Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 3 (2025) pp: 2823-2831
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Analisis Sentimen Opini Publik Terhadap Peristiwa Bitcoin Halving Pada Data Teks Twitter Menggunakan Metode Naïve Bayes Dan Pembobotan Fitur TF-IDF

Andi Nur Halim[1], Rudiman[2], Nauval Azmi Verdikha[3]
[1,2,3]Universitas Muhammadiyah Kalimantan Timur
2011102441038@umkt.ac.id[1], rudiman@umkt.ac.id[2], nav651@umkt.ac.id[3]

Abstrak

Penelitian ini bertujuan menganalisis sentimen opini publik terhadap peristiwa Bitcoin Halving pada data teks Twitter menggunakan Naïve Bayes Classifier dan pembobotan fitur TF-IDF. Latar belakang penelitian ini adalah pesatnya pertumbuhan media sosial sebagai sumber data opini publik yang dinamis, khususnya terkait peristiwa finansial seperti Bitcoin Halving. Pendekatan kuantitatif dengan metode deskriptif analitis digunakan untuk mengklasifikasikan sentimen. Populasi penelitian adalah seluruh tweet yang berkaitan dengan topik tersebut, dan sampelnya berjumlah 538 tweet setelah melalui proses *crawling* dan *preprocessing*. Instrumen yang digunakan adalah bahasa pemrograman Python dan *library* tweet-harvest. Hasil penelitian menunjukkan bahwa model Naïve Bayes efektif, dengan akurasi tertinggi sebesar 74% pada rasio pembagian data 80:20. Kesimpulan dari penelitian ini adalah bahwa kombinasi metode tersebut mampu memberikan wawasan berharga mengenai sentimen pasar secara *real-time*.

Kata Kunci: *Analisis Sentimen, Bitcoin Halving, Naïve Bayes Classifier, TF-IDF, Twitter*

1. Latar Belakang

Kemajuan pesat teknologi informasi telah memicu pertumbuhan eksponensial media sosial, menciptakan fenomena yang sering disebut "banjir data" (Julianto et al., 2022). Di antara berbagai platform, Twitter menjadi salah satu yang paling dominan, berfungsi sebagai sumber data real-time yang kaya akan opini publik (Asmara et al., 2020). Data ini seringkali mencerminkan topik yang sedang hangat diperbincangkan, seperti peristiwa ekonomi, politik, atau teknologi. Seiring dengan popularitas mata uang kripto (cryptocurrency), pembahasan mengenai Bitcoin, yang merupakan pionir di bidang ini, juga menjadi sorotan. Bitcoin, sebagai mata uang digital terdesentralisasi pertama, beroperasi tanpa pengawasan lembaga keuangan (Handaya et al., 2020), menjadikannya topik yang sangat dinamis dan sensitif terhadap sentimen publik. Salah satu peristiwa yang paling dinantikan dan sering menjadi topik tren di Twitter adalah Bitcoin Halving. Peristiwa ini, yang terjadi setiap empat tahun, secara historis telah dikaitkan dengan fluktuasi pasar yang signifikan dan memicu spekulasi luas (Meynkhard, 2019). Oleh karena itu, menganalisis sentimen dari data Twitter terhadap Bitcoin Halving menjadi sangat relevan untuk memahami dinamika pasar dan opini investor.

Peristiwa Bitcoin Halving secara historis telah terbukti menjadi katalisator bagi pergerakan harga, yang dapat dijelaskan melalui hukum dasar ekonomi permintaan dan penawaran (Meynkhard, 2019). Ketika imbalan untuk penambang berkurang, pasokan Bitcoin baru pun melambat, meningkatkan kelangkaannya dan berpotensi mendorong kenaikan harga (M. H. Z. K. Ramadhani, 2022). Tingginya volatilitas dan fluktuasi harga yang cepat membuat topik ini menjadi daya tarik utama bagi para investor dan spekulan. Akibatnya, Twitter dibanjiri oleh komentar, prediksi, dan spekulasi yang sangat bervariasi. Menganalisis ribuan tweet ini secara manual untuk menentukan sentimennya —apakah positif, negatif, atau netral— adalah tugas yang sangat sulit dan tidak efisien. Diperlukan sebuah pendekatan otomatis untuk mengolah data teks dalam jumlah besar, yang dalam konteks ini dikenal sebagai analisis sentimen atau sentiment analysis (Julianto et al., 2022). Analisis sentimen, sebagai salah satu cabang dari text mining, memungkinkan ekstraksi penilaian dan opini dari data teks untuk mengklasifikasikannya ke dalam kategori yang telah ditentukan.

Sejumlah penelitian terdahulu telah menunjukkan keefektifan berbagai metode dalam klasifikasi sentimen. Misalnya, penelitian yang dilakukan oleh Srividya & Mary Sowjanya (2019) membandingkan model Naive Bayes dan Support Vector Machine (SVM) untuk analisis sentimen berbasis aspek, menunjukkan bahwa penggunaan TF-IDF sebagai pembobotan fitur memberikan hasil yang lebih unggul. Sementara itu, Firasari et al. (2020) menemukan bahwa Naive Bayes mengungguli K-Nearest Neighbor (K-NN) dalam klasifikasi sentimen, dengan akurasi yang lebih tinggi. Penelitian lain oleh Yuliana et al. (2022) juga memperkuat temuan ini, menunjukkan bahwa metode klasifikasi probabilitas seperti Naive Bayes memiliki performa yang kompetitif dan efisien untuk dataset teks yang tidak terlalu kompleks. Namun, masih ada keterbatasan dalam penelitian yang secara spesifik berfokus pada topik yang sangat dinamis seperti Bitcoin Halving, di mana sentimen dapat berubah dengan cepat. Adrian & Ikhsan (2024) telah mengaplikasikan TF-IDF dan Naive Bayes untuk deteksi berita palsu, membuktikan relevansi metode ini dalam konteks data teks yang rentan terhadap bias dan opini.

Celah penelitian (research gap) yang teridentifikasi adalah kurangnya studi yang secara khusus menganalisis sentimen opini publik di Twitter terhadap peristiwa Bitcoin Halving dengan kombinasi metode Naïve Bayes Classifier dan pembobotan fitur TF-IDF (Term Frequency-Inverse Document Frequency) pada data teks Bahasa Indonesia yang terkini. Meskipun metode ini telah digunakan secara luas, aplikasinya pada topik yang spesifik dan real-time seperti Bitcoin Halving masih dapat dieksplorasi lebih dalam untuk memahami dinamika sentimen pasar. Kombinasi Naive Bayes yang sederhana dan efisien dengan TF-IDF yang efektif dalam menentukan bobot kata penting diharapkan mampu memberikan hasil yang optimal. Br Sinulingga & Sitorus (2024) telah membuktikan bahwa TF-IDF sangat efektif dalam mengubah data teks menjadi representasi numerik yang relevan untuk klasifikasi, sementara Berliani & Lestari (2024) menegaskan bahwa Naive Bayes membutuhkan data latih yang relatif sedikit, menjadikannya pilihan yang ideal untuk dataset yang terbatas.

Berdasarkan latar belakang dan permasalahan yang telah diuraikan, penelitian ini memiliki tujuan untuk menganalisis sentimen opini publik terhadap peristiwa Bitcoin Halving menggunakan kombinasi metode Naïve Bayes Classifier dan pembobotan fitur TF-IDF. Urgensi penelitian ini terletak pada kemampuannya untuk menyediakan wawasan berharga mengenai sentimen pasar yang real-time, yang dapat dimanfaatkan oleh investor, analis pasar, atau peneliti untuk memprediksi tren dan membuat keputusan yang lebih informatif. Kebaruan (novelty) dari penelitian ini adalah penerapan metode yang telah teruji pada topik yang sangat spesifik dan terkini, yaitu Bitcoin Halving 2024, serta memberikan analisis yang mendalam terhadap kinerja model klasifikasi dengan variasi rasio data latih dan data uji, yang jarang dieksplorasi secara komprehensif dalam konteks serupa. Diharapkan, hasil dari penelitian ini dapat menjadi referensi yang kuat dan memberikan kontribusi nyata bagi bidang analisis sentimen, khususnya pada topik-topik finansial yang terdesentralisasi.

2. Metode Penelitian

2.1 Jenis dan Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif analitis, yang bertujuan untuk menganalisis dan mendeskripsikan secara sistematis proses serta hasil klasifikasi sentimen opini publik. Metode ini dipilih karena memungkinkan pengolahan data teks dalam jumlah besar untuk diukur dan diinterpretasikan secara objektif (Sugiyono, 2021). Fokus utama penelitian ini adalah penerapan metode text mining untuk analisis sentimen, di mana data teks dari media sosial diubah menjadi informasi terstruktur yang dapat dianalisis. Analisis sentimen sendiri merupakan metode yang sangat relevan dalam disiplin ilmu informatika untuk memahami kecenderungan penilaian terhadap suatu objek secara otomatis (Julianto et al., 2022). Penelitian ini secara spesifik mengaplikasikan algoritma Naïve Bayes Classifier yang dikenal efisien dan efektif untuk klasifikasi teks, terutama saat dikombinasikan dengan teknik pembobotan fitur (Imelda & Arief Ramdhan Kurnianto, 2023).

2.2 Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh opini publik yang disebarkan melalui tweet di platform Twitter yang berkaitan dengan peristiwa Bitcoin Halving. Mengingat populasi yang sangat besar dan dinamis, penelitian ini menggunakan teknik pengambilan sampel. Sampel penelitian diambil dengan metode crawling, yaitu proses pengumpulan data otomatis dari Twitter menggunakan kata kunci "Bitcoin Halving". Selama proses crawling ini, data yang diperoleh berjumlah 558 tweet. Namun, setelah melalui tahap pra-pemrosesan data (pre-processing) untuk menghilangkan duplikasi dan konten tidak relevan, jumlah sampel akhir yang digunakan dalam analisis adalah 538 tweet. Pemilihan sampel ini sudah memenuhi kriteria untuk mendapatkan representasi yang memadai

DOI: https://doi.org/ 10.31004/riggs.v4i3.2291

dari opini yang beredar (Sudaryono, 2022). Sesuai dengan kaidah metodologi penelitian, penentuan sampel yang representatif sangat krusial untuk memastikan hasil penelitian dapat digeneralisasi (Emzir, 2021).

2.3 Instrumen dan Teknik Analisis Data

Instrumen yang digunakan dalam penelitian ini adalah bahasa pemrograman Python dengan bantuan pustaka (library) seperti tweet-harvest yang diintegrasikan dalam lingkungan Google Colab. Pustaka ini berfungsi sebagai alat utama untuk melakukan crawling data dari Twitter. Teknik analisis data yang diterapkan mengikuti serangkaian tahapan yang terstruktur. Tahapan ini dimulai dengan pra-pemrosesan data (pre-processing) untuk membersihkan data mentah dari noise seperti tautan URL, tagar, dan simbol, serta mengubahnya menjadi format yang siap untuk dianalisis (Barus, 2022). Selanjutnya, dilakukan ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk memberikan bobot numerik pada setiap kata, yang mencerminkan tingkat kepentingannya dalam dataset (Br Sinulingga & Sitorus, 2024). Setelah itu, data yang telah diproses dan dibobotkan diklasifikasikan menggunakan algoritma Naïve Bayes Classifier. Kinerja model dievaluasi menggunakan metrik akurasi yang dihitung dari confusion matrix, sebuah teknik standar untuk menilai performa model klasifikasi (Adrian & Ikhsan, 2024).

2.4 Prosedur Penelitian

Prosedur penelitian ini dirancang secara sistematis melalui beberapa tahapan yang logis, yang dapat digambarkan sebagai berikut:

Pengumpulan Data: Data tweet dikumpulkan secara otomatis dari Twitter menggunakan teknik crawling dengan kata kunci "Bitcoin Halving lang:id" untuk memastikan data berbahasa Indonesia. Sebanyak 558 tweet berhasil dikumpulkan pada awal Mei 2024.

Pelabelan Data (Labeling): Seluruh data tweet yang telah dikumpulkan dilabeli secara manual oleh seorang ahli (expert) ke dalam dua kategori sentimen: positif atau negatif. Kriteria pelabelan didasarkan pada makna, konotasi, dan ekspresi yang terkandung dalam tweet.

Pra-pemrosesan Data (Pre-processing): Data yang telah dilabeli melalui serangkaian tahapan: case folding (mengubah semua teks menjadi huruf kecil), cleansing (menghilangkan karakter tidak relevan seperti URL, tagar, angka, dan simbol), tokenizing (memecah kalimat menjadi kata), stopword removal (menghilangkan kata-kata umum yang tidak memiliki makna penting seperti "dan," "ini," "itu"), stemming (mengembalikan kata ke bentuk dasarnya), dan penghapusan data duplikat untuk menjaga kualitas data (S. Ramadhani et al., 2022).

Pembobotan Fitur (TF-IDF): Setelah pra-pemrosesan, kata-kata dalam tweet diubah menjadi representasi numerik menggunakan pembobotan TF-IDF. Proses ini melibatkan perhitungan Term Frequency (TF), Inverse Document Frequency (IDF), dan akhirnya nilai TF-IDF untuk setiap kata.

Pembagian Data: Dataset dibagi menjadi data latih (training data) dan data uji (testing data) dengan tiga rasio yang berbeda: 90:10, 80:20, dan 70:30. Pembagian ini bertujuan untuk menguji bagaimana proporsi data mempengaruhi kinerja dan akurasi model klasifikasi.

Klasifikasi dan Evaluasi: Data latih digunakan untuk melatih model Naïve Bayes Classifier. Kemudian, data uji digunakan untuk memprediksi sentimen, dan hasilnya dievaluasi menggunakan confusion matrix dan metrik akurasi untuk menentukan performa model terbaik dari ketiga rasio yang telah diuji.

DOI: https://doi.org/ 10.31004/riggs.v4i3.2291

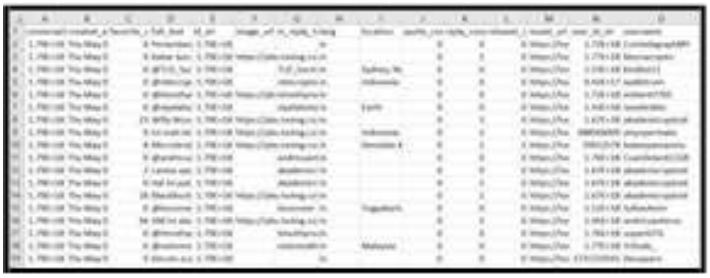
Prosedur penelitian ini dirangkum dalam diagram alir berikut:

Gambar 1 Alur Penelitian

3. Hasil dan Diskusi

3.1 Input Dataset

Penelitian ini dimulai dengan pengumpulan data dari platform Twitter. Data yang diambil berkaitan dengan topik "Bitcoin Halving" yang terjadi pada bulan April 2024. Teknik yang digunakan untuk mengumpulkan data adalah *crawling* dengan bantuan *library* tweet-harvest yang dioperasikan melalui Python dan Google Colab. Proses ini berhasil mengumpulkan total 558 tweet. Data mentah tersebut memiliki 15 kolom, termasuk full_text, yang merupakan atribut utama dan berisi teks komentar tweet. Setelah melalui tahapan pra-pemrosesan, jumlah data yang digunakan dalam analisis akhir berkurang menjadi 538 tweet. Data ini kemudian disimpan dalam format CSV.



Gambar 2. Dataset

3.2 Labelling

Dataset kemudian melalui tahap pemberian label secara manual oleh ahli bahasa (*expert*). Dengan mengkategorikan tweet menjadi dua kategori yaitu positif dan negatif berdasarkan sentimen. Tabel 2 adalah sampel dataset yang telah dilabeli sebagai berikut :

Tabel 1. Labelling

Full_text	label
@TLD_Survive kalau saya dikasih 100juta dalam kondisi ini uang nganggur ya. saya all in beli bitcoin	Positif
saat sentiment lagi fear mungkin beberapa bulan kedepan saat harga sedang dibottomnya tunggu 2025 atau bitcoin scam. jgn beli. ponzi	Negatif

3.3 Pre-Processing

Gambar 3 menunjukkan visualisasi kata dari dataset sebelum dilakukan preprocessing. Dalam gambar tersebut, banyak kata tidak relevan terlihat, seperti 'dan', 'ini', 'itu', 'yg', 'di', 'ke', 'gak'. Kata-kata ini sering muncul tetapi tidak memberikan informasi penting terkait topik yang dibahas. Selain itu, terdapat 'URL' serta berbagai kata dalam bentuk tidak baku dan masih memiliki imbuhan. Kondisi ini mencerminkan karakteristik data mentah yang belum diolah, di mana noise dan kata-kata tidak relevan masih terdapat dalam dataset. Keberadaan kata-kata yang tidak relevan ini menunjukkan pentingnya tahap *preprocessing* dalam pengolahan data teks untuk menghilangkan elemen-elemen yang tidak informatif, sehingga analisis sentimen dapat lebih

DOI: https://doi.org/ 10.31004/riggs.v4i3.2291

akurat dan representatif terhadap opini publik yang sebenarnya. Tanpa *preprocessing*, hasil analisis bisa menjadi bias dan kurang bermakna. Oleh karena itu, *preprocessing* merupakan langkah kritis dalam memastikan kualitas data sebelum memasuki tahap analisis lebih lanjut.

Gambar 3. Wordcloud Sebelum Pre-Processing

Selanjutnya pada gambar 4 menampilkan hasil visualisasi kata (*wordcloud*) dari dataset setelah dilakukan text *preprocessing*. Dalam visualisasi ini, kata-kata yang tidak relevan seperti 'dan', 'ini', 'itu', 'yg', 'di', 'ke', 'gak', serta 'URL' dan berbagai kata dalam bentuk tidak baku telah dihilangkan. Proses *preprocessing* yang melibatkan *case folding*, *cleansing*, *tokenizing*, *stopword removal*, *stemming* dan penghapusan data duplikat (*delete duplicate*) telah berhasil mengurangi noise dan meningkatkan relevansi kata-kata yang tampil. Hasil ini menunjukkan bahwa kata-kata yang lebih informatif dan berkaitan langsung dengan topik analisis, seperti 'peristiwa', 'harga', 'dunia', 'kripto', dan 'investor', kini mendominasi *wordcloud*, mencerminkan opini publik yang lebih akurat dan representatif terhadap peristiwa Bitcoin halving. *Preprocessing* ini merupakan langkah penting dalam memastikan kualitas data sehingga analisis sentimen yang dilakukan dapat memberikan hasil yang lebih valid dan bermakna.

Gambar 4. Wordcloud Sesudah Pre-Processing

3.5 TF-IDF

Untuk dapat diproses oleh mesin, kata-kata dari dataset perlu diubah menjadi representasi numerik. Proses ini menggunakan metode pembobotan TF-IDF (*Term Frequency – Inverse Document Frequency*).
TF-IDF adalah teknik yang memberikan bobot pada setiap kata untuk menentukan tingkat kepentingannya dalam sebuah dokumen. Prosesnya melibatkan tiga langkah utama:

1. Term Frequency (TF): Menghitung frekuensi kemunculan sebuah kata dalam satu dokumen.
2. Inverse Document Frequency (IDF): Mengukur seberapa unik atau penting sebuah kata di seluruh dokumen dalam dataset.
3. TF-IDF: Mengalikan nilai TF dan IDF untuk mendapatkan bobot akhir, yang mencerminkan pentingnya kata tersebut dalam konteks dataset.

Dengan demikian, nilai TF-IDF yang dihasilkan akan menjadi representasi numerik dari setiap kata, yang siap digunakan oleh model klasifikasi.

DOI: https://doi.org/ 10.31004/riggs.v4i3.2291

Tabel 2. Dataset Terlabeli

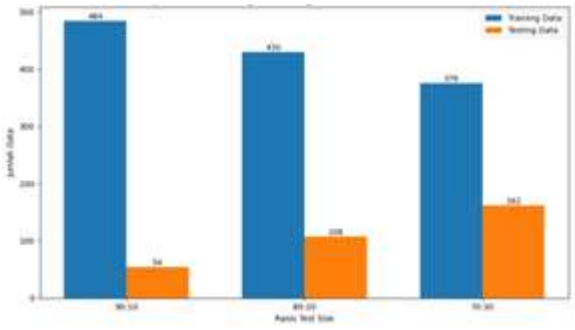
Array Position	*TF*	*IDF*	*TF-IDF*	*Term*
342	0.050000	6.633002	0.291165	Buruk
1088	0.050000	6.633002	0.291165	Kolaps
397	0.050000	6.227537	0.273367	Cinta
685	0.050000	6.227537	0.273367	Ftx
987	0.050000	6.227537	0.273367	kabar
464	0.050000	5.939855	0.260738	Dalam
1328	0.050000	5.939855	0.260738	Merosot
1686	0.050000	5.534390	0.242940	Prospek
1605	0.050000	5.380239	0.236173	Persen
332	0.050000	5.380239	0.236173	Bulan
2013	0.050000	5.023564	0.220516	Suku
337	0.050000	4.928254	0.216333	Bunga
1109	0.050000	4.687092	0.205747	Koreksi
521	0.050000	4.618099	0.202718	Digital
128	0.050000	3.892162	0.170852	Aser
116	0.050000	3.688563	0.161915	April
2185	0.050000	3.374906	0.148146	Turun
1117	0.050000	3.318816	0.145684	Kripto
803	0.050000	3.092043	0.135730	Harga
268	0.050000	1.415353	0.062129	bitcoin

Hasil perhitungan TF-IDF menunjukkan pentingnya sebuah kata dalam dataset. Kata-kata seperti "buruk" dan "kolaps" memiliki nilai IDF tertinggi (6.633002). Ini menandakan bahwa kedua kata tersebut sangat jarang muncul dalam dokumen, namun memiliki bobot tinggi karena penting untuk klasifikasi sentimen negatif. Nilai TF-IDF yang tinggi, yaitu 0.291165, menegaskan signifikansi kata-kata tersebut, meskipun frekuensi kemunculannya rendah. Kesimpulannya, kata-kata ini memiliki pengaruh besar dalam menentukan sentimen pada dokumen tersebut.

3.6 Split Data

Untuk mengevaluasi kinerja model klasifikasi secara komprehensif, dataset dibagi menjadi dua bagian: data latih dan data uji. Dalam penelitian ini, pembagian data dilakukan menggunakan tiga rasio yang berbeda: 90:10, 80:20, dan 70:30.

Tujuan dari pembagian ini adalah untuk menganalisis bagaimana perbedaan proporsi data latih dan data uji memengaruhi hasil analisis sentimen. Dengan menguji beberapa rasio, peneliti dapat mengidentifikasi rasio optimal yang mampu menyeimbangkan bias dan varians pada model, serta meningkatkan kemampuan model untuk melakukan generalisasi terhadap data baru.



Gambar 4. Split Data

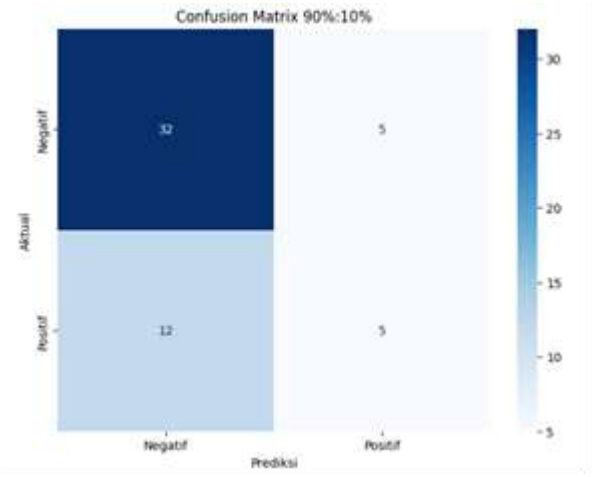
3.7 Klasifikasi

Hasil analisis model klasifikasi menggunakan Naïve Bayes divisualisasikan melalui *Confusion Matrix* untuk setiap rasio pembagian data. Berikut adalah ringkasan hasil klasifikasi untuk rasio 90:10, 80:20, dan 70:30:

1. **Rasio 90:10:** Hasil klasifikasi menunjukkan bahwa model ini mampu mengidentifikasi 32 *true negatives* dan 5 *true positives*, tetapi juga memiliki 12 *false negatives* dan 5 *false positives*.
2. **Rasio 80:20:** Pada rasio ini, terdapat peningkatan jumlah *true positives* menjadi 8, sementara *false negatives* juga meningkat menjadi 20.
3. **Rasio 70:30:** Pengujian dengan rasio ini menghasilkan 11 *true positives* dan 13 *false positives* yang relatif rendah, namun *false negatives* meningkat menjadi 30.

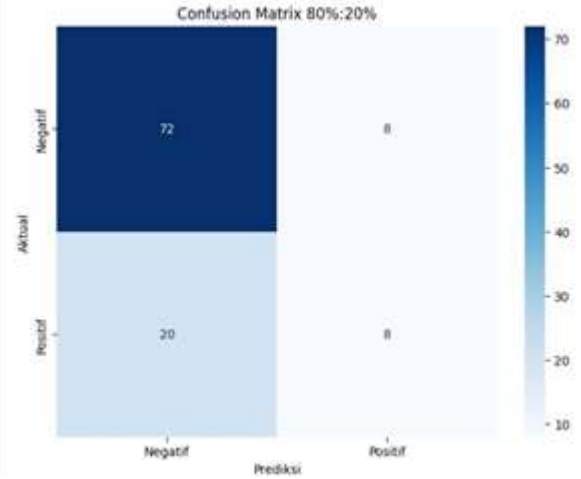
Secara keseluruhan, visualisasi melalui *Confusion Matrix* ini membantu mengevaluasi kemampuan model dalam mengklasifikasikan sentimen positif dan negatif untuk setiap rasio yang diuji.

DOI: https://doi.org/ 10.31004/riggs.v4i3.2291



Gambar 5. Confusion Matrix 90:10

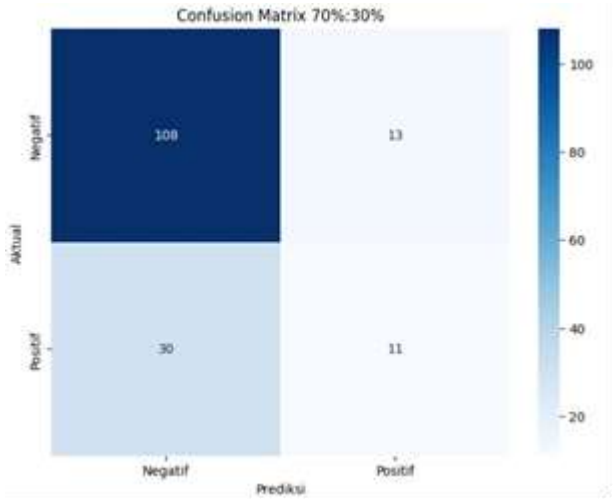
Berdasarkan hasil pengujian rasio 90:10, model Naive Bayes berhasil mengidentifikasi 32 true negatives dan 5 true positives. Namun, model juga memiliki 12 false negatives dan 5 false positives. Hasil ini menunjukkan bahwa model tersebut cenderung lebih efektif dalam mengklasifikasikan sentimen negatif, tetapi kemampuannya untuk mengenali sentimen positif masih belum optimal.



Gambar 6 Confusion Matrix 80:20

Berdasarkan pengujian dengan rasio 80:20, model Naive Bayes menunjukkan peningkatan dalam mengidentifikasi sentimen positif, dengan jumlah *true positives* naik menjadi 8. Namun, peningkatan ini juga disertai dengan kenaikan *false negatives* menjadi 20. Ini menunjukkan bahwa ketika jumlah data uji diperluas, model memiliki kemampuan yang lebih baik dalam mengklasifikasikan kasus positif dibandingkan pengujian sebelumnya.

DOI: https://doi.org/ 10.31004/riggs.v4i3.2291



Gambar 7. Confusion Matrix 70:30

Berdasarkan hasil pengujian dengan rasio 70:30, model berhasil meningkatkan jumlah true positives menjadi 11. Meskipun ada peningkatan false negatives menjadi 30, model menunjukkan jumlah false positives yang relatif rendah, yaitu 13. Hasil ini mengindikasikan bahwa penambahan data uji memiliki potensi untuk meningkatkan kemampuan model dalam mengidentifikasi sentimen positif dengan lebih baik.

3.7 Evaluasi

Evaluasi performa model Naïve Bayes menunjukkan hasil yang bervariasi berdasarkan tiga rasio pembagian data latih dan uji yang digunakan, yaitu 90:10, 80:20, dan 70:30. Hasil pengujian menunjukkan bahwa model memberikan performa yang cukup baik, dengan tingkat akurasi sebagai berikut:

1. **Rasio 90:10:** Akurasi model adalah **68,5%**.
2. **Rasio 80:20:** Akurasi model meningkat menjadi **74%**, yang merupakan akurasi tertinggi di antara ketiga rasio.
3. **Rasio 70:30:** Akurasi model sedikit menurun menjadi **73,4%**.

Perbedaan akurasi ini mengindikasikan bahwa kinerja model sangat dipengaruhi oleh proporsi data yang digunakan untuk pelatihan dan pengujian. Rasio 80:20 memberikan hasil terbaik karena proporsi data latih yang cukup besar (80%) memungkinkan model untuk mempelajari pola sentimen secara lebih komprehensif. Pada saat yang sama, data uji yang memadai (20%) memastikan evaluasi yang representatif terhadap kemampuan generalisasi model.

Sebaliknya, akurasi yang lebih rendah pada rasio 90:10 (68,5%) kemungkinan disebabkan oleh jumlah data uji yang terlalu sedikit, sehingga hasil evaluasi menjadi kurang representatif. Temuan ini menegaskan bahwa pemilihan rasio pembagian data yang tepat memiliki peran krusial dalam memaksimalkan kinerja model dan menghasilkan prediksi yang lebih akurat dan dapat diandalkan.

4. Kesimpulan

Berdasarkan analisis yang telah dilakukan, penelitian ini menyimpulkan bahwa kombinasi metode Naïve Bayes Classifier dan pembobotan fitur TF-IDF efektif untuk menganalisis sentimen publik terkait peristiwa Bitcoin Halving. Temuan utama menunjukkan bahwa kinerja model sangat dipengaruhi oleh proporsi pembagian data, dengan rasio 80:20 menghasilkan akurasi tertinggi sebesar 74%. Meskipun demikian, penelitian ini memiliki keterbatasan karena hanya berfokus pada dataset dari Twitter dan sentimen dalam konteks Bitcoin Halving 2024, yang dapat berubah dengan cepat. Oleh karena itu, disarankan untuk penelitian selanjutnya agar dapat mengaplikasikan metode serupa pada topik finansial dinamis lainnya, serta mengeksplorasi penggunaan rasio pembagian data yang lebih bervariasi atau menguji algoritma klasifikasi lain untuk membandingkan dan mengoptimalkan hasil.

DOI: https://doi.org/ 10.31004/riggs.v4i3.2291

Referensi

Adrian, M. I., & Ikhsan, R. M. (2024). *Asuhan keperawatan keluarga pada keluarga TN.A khususnya NY.Y dengan diabetes melitus tipe II tanpa luka*. Undergraduate thesis, Sriwijaya University.

Asmara, G., Handayani, R. T., & Utami, T. N. (2020). Faktor yang berhubungan dengan perilaku pengelolaan sampah pada followers Instagram Males.Nyampah. *Jurnal Universitas Padjadjaran*.

Barus, F. A. S. (2022). *Entitas masyarakat islam di Barus 1945-2022*. Skripsi, UIN Imam Bonjol.

Berliani, T., & Lestari, Y. E. (2024). Manajemen pembelajaran individual peserta didik berkebutuhan khusus. *Equity In Education Journal*, *6*(2), 53–60. https://doi.org/10.37304/eej.v6i2.14611

Br Sinulingga, J. R., & Sitorus, H. (2024). Analisis opini publik tentang kualitas pelayanan E-KTP di kantor camat Merdeka Kabupaten Karo. *Jurnal Social Opinion: Jurnal Ilmiah Ilmu Komunikasi*, *5*(2), 134–143.

Emzir. (2021). *Metodologi penelitian pendidikan kuantitatif dan kualitatif* (Edisi Revisi). Rajawali.

Firasari, E., & Anggriana, F. (2020). Analisis sentimen pada beasiswa menggunakan algoritma naïve bayes dan k-nearest neighbor. *Jurnal Riset dan Aplikasi Komputer*, *2*(1), 1-8.

Handaya, A., Hidayat, R. T., & Firmansyah. (2020). Analisis persepsi dan adopsi cryptocurrency di Indonesia. *Jurnal Bisnis dan Keuangan*, *7*(1), 15-28.

Imelda, I., & Kurnianto, A. R. (2023). Naïve Bayes and TF-IDF for sentiment analysis of the Covid-19 booster vaccine. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *7*(1), 1–6. https://doi.org/10.21108/ijoict.v8i2.681

Julianto, N., Suwandi, R., & Gunawan. (2022). Analisis sentimen pada twitter dengan metode naïve bayes classifier. *Jurnal Sains dan Teknologi*, *2*(1), 1-8.

Meynkhard, A. (2019). Fair market value of bitcoin: halving effect. *Investment Management and Financial Innovations*, *16*(4), 72–85. https://doi.org/10.21511/imfi.16(4).2019.06

Ramadhani, M. H. Z. K., Ulfah, Y., & Rinaldi, M. (2022). The impact of bitcoin halving day on stock market in Indonesia. *Jurnal International Conference Proceedings*, *5*(3), 127–137. https://doi.org/10.32535/jicp.v5i3.1800

Ramadhani, S., Azzahra, D., & Z, T. (2022). Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, *13*(1), 24–33. https://doi.org/10.31849/digitalzone.v13i1.9292

Srividya, K., & Mary Sowjanya, A. (2019). Aspect based sentiment analysis using POS tagging and TFIDF. *International Journal of Engineering and Advanced Technology*, *8*(6), 1960–1963. https://doi.org/10.35940/ijeat.F7935.088619

Sudaryono. (2022). *Metodologi penelitian: Pengenalan dasar dan aplikasinya*. Penerbit Andi.

Sugiyono. (2021). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Yuliana, T. M., Haryani, S., & Rosiadi, Y. S. (2022). Analisis sentimen ulasan film menggunakan naïve bayes. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, *3*(2), 99-106.

DOI: https://doi.org/ 10.31004/riggs.v4i3.2291