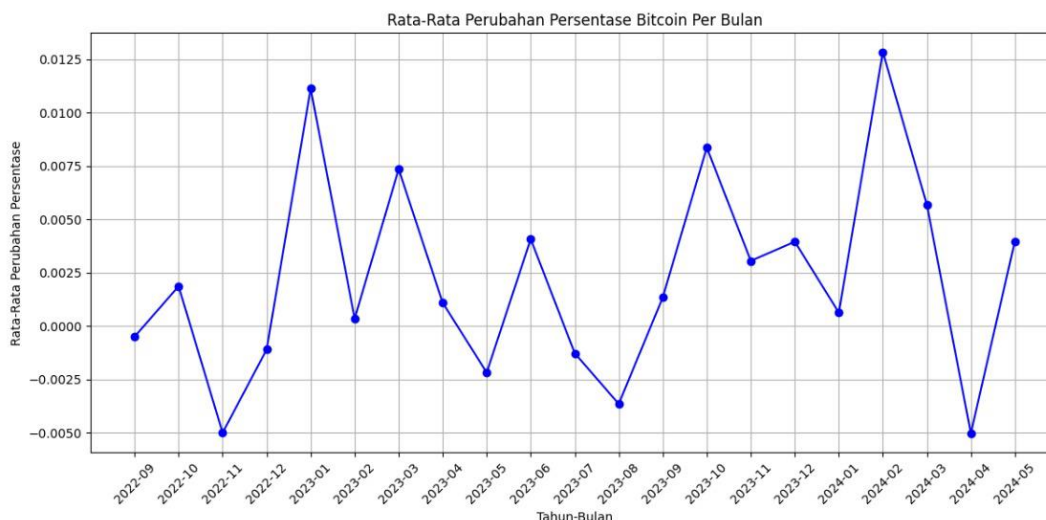
Gambar 4. Perbandingan nilai akurasi model klasifikasi

Evaluasi pengukuran untuk memastikan bahwa model klasifikasi yang telah dibuat berkinerja dengan cukup baik dikarenakan nilai rata-rata akurasi, precision dan recall untuk metode klasifikasi tersebut. Hasil evaluasi pengukuran untuk metode K-Nearest Neighbor menunjukkan bahwa rata-rata akurasi sebesar 74,40%, sedangkan precision untuk kelas opini positif adalah 76,44%, untuk kelas opini negatif adalah 67,01%, sedangkan recallnya kelas positif sebesar 89,45% dan kelas negatif adalah 43,65%.

K-Nearest Neighbor cukup bagus dalam menganalisis sentiment berita sebagai berpengaruh atau tidak terhadap pergerakan harga Bitcoin. Metode K-Nearest Neighbor mencapai rata-rata akurasi sebesar 74,40%, hasil ini menunjukkan bahwa pengklasifikasian menggunakan metode ini efektif. Selain itu, nilai rata-rata precision score untuk kelas opini positif menunjukkan tingkat klasifikasi yang lumayan tinggi dengan nilai 76,44%, sedangkan precision score untuk kelas opini negatif mencapai nilai 67,01%. Hal ini menandakan bahwa metode K-Nearest Neighbor cukup baik dianjurkan dalam analisis sentimen berita terhadap pergerakan harga Bitcoin. Selain memiliki tingkat klasifikasi yang tinggi untuk kedua kelas, metode ini juga memiliki sifat sederhana namun memiliki performa yang kuat secara keseluruhan.

Id ↑	label	prediction(la...	text	Tanggal
1	NEGATIVE	NEGATIVE	human error blame deadli train crash greek	Mar 01, 2023
10	NEGATIVE	NEGATIVE	stamped yemen ramadan chariti event kill lea...	Apr 20, 2023
100	NEGATIVE	NEGATIVE	niger soldier declar coup nation	Jul 26, 2023
101	POSITIVE	NEGATIVE	niger heart sahel crisi	Sep 18, 2023
102	NEGATIVE	NEGATIVE	gunmen abduct student northwestern nigeria...	Mar 07, 2024
103	NEGATIVE	NEGATIVE	kuriga kidnap more than nigerian pupil releas	Mar 24, 2024
104	NEGATIVE	NEGATIVE	acinda ardem resign reaction from around w...	Jan 19, 2023
105	POSITIVE	NEGATIVE	zealand chri hipkin take over from jacinda ard...	Jan 22, 2023
106	POSITIVE	NEGATIVE	zealand elect conserv christoph luxon premie...	Oct 14, 2023

Gambar. 5 Contoh id diindikasikan sebagai sentimen berita terhadap pergerakan harga Bitcoin.



Gambar. 6 Grafik bitcoin tahun 2022 hingga 2024.

Hasil klasifikasi yang diperoleh ditunjukkan pada Gambar 5. Terdapat lima kolom yaitu kolom pertama menampilkan ID, kedua menampilkan label, ketiga menampilkan prediksi klasifikasi, keempat menampilkan teks atau berita tentang Bitcoin, dan kelima menampilkan tanggal. Sebagai contoh, data pada gambar menunjukkan adanya berita yang masuk dalam kategori negatif. Hasil penelitian menunjukkan bahwa berita dengan ID 100 terindikasi mengandung sentimen negatif pada tanggal juli 2024, pada bulan tersebut itu dapat dilihat pada gambar 6 itu menunjukkan bahwa harga bitcoin juga mengalami penurunan. Informasi ini dapat digunakan sebagai acuan untuk melihat nilai pasar dari harga Bitcoin pada tanggal, bulan, dan tahun tersebut serta memantau pergerakan harga Bitcoin pada gambar 6. Pada gambar 6 Grafik pergerakan Bitcoin menunjukkan fluktuasi yang signifikan dari waktu ke waktu. Perubahan harga secara berkala tercermin dalam pola naik-turun yang terlihat jelas dalam grafik tersebut. Analisis sentimen berita mengungkapkan bahwa pergerakan harga Bitcoin dipengaruhi oleh sentimen positif dan negatif yang berkembang dalam komunitas crypto. Terdapat korelasi yang kuat antara perubahan sentimen dan pergerakan harga, dengan penurunan dan kenaikan harga yang seringkali terjadi setelah munculnya sentimen positif dan negatif pada bulan tertentu. Ketika sentimen positif mendominasi, harga Bitcoin cenderung mengalami kenaikan yang signifikan, menciptakan peluang bagi para investor untuk mendapatkan keuntungan. Namun, di sisi lain, ketika sentimen negatif merajalela, harga Bitcoin dapat mengalami penurunan tajam, menimbulkan kekhawatiran di antara para pemegang aset kripto. Dengan memahami pola pergerakan harga Bitcoin dan memonitor sentimen berita secara cermat, para investor dapat mengambil keputusan yang lebih terinformasi dan mengantisipasi perubahan pasar yang potensial. Hal ini menegaskan pentingnya analisis sentimen dalam strategi perdagangan dan investasi di pasar kripto yang dinamis.

V. KESIMPULAN

Dalam penelitian ini, telah dilakukan perbandingan pengujian menggunakan 10-fold cross-validation pada metode klasifikasi untuk analisis beritas terhadap pergerakan harga Bitcoin. Tahap pra-pemrosesan dan pelabelan data dilakukan secara seragam pada seluruh data. Setelah tahap pra-pemrosesan dan pelabelan, data diproses dengan KNN untuk membandingkan performa 10-fold cross-validation dan menemukan rata-rata akurasi. Hasil analisis menunjukkan bahwa Metode KNN mencapai rata-rata akurasi 74,40%, hasil ini menunjukkan bahwa pengklasifikasian menggunakan metode ini efektif. Selain itu, nilai rata-rata precision score untuk kelas opini positif menunjukkan tingkat klasifikasi yang lumayan tinggi dengan nilai 76,44%, sedangkan precision score untuk kelas opini negatif mencapai nilai 67,01%. Hal ini menandakan bahwa metode K-Nearest Neighbor cukup baik dianjurkan dalam analisis sentimen berita terhadap pergerakan harga Bitcoin konteks penelitian ini.

Selain itu, dalam mengidentifikasi beritas terhadap pergerakan harga Bitcoin, sebagai contoh, beritas dengan ID 100 terindikasi mengandung sentimen negatif pada tanggal juli 2024, pada bulan tersebut itu menunjukkan bahwa harga bitcoin juga mengalami penurunan.. Informasi ini dapat memberikan wawasan yang berguna dalam analisis tren dan dinamika pasar Bitcoin.

DAFTAR PUSTAKA

- [1] A. Wijayanto, I. Riadi, Y. Prayudi, and T. Sudinugraha, "Network Forensics Against Address Resolution Protocol Spoofing Attacks Using Trigger , Acquire , Analysis , Report , Action Method," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 8, no. 2, pp. 156–169, 2022, [Online]. Available: <http://www.journal.unipdu.ac.id/index.php/register/article/view/2953/1497>
- [2] S. Alam, M. Jamil, and A. Syamsir, "Digital Currency in Indonesia (Prospects and Challenges in Inclusive Financial Reviews)," *J. Ad'ministrare*, vol. 9, no. 2, p. 515, 2022, doi: 10.26858/ja.v9i2.39498.
- [3] R. Iyer, "New Evidence on Spillovers Between Crypto Assets and Financial Markets," *IMF Work. Pap.*, vol. 2023, no. 213, p. 1, 2023, doi: 10.5089/9798400256622.001.
- [4] R. Shewale, "blockchain statistics," *demandsage*, 2024.
- [5] O. Sattarov, H. S. Jeon, R. Oh, and J. D. Lee, "Forecasting bitcoin price fluctuation by twitter sentiment analysis," *2020 Int. Conf. Inf. Sci. Commun. Technol. ICISCT 2020*, vol. 02, no. 1, pp. 1–12, 2020, doi: 10.1109/ICISCT50599.2020.9351527.
- [6] X. Xu, B. Zhang, W. Lv, and Weiwei, "Dynamic Crypto Currency Transaction Index Modelling Using Blockchain and Community Detection Method," in *2021 3rd International Conference on Intelligent Control, Measurement and Signal Processing and Intelligent Oil Field (ICMSP)*, 2021, pp. 234–238. doi: 10.1109/ICMSP53480.2021.9513352.
- [7] L. Liestyowati, E. Sudarmanto, H. Ramadhani, S. Rijal, and T. W. Nurdiani, "Tren Investasi Aset Digital: Studi tentang Perilaku Investor Muda terhadap Cryptocurrency di Tengah Perubahan Pasar Keuangan di Kota Bandung," *J. Akunt. Dan Keuang. West Sci.*, vol. 2, no. 03, pp. 142–149, 2023, doi: 10.58812/jakws.v2i03.639.
- [8] F. Kjaerland, M. Meland, A. Oust, and V. Øyen, "How can Bitcoin Price Fluctuations be Explained?," *Int. J. Econ. Financ. Issues*, vol. 8, no. 3, pp. 323–332, 2018.
- [9] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 267–281, 2022.
- [10] E. T. Yolanda, L. Junaedi, and A. Bimo Gumelar, "Analisis Sentimen Pergerakan Harga Mata Uang Kripto (Cryptocurrency) Menggunakan TextBlob-NLTK (Natural Language Toolkit)," *Jlk*, vol. 5, no. 2, pp. 44–50, 2022.
- [11] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
- [12] A. O. Indarso, H. N. Irmanda, and R. Astriratma, "Sentiment Analysis of Cryptocurrency Exchange Application on Twitter Using Naïve Bayes Classifier Method," *Telemat. J. Inform. dan Teknol. Inf.*, vol. 20, no. 1, p. 15, 2023, doi: 10.31315/telematika.v20i1.9044.
- [13] G. I. Mumtazah and Y. T. Permadhy, "ANALISIS KINERJA REKSA DANA SAHAM DENGAN METODE SHARPE, TREYNOR, DAN JENSEN SELAMA MASA PANDEMI," *J. Ris. Manaj. Sains Indones.*, vol. 13, no. 1, pp. 53–75, 2022, doi: <http://doi.org/10.21009/JRMSI>.
- [14] W. M. Al-rahmi, N. Yahaya, M. M. Alamri, and N. Ali, "A Model of Factors Affecting Cyber Bullying Behaviors among University Students," *IEEE Access*, vol. PP, no. November, p. 1, 2018, doi: 10.1109/ACCESS.2018.2881292.
- [15] Yoannes Romando, R. Sulistyowati, and I. S. Wibisono, "Identifikasi Komentar Negatif Berbahasa Indonesia Pada Instagram Dengan K-Means," *Multimatrix*, vol. II, no. 1, pp. 6–8, 2019.
- [16] M. Andriansyah *et al.*, "Cyberbullying comment classification on Indonesian Selebgram using support vector machine method," *Proc. 2nd Int. Conf. Informatics Comput. ICIC 2017*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/IAC.2017.8280617.
- [17] F. Guo, X. Zhao, B. Zou, and Y. Liang, "Automatic retinal image registration using blood vessel segmentation and SIFT feature," *Int. J. Pattern ...*, 2017, doi: 10.1142/S0218001417570063.
- [18] I. Riadi, Sunardi, and P. Widiandana, "Cyberbullying Detection on Instant Messaging Services Using Rocchio and Digital Forensics Research Workshop Framework," *J. Eng. Sci. Technol.*, vol. 17, no. 2, pp. 1408–1421, 2022.
- [19] P. Widiandana and I. Riadi, "ANALISIS INVESTIGASI FORENSIK CYBERBULLYING PADA WHATSAPP MESSENGER MENGGUNAKAN METODE NATIONAL INSTITUTE of STANDARDS and TECHNOLOGY (NIST)," pp. 488–493, 2019.
- [20] RSA, "2016: Current State of Cybercrime," p. 7, 2016.

- [21] H. S. Christopher D. Manning, Prabhakar Raghavan, *Introduction to Modern Information Retrieval*. 2009.
- [22] I. Ruthven and M. Lalmas, "A survey on the use of relevance feedback for information access systems," *Knowl. Eng. Rev.*, vol. 18, no. 2, pp. 95–145, 2003, doi: 10.1017/S0269888903000638.
- [23] S. K. Sahu, S. Sarangi, and S. K. Jena, "A detail analysis on intrusion detection datasets," *Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014*, pp. 1348–1353, 2014, doi: 10.1109/IAdCC.2014.6779523.
- [24] E. W. Selberg, "Information Retrieval Advances using Relevance Feedback," *UW Dept CSE Gen. Exam*, 1997.
- [25] P. Yugianus, H. S. Dachlan, and R. N. Hasanah, "Pengembangan Sistem Penelusuran Katalog Perpustakaan Dengan Metode Rocchio Relevance Feedback," *J. EECCIS*, vol. 7, no. 1, pp. 47–52, 2013.
- [26] M. Y. R. Rangkuti, M. V. Alfansyuri, and W. Gunawan, "Penerapan Algoritma K-Nearest Neighbor (Knn) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia," *Hexag. J. Tek. dan Sains*, vol. 2, no. 2, pp. 11–16, 2021, doi: 10.36761/hexagon.v2i2.1082.