

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP FLUKTUASI HARGA ETHEREUM MENGGUNAKAN NLP DAN MACHINE LEARNING

Audi Rahma Firdasari, Sri Eniyati
Teknik Informatika, Universitas Stikubank Semarang
Jalan Kendeng No. 5 Semarang, Indonesia
audirahmaf@gmail.com

ABSTRAK

Perkembangan *cryptocurrency*, khususnya *ethereum* telah menarik perhatian banyak kalangan karena volatilitas harga yang tinggi karena dipengaruhi oleh faktor-faktor ekonomi dan sentimen pasar. Penelitian ini bertujuan untuk menganalisis hubungan antara sentimen pengguna twitter dengan fluktuasi harga *ethereum* menggunakan teknologi *Natural Language Processing (NLP)* dan *Machine Learning (ML)*. Metode yang digunakan mencakup pengumpulan data tweet tentang *ethereum*, pra-pemrosesan data, serta analisis sentimen menggunakan algoritma *Naïve Bayes*. Data harga *ethereum* diperoleh dari sumber informasi kripto. Hasil analisis menunjukkan bahwa meskipun terdapat fluktuasi antara sentimen positif dan negatif di twitter, korelasi antara sentimen publik dan pergerakan harga *ethereum* sangat lemah, dengan nilai koefisien korelasi yang rendah.

Kata kunci : Sentimen, Twitter, Ethereum, NLP, Machine Learning, Naïve Bayes, Fluktuasi Harga

1. PENDAHULUAN

Perkembangan aset kripto telah membawa perubahan besar terhadap sistem keuangan global. *Ethereum* merupakan salah satu aset digital yang berpengaruh, yang tidak hanya diperdagangkan sebagai instrumen investasi, tetapi juga dimanfaatkan sebagai platform untuk penerapan smart contract dan berbagai aplikasi terdesentralisasi. Harga *Ethereum* dikenal memiliki volatilitas yang tinggi, yaitu tingkat perubahan harga yang cepat dan fluktuatif dalam kurun waktu singkat. Faktor-faktor seperti permintaan pasar, kebijakan ekonomi, hingga tingkat adopsi teknologi berperan besar dalam menentukan arah pergerakan nilainya [1]. Karakteristik *ethereum* menjadikannya menarik untuk dikaji, khususnya dalam konteks analisis perilaku pasar dan psikologi investor.

Twitter merupakan media sosial yang kini menjadi salah satu sumber utama opini publik terkait aset kripto. Perasaan masyarakat seperti optimisme, ketakutan, atau sosialisasi terhadap kondisi pasar tercermin dalam unggahan tweet yang diunggah. Pola interaksi semacam ini sering kali menunjukkan kaitan dengan perubahan harga aset digital di pasar global. Beberapa penelitian terdahulu mengungkap bahwa sentimen yang muncul di media sosial dapat memengaruhi fluktuasi harga aset kripto tertentu [2][3]. Hal tersebut menjadikan analisis sentimen twitter sebagai pendekatan potensial untuk memahami dinamika harga *ethereum* secara lebih komprehensif.

Pemanfaatan teknologi *Natural Language Processing (NLP)* mempermudah analisis opini publik dalam jumlah besar secara otomatis. Dengan metode ini, data teks dari media sosial dapat diolah menjadi informasi kuantitatif yang menggambarkan arah sentimen pengguna. Melalui tahapan seperti pembersihan data, tokenisasi, normalisasi, dan ekstraksi fitur, data yang semula tidak terstruktur dapat

diubah menjadi bentuk yang dapat diproses oleh algoritma pembelajaran mesin. Beberapa metode *machine learning* seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, *Random Forest*, hingga model berbasis transformer seperti BERT dan FinBERT, terbukti meningkatkan akurasi analisis sentimen pada bidang keuangan [4].

Penelitian tentang hubungan antara sentimen publik dan harga aset kripto telah banyak dilakukan, namun sebagian besar masih berfokus pada Bitcoin. Penelitian terhadap *Ethereum* masih terbatas, padahal karakteristiknya berbeda baik dari sisi teknologi maupun perilaku pasar. Bahasa informal di Twitter, seperti singkatan, slang, dan penggunaan emoji, juga menimbulkan tantangan tersendiri dalam proses pra-pemrosesan dan pemilihan fitur, yang dapat mempengaruhi hasil klasifikasi sentimen [5]. Kondisi tersebut menunjukkan perlunya penelitian yang lebih mendalam dan terarah untuk memahami hubungan antara sentimen pengguna twitter dengan fluktuasi harga *ethereum* melalui pendekatan NLP dan *machine learning*.

Penelitian mengenai keterkaitan sentimen media sosial terhadap pergerakan harga *ethereum* memiliki manfaat praktis maupun akademis. Dari sisi praktis, hasilnya dapat menjadi dasar bagi investor dan pelaku pasar dalam merumuskan strategi analisis risiko dan pengambilan keputusan investasi. Dari sisi akademis, penelitian ini diharapkan memperkaya literatur mengenai *financial technology* serta memberikan kontribusi empiris tentang peran data media sosial dalam memprediksi pergerakan harga aset digital [1].

2. TINJAUAN PUSTAKA

Kemajuan teknologi informasi dan komunikasi telah memudahkan analisis data digital, termasuk data media sosial. Penelitian ini menggunakan teori-teori utama tentang *cryptocurrency*, analisis sentimen, *natural language processing (NLP)*, *machine learning*

(ML), dan hubungan sentimen dengan fluktuasi harga Ethereum. Berikut ini adalah teori-teori yang menjadi dasar konteks dan kerangka penelitian ini.

2.1. Cryptocurrency dan Ethereum

Cryptocurrency, yang merupakan jenis aset digital yang sangat bergantung pada teknologi blockchain, bertujuan untuk memastikan keamanan dan transparansi dalam setiap transaksi yang dilakukan [6]. Ethereum dikembangkan oleh Vitalik Buterin pada tahun 2015 sebagai platform terbuka berbasis blockchain yang memungkinkan implementasi *smart contract* dan aplikasi terdesentralisasi (*decentralized applications*).

Ethereum sangat penting dalam ekosistem aset digital karena mendukung berbagai aktivitas keuangan berbasis blockchain. Banyak faktor, termasuk tingkat permintaan pasar, persepsi publik, volume transaksi, dan perubahan aturan, mempengaruhi harga Ethereum [1]. Sentimen masyarakat di media sosial adalah indikator nonteknis yang sering dikaitkan dengan pergerakan harga karena menunjukkan tingkat kepercayaan investor terhadap aset kripto. Memahami faktor sentimen sangat penting untuk memahami bagaimana harga ethereum bergerak.

2.2. Analisis Sentimen

Analisis sentimen digunakan untuk menentukan orientasi emosional suatu teks positif, negatif, atau netral [7]. Pendekatan berbasis leksikon dan pendekatan berbasis pembelajaran mesin adalah dua pendekatan utama yang digunakan. Pendekatan pertama menggunakan daftar kata yang diberi nilai polaritas, sedangkan pendekatan kedua menggunakan mesin pembelajaran untuk menemukan pola linguistik yang terkait dengan emosi melalui data berlabel.

2.3. Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan bagian dari kecerdasan buatan yang membuat bahasa manusia dapat dipahami oleh komputer [8]. Dalam analisis sentimen, NLP mengubah data mentah menjadi representasi numerik yang dapat dipahami oleh algoritma machine learning. Natural Language Processing (NLP) mencakup berbagai proses, termasuk pengumpulan data, pembersihan teks, tokenisasi, stemming, lemmatisasi, dan penghapusan stopwords [9].

2.4. Machine Learning (ML)

Machine Learning (ML) memungkinkan sistem untuk membuat prediksi atau membuat keputusan dengan mempelajari pola data [10]. Dalam konteks analisis sentimen, algoritma *machine learning* digunakan untuk mengidentifikasi pola dalam teks dan mengklasifikasikan emosi atau perasaan berdasarkan fitur-fitur yang diperoleh dari pemrosesan *natural language processing*. Beberapa algoritma yang sering digunakan untuk klasifikasi sentimen antara lain *Naïve Bayes*, *Support Vector Machine (SVM)*, *Random*

Forest, *Logistic Regression*, serta *model-model deep learning* seperti *Long Short-Term Memory (LSTM)* dan *BERT*. *LSTM*, yang dirancang khusus untuk menangani hubungan antar kata dalam suatu kalimat melalui memori jangka panjang, memiliki kemampuan yang sangat baik dalam memahami konteks temporal [11]. Selain itu, *SVM* telah terbukti berguna untuk mengklasifikasikan data yang sangat besar dan kompleks dengan akurasi yang tinggi.

Naïve Bayes adalah salah satu algoritma klasifikasi teks yang paling banyak digunakan. Algoritma ini didasarkan pada teorema Bayes dan dapat menghitung probabilitas kelas atau kategori berdasarkan karakteristik data yang diketahui. Secara teoritis, *Naïve Bayes* mengasumsikan bahwa setiap karakteristik tidak bergantung pada karakteristik lainnya, dan mencoba memprediksi kelas atau kategori data berdasarkan konsep independensi ini. Terlepas dari asumsi sederhana tentang independensi karakteristik ini, *Naïve Bayes* sering memberikan hasil yang memuaskan dalam berbagai aplikasi, seperti analisis opini [12].

Secara matematis, Teorema Bayes digunakan untuk menghitung probabilitas suatu kelas C berdasarkan fitur X dengan rumus berikut:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Dimana:

- $P(C|X)$ adalah probabilitas bahwa data X termasuk dalam kelas C
- $P(X|C)$ adalah probabilitas melihat fitur X dalam kelas C
- $P(C)$ adalah probabilitas kelas C sebelum kita mengetahui fitur X
- $P(X)$ adalah probabilitas total dari fitur X

2.5. Hubungan Sentimen dengan Fluktuasi Harga Ethereum

Perilaku pasar kripto dijelaskan tidak hanya oleh faktor ekonomi rasional tetapi juga oleh faktor psikologis investor yang dikenal dalam teori *behavioral finance* [13]. Sentimen publik yang positif terhadap ethereum dapat meningkatkan kepercayaan investor sehingga mendorong aktivitas beli, sedangkan dominasi sentimen negatif dapat memperkuat tekanan jual dan menurunkan harga pasar.

Sentimen pasar memiliki pengaruh yang signifikan terhadap volatilitas harga ethereum di samping faktor ekonomi makro dan kebijakan moneter [1]. Hubungan ini memastikan bahwa opini publik di media sosial tidak hanya didasarkan pada informasi yang akurat, tetapi juga memengaruhi perilaku investasi. Analisis berbasis *natural language processing* dan *machine learning* dapat memetakan kecenderungan sentimen tersebut dan menghubungkannya dengan perubahan harga ethereum. Konsep ini menjadi dasar konseptual utama penelitian yang menelaah pengaruh ekspresi publik di Twitter terhadap fluktuasi harga aset digital menggunakan pendekatan kecerdasan buatan.

2.6. Penelitian Terdahulu

Penelitian sebelumnya telah memberikan referensi untuk pengembangan model analisis sentimen di bidang keuangan, dan panduan tersebut dapat diterapkan dalam penelitian ini. Aplikasi khusus untuk melakukan analisis sentimen pada artikel keuangan menggunakan model *FinBERT* yang telah dikembangkan. Model ini didasarkan pada arsitektur *BERT* dan dilatih berdasarkan basis data *Reuters Text Classification 2 (TRC2)*. Eksperimen yang dilakukan pada basis data *Financial Phrasebank* dan *FiQA Financial Opinion* menemukan akurasi sebesar 94,1%, kinerja yang lebih baik daripada basis referensi seperti *ELMo* dan *ULMFiT*. Temuan ini menunjukkan bahwa penerapan *FinBERT* di bidang keuangan dapat semakin meningkatkan akurasi analisis sentimen. Hal ini memberikan dasar yang kuat untuk pengembangan *NLP* untuk analisis dokumen keuangan.[4]

Penelitian lain menunjukkan bahwa sentimen yang diungkapkan oleh pengguna twitter memiliki korelasi yang kuat dengan pergerakan harga *cryptocurrency*. Dengan menggunakan model *NLP* dan *Deep Learning* yang digabungkan, hasil penelitian ini menegaskan bahwa penggunaan kedua pendekatan tersebut dalam analisis sentimen mampu memberikan hasil yang sangat akurat dalam memprediksi fluktuasi harga. Akurasi tinggi yang dicapai oleh model *FinBERT-GRU* menunjukkan bahwa kombinasi teknik tersebut dapat secara efektif memanfaatkan data sentimen yang berasal dari platform media sosial seperti twitter untuk analisis pasar *cryptocurrency*. [5]

Penelitian lain, menggunakan regresi linier dan regresi vektor otoregresif untuk menganalisis hubungan antara opini publik di Twitter dan pergerakan harga dua *cryptocurrency*. Model *VAR* digunakan untuk menerapkan metode penelitian kuantitatif. Data yang digunakan dalam studi ini diperoleh dari 90.000 tweet, yang dikumpulkan menggunakan *Scrapy Python* dan diklasifikasikan menggunakan *leksikon Loughran-McDonald*. Analisis menunjukkan bahwa sentimen keseluruhan terkait pergerakan harga ethereum di twitter bervariasi antara 6% dan 7%, dengan R^2 sebesar 6,76%. Selain itu, analisis kausal menunjukkan bahwa pergerakan harga memiliki pengaruh yang lebih besar terhadap perubahan sentimen keseluruhan daripada sentimen terhadap harga. Penelitian ini memberikan wawasan berharga tentang dinamika ulasan dan harga aset di platform sosial, mengungkapkan efektivitas perilaku mereka melalui kecerdasan kolektif yang dihasilkan dalam saluran dan jaringan pasar.[2]

Peneliti lain menganalisis faktor-faktor yang mempengaruhi volatilitas harga bitcoin dan ethereum, dengan fokus khusus pada kondisi ekonomi global dan kepercayaan pasar. Dengan menerapkan metode *ARCH/GARCH* dan analisis regresi sederhana pada data dari tahun 2018 hingga 2023, ditemukan bahwa tingkat volatilitas harga ethereum sebesar 17,2%. Ditemukan bahwa kepercayaan investor secara signifikan mempengaruhi volatilitas harga ethereum (β

$= 0,29$; $p < 0,01$), dengan model mencapai akurasi prediksi 89,7%. Menggunakan model *GARCH*, penulis mengkategorikan fluktuasi harga menjadi tiga kelompok: tinggi, sedang, dan rendah.[1]

3. METODE PENELITIAN

Data utama yang digunakan dalam penelitian ini berasal dari harga harian ethereum yang diambil dari situs penyedia informasi kripto serta unggahan atau tweet pengguna twitter yang membahas ethereum. Melalui proses pra-pemrosesan teks, analisis sentimen menggunakan teknik *natural language processing (NLP)* dan algoritma *machine learning (ML)*, data ini kemudian dihubungkan dengan pergerakan harga.

Penelitian ini bertujuan untuk mengungkap bagaimana sentimen masyarakat di twitter bisa berkaitan dengan naik-turunnya harga ethereum. Alur penelitian ini dibagi menjadi beberapa tahap yang terstruktur mulai dari analisis masalah, pengumpulan data, analisis dan evaluasi model.

3.1. Analisis Masalah

Twitter telah menjadi platform yang digunakan oleh berbagai lapisan masyarakat untuk menyampaikan opini dan pandangan mereka. Masyarakat sering memanfaatkan platform ini untuk mengekspresikan sentimen secara bebas, baik positif maupun negatif, terkait isu-isu terkini seperti pergerakan pasar aset kripto seperti ethereum. Dengan meningkatnya penggunaan media sosial, twitter menjadi sumber yang berharga untuk menganalisis sentimen publik. Dalam penelitian ini, fokus peneliti adalah sentimen masyarakat terhadap harga ethereum, yang dijelaskan melalui pengumpulan data tweet menggunakan teknik *web scraping*. Dengan menggunakan teknik ini, peneliti dapat mengambil tweet yang relevan untuk mengeksplorasi sentimen yang terkandung di dalamnya.

Metode yang digunakan untuk mengklasifikasikan tweet ke dalam dua kategori sentimen utama yaitu positif dan negatif adalah metode *Naïve Bayes*. Sebelum klasifikasi, data yang dikumpulkan melalui *scraping* harus melalui serangkaian proses pra-pemrosesan, seperti pembersihan data, tokenisasi, dan penghapusan stopwords. Proses ini bertujuan untuk menyiapkan data agar dapat diproses oleh algoritma dengan efektif. Setelah itu, tweet yang telah diproses untuk mengidentifikasi hubungan antara sentimen publik dan fluktuasi harga ethereum.

3.2. Pengumpulan Data

Proses pertama dalam penelitian ini adalah mengumpulkan data dari dua sumber utama yaitu data harga yang menunjukkan pergerakan harga Ethereum dikumpulkan melalui API seperti *CoinGecko*, dan data tweet yang menunjukkan pendapat publik tentang ethereum melalui proses *scraping* menggunakan *tweepy* di google colab dengan memasukkan kata kunci yang relevan seperti ethereum, ETH, #ethereum,

#eth, crypto, #crypto, cryptocurrency, #ETHprice, #cryptocurrency, blockchain, #blockchain, smart contract, DeFi, #DeFi, NFT, #NFT, altcoin, ETHprice, #ethereumcommunity. Agar analisis tren harga dan diskusi komunitas tetap terkini dan terstruktur gunakan rentang waktu dari 1 Januari 2025 hingga 30 Agustus 2025, sambil memastikan semua data dalam bahasa Inggris untuk menjangkau global yang lebih luas. Setelah itu, simpan hasilnya dalam file bernama eth.csv.

3.3. Analisis dan Evaluasi Model

Pada tahap ini, akan dijelaskan proses analisis data tweet dan evaluasi model dalam penelitian ini. Proses dimulai dengan tahap pra-pemrosesan data, diikuti oleh analisis sentimen, evaluasi model, dan analisis korelasi. Tahapan terakhir menandakan bahwa proses telah selesai. Setiap tahapan dalam diagram ini saling terhubung untuk memastikan kelancaran alur kerja yang sistematis dan terstruktur.

3.4. Pra-pemrosesan data

Pra-pemrosesan data adalah langkah awal yang sangat penting dalam analisis data untuk memastikan kualitas dan keberhasilan analisis yang akan dilakukan. Proses dimulai dengan pembersihan teks yang berfungsi mengubah semua huruf kapital menjadi huruf kecil, menghapus tanda baca, angka, simbol, dan karakter khusus. Diikuti oleh tokenisasi yang berfungsi untuk memecah teks menjadi kata atau token yang lebih kecil untuk menganalisis dan memahami emosi atau opini yang terkandung dalam teks. Selanjutnya, penghapusan *stopwords* yang berfungsi menghapus kata yang sering muncul tetapi tidak memiliki informasi penting. Dilanjutkan dengan stemming yang berfungsi untuk mengubah kata menjadi bentuk kata dasar dengan menghilangkan awalan dan akhiran agar mempermudah proses analisis. Tahapan terakhir menandakan bahwa proses telah selesai.

3.5. Analisis Sentimen

Analisis sentimen berfungsi untuk memahami perasaan pengguna twitter terhadap topik atau isu tertentu. Dengan menganalisis tweet yang berisi berbagai pendapat, analisis ini dapat mengidentifikasi apakah sentimen yang muncul bersifat positif, negatif, atau netral. Hal ini berguna untuk memantau opini publik secara *real-time*. Dengan teknologi ini, analisis dapat dilakukan dengan cepat dan efektif pada data dalam jumlah besar.

3.6. Evaluasi Model

Setelah analisis dilakukan, hasilnya akan dievaluasi dengan menggunakan berbagai metrik untuk memastikan validitas dan reliabilitas model yang digunakan. Untuk model sentimen, evaluasi dilakukan dengan mengukur akurasi, *precision*, *recall*, dan *f1-score* untuk menilai kemampuan model dalam mengklasifikasikan tweet dengan benar. Selain itu,

cross-validation dengan *10-fold* akan digunakan untuk mengurangi overfitting dan memastikan model dapat digeneralisasi dengan baik pada data yang belum terlihat. *Confusion matrix* juga akan digunakan untuk memvisualisasikan kesalahan klasifikasi dalam setiap kategori sentimen.

Rumus akurasi dihitung sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Rumus *precision* dihitung sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Rumus *recall* dihitung sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Rumus *f1-score* dihitung sebagai berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Untuk memvisualisasikan kesalahan klasifikasi, digunakan *confusion matrix*. Matriks ini menggambarkan bagaimana kesalahan klasifikasi terjadi antara dua kategori, yaitu positif dan negatif.

3.7. Analisis Korelasi

Setelah pra-pemrosesan, *Naïve Bayes* akan digunakan untuk mengklasifikasikan tweet ke dalam kategori sentimen positif, negatif, dan netral. Algoritma ini dipilih karena kemampuannya dalam menangani data teks yang besar dan tidak terstruktur dengan efisien. Pada saat yang sama, data harga ethereum akan dipersiapkan dengan menggunakan statistik deskriptif untuk memahami pola fluktuasi harga, yang meliputi perhitungan rata-rata harga, standar deviasi, dan analisis variabilitas harga.

3.8. Visualisasi Hasil

Pada tahap terakhir, visualisasi hasil analisis akan dilakukan untuk memperjelas dan mempermudah pemahaman mengenai temuan penelitian. Diagram garis akan digunakan untuk menggambarkan fluktuasi harga ethereum selama periode penelitian, dan menggambarkan distribusi sentimen harian. Visualisasi ini akan membantu mengidentifikasi pola hubungan antara sentimen dan harga, serta memberikan gambaran yang lebih jelas tentang bagaimana sentimen publik dapat memengaruhi harga ethereum.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan *tweepy* dengan kata kunci ethereum, ETH, #ethereum, #eth, crypto, #ETHprice, #crypto, cryptocurrency, #cryptocurrency, blockchain, smart contract, DeFi, #DeFi, NFT, #NFT, altcoin, ETHprice, #ethereumcommunity, #blockchain since:2025-01-01 until:2025-08-30 lang:en". Hasil data yang diperoleh sebanyak 1990 tweet, dan disimpan dalam format .csv. Selanjutnya, dilakukan tahap pra-pemrosesan dari hasil pengumpulan data.

4.2. Hasil Pra-Pemrosesan Data

Hasil pengumpulan data yang telah didapatkan selanjutnya difilter untuk menghilangkan *noise* menggunakan metode *pre-processing*. Metode *pre-processing* dibagi menjadi 4 tahap.

4.3. Pembersihan Teks

Hasil pengumpulan data yang difilter pada tahap pembersihan data untuk mengubah semua huruf kapital menjadi huruf kecil, menghapus tanda baca, angka, simbol, dan karakter khusus.

Tabel 1. Hasil Pembersihan Data

full_text	Pembersihan Data
@TheWhiteWhaleV2 Good to see the disclaimer but the real scam is that people need it in the first place. Why does crypto culture make founder launches token the default expectation? That's the actual problem.	good to see the disclaimer but the real scam is that people need it in the first place why does crypto culture make founder launches token the default expectation thats the actual problem

4.4. Tokenisasi

Tahapan selanjutnya adalah tahap tokenisasi untuk memecah data tweet menjadi kata atau token yang lebih kecil untuk menganalisis dan memahami emosi atau opini yang terkandung dalam teks.

Tabel 2. Hasil Tokenisasi

Pembersihan Data	Hasil Tokenisasi
good to see the disclaimer but the real scam is that people need it in the first place why does crypto culture make founder launches token the default expectation thats the actual problem	['good', 'to', 'see', 'the', 'disclaimer', 'but', 'the', 'real', 'scam', 'is', 'that', 'people', 'need', 'it', 'in', 'the', 'first', 'place', 'why', 'does', 'crypto', 'culture', 'make', 'founder', 'launches', 'token', 'the', 'default', 'expectation', 'thats', 'the', 'actual', 'problem']

4.5. Stopwords Removal

Tahapan selanjutnya adalah tahap *stopwords removal* untuk menghapus kata yang sering muncul tetapi tidak memiliki informasi penting menggunakan *library nltk*.

Tabel 3. Hasil Stopwords Removal

Hasil Tokenisasi	Hasil Stopwords Removal
['good', 'to', 'see', 'the', 'disclaimer', 'but', 'the', 'real', 'scam', 'is', 'that', 'people', 'need', 'it', 'in', 'the', 'first', 'place', 'why', 'does', 'crypto', 'culture', 'make', 'founder', 'launches', 'token', 'the', 'default', 'expectation', 'thats', 'the', 'actual', 'problem']	['good', 'see', 'disclaimer', 'real', 'scam', 'people', 'need', 'first', 'place', 'crypto', 'culture', 'make', 'founder', 'launches', 'token', 'default', 'expectation', 'thats', 'actual', 'problem']

4.6. Stemming

Tahapan selanjutnya adalah tahap *stemming* untuk mengubah kata menjadi bentuk kata dasar dengan menghilangkan awalan dan akhiran agar mempermudah proses analisis.

Tabel 4. Hasil Stemming

Hasil Stopwords Removal	Hasil Stemming
['good', 'see', 'disclaimer', 'real', 'scam', 'people', 'need', 'first', 'place', 'crypto', 'culture', 'make', 'founder', 'launches', 'token', 'default', 'expectation', 'thats', 'actual', 'problem']	good see disclaim real scam peopl need first place crypto cultur make founder launch token default expect that actual problem

4.7. Hasil Analisis Sentimen

Setelah hasil *preprocessing* diperoleh, *VADER SentimentIntensityAnalyzer* digunakan untuk menganalisis teks sentimen ke dalam dua kelas: negatif dan positif. Analisis ini didasarkan pada skor campuran *VADER*, yang berkisar dari -1, yang menunjukkan sentimen negatif, hingga +1, yang menunjukkan sentimen positif.

Setelah analisis sentimen menggunakan *VADER*, kolom hasil *stemming* diubah menjadi representasi numerik dengan menggunakan teknik *TF-IDF (Term Frequency-Inverse Document Frequency)*, yang mengubah teks menjadi vektor fitur yang siap diproses lebih lanjut. Selanjutnya, model *Multinomial Naïve Bayes* diterapkan untuk melatih klasifikasi sentimen, dengan menggunakan data latih yang telah tertransformasi menjadi vektor *TF-IDF*. Selanjutnya, model ini digunakan untuk menentukan apakah sentimen yang dimasukkan ke dalam tulisan bersifat positif atau negatif.

Tabel 5. Hasil Analisis Sentimen

Hasil Stemming	Sentiment Score	Sentiment Label
good see disclaim real scam peopl need first place crypto cultur make founder launch token default expect that actual problem	-0.5423	Negatif

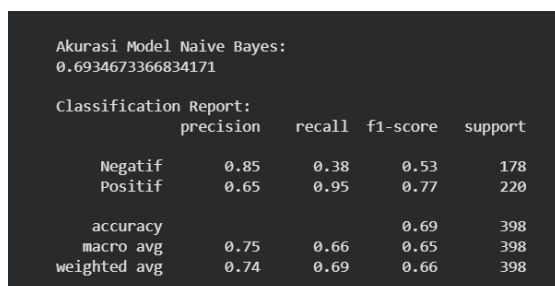
4.8. Hasil Evaluasi Model

Tahap evaluasi model dimulai dengan pengukuran akurasi dan *classification report* pada model *Naïve Bayes* yang telah dilatih menggunakan data pelatihan, bertujuan untuk memperoleh gambaran awal kinerja model. Selanjutnya, dilakukan *cross-validation* dengan teknik *10-fold* untuk menilai kemampuan generalisasi model serta menghindari *overfitting*. Setelah pelatihan ulang dengan nilai alpha terbaik untuk meningkatkan akurasi model, tahap berikutnya adalah optimasi parameter menggunakan *GridSearchCV*. Setelah itu, evaluasi lanjutan dilakukan menggunakan data uji untuk menerangi kinerja model. Sebagai langkah akhir, visualisasi *confusion matrix* digunakan untuk menganalisis

kesalahan klasifikasi dan memahami bagaimana model berfungsi untuk membedakan kelas-kelas yang ada.

4.9. Evaluasi Model Awal

Pada tahap pertama, model *Naïve Bayes* yang telah dilatih menggunakan data pelatihan dievaluasi. Tujuan evaluasi ini adalah untuk mendapatkan gambaran awal tentang kinerja model dengan mengukur *classification report* dan akurasi. *Classification report* memberikan metrik evaluasi yang lebih komprehensif seperti *precision*, *recall*, dan *f1-score*, yang menunjukkan seberapa baik model mengklasifikasikan setiap kategori sentimen. Hasil evaluasi awal ini akan menjadi dasar untuk langkah selanjutnya dalam perbaikan model.



```

Akurasi Model Naive Bayes:
0.6934673366834171

Classification Report:
      precision    recall  f1-score   support

Negatif      0.85      0.38      0.53       178
Positif      0.65      0.95      0.77       220

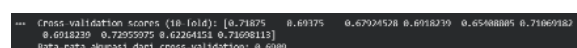
accuracy      0.69      0.69      0.69       398
macro avg      0.75      0.66      0.65       398
weighted avg      0.74      0.69      0.66       398
  
```

Gambar 1. Hasil Evaluasi Model Awal

Gambar 1. menunjukkan hasil evaluasi model *Naïve Bayes* yang mencakup akurasi dan *classification report*. Akurasi model menunjukkan tingkat akurasi model sebesar 0,694, yang menunjukkan bahwa model ini dapat memprediksi sentimen dengan tingkat kebenaran sekitar 69%. Untuk kategori sentimen negatif, tingkat *precision* sebesar 0.85, *recall* sebesar 0.38, dan *f1-score* sebesar 0.53, untuk kategori sentimen positif, tingkat *precision* sebesar 0.65, *recall* sebesar 0.95, dan *f1-score* sebesar 0.77.

4.10. Cross Validation untuk Mengukur Generalisasi Model

Setelah evaluasi awal selesai, langkah berikutnya adalah meluncurkan kemampuan model untuk diterapkan secara luas dengan menggunakan *cross-validation*. Teknik *cross-validation 10-fold* diterapkan untuk menilai stabilitas model dan mengurangi risiko *overfitting*, yakni kondisi di mana model terlalu cocok dengan data pelatihan dan kurang mampu menggeneralisasi pada data yang belum pernah dilihat. Pada *cross-validation*, data pelatihan dibagi menjadi 10 bagian, di mana model dilatih pada 9 bagian dan diuji pada 1 bagian, proses ini diulang sebanyak 10 kali. Hasil dari *cross-validation* memberikan gambaran yang lebih objektif mengenai seberapa baik model akan bekerja pada data yang tidak terlihat sebelumnya.



```

--- (cross validation scores (10 fold)): [0.71875 0.69175 0.67924528 0.6718279 0.6580885 0.71069182
0.6918239 0.72952975 0.62664153 0.71698113]
Rata rata akurasi dari cross validation: 0.6989
  
```

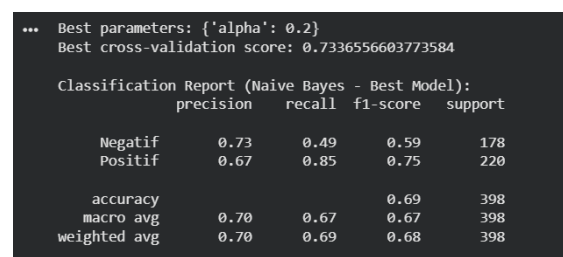
Gambar 2. Hasil Cross-Validation

Gambar 2. menampilkan hasil dari *cross-validation* model menggunakan metode *10-fold*. Setiap angka dalam urutan menunjukkan nilai akurasi pada masing-masing *fold*, dengan rentang nilai antara 0.62 hingga 0.73. Rata-rata akurasi dari seluruh *10-fold* adalah 0.6909, yang menunjukkan kinerja model yang relatif konsisten dalam memprediksi data yang belum terlihat. *Cross-validation* ini bertujuan untuk mengukur kemampuan model untuk menggeneralisasi pada berbagai jenis data. Hasilnya dapat memberikan gambaran yang lebih akurat tentang bagaimana model bekerja secara keseluruhan.

4.11. Mencari dan Melatih Parameter Terbaik dengan GridSearchCV

Setelah melakukan *cross-validation*, langkah selanjutnya adalah mengoptimalkan performa model dengan mencari parameter terbaik menggunakan *GridSearchCV*. *GridSearchCV* mencoba beberapa nilai *alpha* untuk model *Multinomial Naïve Bayes* dan memilih parameter yang memberikan akurasi tertinggi melalui *cross-validation*. Parameter terbaik yang ditemukan kemudian digunakan untuk melatih model ulang, dengan tujuan untuk meningkatkan akurasi dan mengurangi kemungkinan *overfitting*. Proses ini memastikan bahwa model yang digunakan untuk prediksi sudah dioptimalkan untuk performa terbaik.

Setelah model dengan parameter terbaik dipelajari, kinerjanya diuji kembali menggunakan data pengujian. *classification report* yang mencakup akurasi, keakuratan, dan penilaian F1, serta presisi, digunakan untuk melakukan penilaian ini. Tahap ini bertujuan untuk memastikan bahwa model yang dioptimalkan dapat menghasilkan prediksi yang lebih berkualitas dan akurat daripada model aslinya.



```

--- Best parameters: {'alpha': 0.2}
Best cross-validation score: 0.7336556603773584

Classification Report (Naive Bayes - Best Model):
      precision    recall  f1-score   support

Negatif      0.73      0.49      0.59       178
Positif      0.67      0.85      0.75       220

accuracy      0.69      0.69      0.69       398
macro avg      0.70      0.67      0.67       398
weighted avg      0.70      0.69      0.68       398
  
```

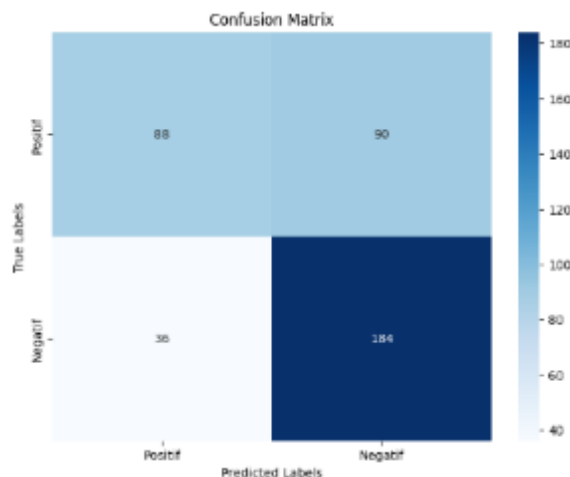
Gambar 3. Hasil Evaluasi Model dengan Parameter Terbaik

Gambar 3. menunjukkan hasil evaluasi model *Naïve Bayes* dengan parameter terbaik ($\alpha = 0.2$) dan skor *cross-validation* terbaik sebesar 0.7337. *Classification report* mencakup nilai *precision*, *recall*, dan *f1-score* untuk masing-masing kategori sentimen (negatif dan positif). Untuk sentimen negatif, *precision* adalah 0.73, *recall* 0.49, dan *f1-score* 0.59, sementara untuk sentimen positif, *precision* adalah 0.67, *recall* 0.85, dan *f1-score* 0.75.

4.12. Visualisasi Confusion Matrix

Confusion matrix merupakan alat yang berguna untuk menilai kesalahan klasifikasi yang dilakukan

oleh model serta memberikan gambaran lebih mendalam mengenai cara model membedakan antara kelas-kelas yang ada. Jumlah prediksi yang benar dan salah untuk masing-masing kelas (positif dan negatif) dapat dilihat dengan jelas dalam *confusion matrix* dalam bentuk *heatmap*. Visualisasi ini membantu memahami kinerja model dan menemukan area yang perlu perbaikan lebih lanjut.



Gambar 4. Visualisasi *Confusion Matrix*

Gambar 4. Menampilkan *confusion matrix* yang digunakan untuk mengevaluasi kinerja model dalam mengklasifikasikan sentimen. Matriks ini membandingkan label prediksi dengan label yang sebenarnya, dengan dua kelas sentimen yang dianalisis yaitu positif dan negatif. Dalam *confusion matrix* ini, terdapat 88 tweet yang benar dikategorikan sebagai positif (*True Positives*) dan 90 tweet yang salah dikategorikan sebagai negatif (*False Negatives*). Selain itu, terdapat 36 tweet yang salah dikategorikan sebagai positif (*False Positives*) dan 184 tweet yang benar dikategorikan sebagai negatif (*True Negatives*). Matriks ini memungkinkan untuk menilai seberapa akurat model dalam membedakan kedua kelas sentimen dan mengidentifikasi kesalahan klasifikasi yang terjadi.

Tabel berikut menunjukkan hasil dari *confusion matrix* yang digunakan untuk mengevaluasi kinerja model dalam mengklasifikasikan data sentimen menjadi dua kelas, yaitu positif dan negatif. *Confusion matrix* ini memperlihatkan jumlah prediksi yang benar dan salah untuk setiap kelas, serta memberikan gambaran yang lebih jelas mengenai bagaimana model membedakan antara kelas-kelas tersebut. Tabel ini menjadi dasar untuk menghitung berbagai metrik evaluasi, seperti *precision*, *recall*, dan *f1-score*, yang digunakan untuk menilai sejauh mana model berhasil dalam mengklasifikasikan data dengan akurat.

Tabel 6. Hasil *Confusion Matrix*

Aktual	Prediksi	
	Positif	Negatif
Positif	88	90
Negatif	36	184

Dari tabel diatas, bisa dilakukan perhitungan manual untuk mengetahui hasil *precision*, *recall*, dan *f1-score*.

Tabel 7. Perhitungan Manual *Confusion Matrix*

Akurasi	$Akurasi = \frac{88+184}{398} = \frac{272}{398} = 0.683$
Precision	$Precision\ Positive = \frac{88}{88+90} = \frac{88}{178} = 0.49$
	$Precision\ Negative = \frac{184}{184+36} = \frac{184}{220} = 0.836$
Recall	$Recall\ Positive = \frac{88}{88+36} = \frac{88}{124} = 0.71$
	$Recall\ Negative = \frac{184}{184+90} = \frac{184}{274} = 0.67$
F1-Score	$F1-Score\ Positive = 2 \times \frac{0.49 \times 0.71}{0.49 + 0.71} = 0.581$
	$F1-Score\ Negative = 2 \times \frac{0.836 \times 0.67}{0.836 + 0.67} = 0.744$

Tabel 8. Hasil Perhitungan Manual *Confusion Matrix*

Matrix	Nilai
Akurasi	68.3%
Precision Positif	49.4%
Precision Negatif	83.6%
Recall Positif	71%
Recall Negatif	67%
F1-Score Positif	58.1%
F1-Score Negatif	74.4%

4.13. Hasil Analisis Korelasi

Hasil analisis korelasi menunjukkan nilai koefisien korelasi antara harga ethereum dan sentimen positif sebesar 0.0071, serta antara harga ethereum dan sentimen negatif sebesar 0.0637. Nilai korelasi yang sangat rendah ini mengindikasikan adanya hubungan yang sangat lemah antara sentimen pasar dan harga ethereum. Dengan nilai korelasi yang mendekati nol, dapat disimpulkan bahwa perubahan sentimen pasar tidak memiliki dampak signifikan terhadap fluktuasi harga ethereum selama periode yang dianalisis.

Meskipun terdapat korelasi yang lemah, volatilitas harga yang tinggi menunjukkan bahwa faktor-faktor lain, selain sentimen pasar, mungkin lebih dominan dalam memengaruhi pergerakan harga ethereum. Faktor eksternal seperti regulasi, sentimen pasar global, dan berita terkait *cryptocurrency* mungkin berperan lebih besar dalam mempengaruhi fluktuasi harga.

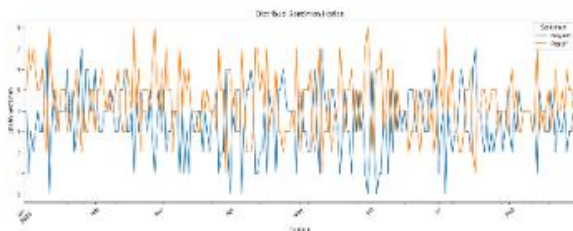
Rata-rata harga ethereum dalam periode yang dianalisis adalah 2.754,3 USD. Angka ini menggambarkan harga rata-rata ethereum yang diperdagangkan selama periode tersebut. Standar deviasi harga yang tercatat sebesar 810,11 USD menunjukkan tingkat volatilitas yang tinggi, yang berarti harga ethereum berfluktuasi cukup tajam dalam periode yang sama. Rentang harga, yang selisih antara harga tertinggi dan terendah, tercatat sebesar 3.357,86 USD, mengindikasikan variasi harga yang signifikan.



Gambar 5. Fluktuasi Harga Ethereum

Gambar 5. Menunjukkan fluktuasi harga Ethereum yang signifikan sepanjang periode dari Januari hingga Agustus 2025. Terdapat penurunan harga yang tajam pada beberapa bulan awal, diikuti dengan lonjakan harga yang cukup pesat pada pertengahan tahun hingga mencapai puncaknya pada Agustus 2025.

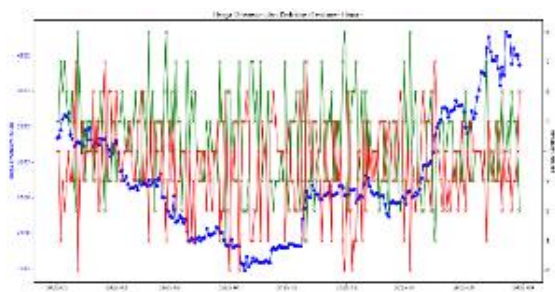
Perubahan perasaan atau reaksi terkait topik yang dianalisis dapat terlihat sepanjang waktu, dengan fluktuasi antara sentimen positif dan negatif.



Gambar 6. Distribusi Sentimen Harian

Gambar 6. menunjukkan dominasi sentimen positif terlihat jelas di sebagian besar periode, meskipun ada beberapa puncak sentimen negatif yang muncul pada hari-hari tertentu.

Pergerakan harga ethereum dan distribusi sentimen pasar tercermin dalam grafik yang menggambarkan fluktuasi harga harian dan jumlah sentimen positif serta negatif yang terdeteksi setiap hari yang terdapat pada gambar 7.



Gambar 7. Hubungan Sentimen dan Harga Ethereum

Gambar 7. menunjukkan hubungan antara distribusi sentimen harian (positif dan negatif) di twitter dan harga ethereum (dalam USD). Hal ini menunjukkan bahwa tidak ada korelasi langsung antara pergerakan harga ethereum dan opini publik. Garis biru yang menunjukkan harga ethereum

cenderung memiliki tren kenaikan atau penurunan yang lebih konsisten, sementara distribusi sentimen harian (ditunjukkan dengan silang hijau untuk sentimen positif dan merah untuk sentimen negatif) terlihat sangat berubah setiap hari, tanpa adanya pola yang konsisten seiring dengan pergerakan harga.

Meskipun jumlah sentimen positif telah meningkat seiring dengan kenaikan harga ethereum dalam beberapa waktu, hubungan antara keduanya tidak jelas. Dengan kata lain, dapat disimpulkan bahwa opini publik di twitter tidak memiliki pengaruh yang signifikan terhadap perubahan harga ethereum selama rentang waktu yang ditampilkan pada grafik tersebut. Ini karena faktor lain cenderung lebih mempengaruhi perubahan harga ethereum.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa sentimen publik di twitter tidak memiliki pengaruh signifikan terhadap fluktuasi harga ethereum selama periode yang dianalisis. Meskipun terdapat fluktuasi antara sentimen positif dan negatif, nilai koefisien korelasi yang sangat rendah antara harga ethereum dan sentimen pasar mengindikasikan hubungan yang lemah. Hal ini menandakan bahwa faktor lain, seperti kebijakan ekonomi, regulasi pasar, dan berita terkait *cryptocurrency*, lebih memengaruhi pergerakan harga ethereum daripada opini publik yang tercermin di media sosial.

Penelitian selanjutnya sebaiknya mempertimbangkan faktor eksternal lainnya, seperti kebijakan pemerintah dan faktor makroekonomi, yang dapat mempengaruhi harga ethereum. Selain itu, menggabungkan analisis sentimen dengan variabel fundamental lainnya akan memberikan gambaran yang lebih komprehensif dalam memprediksi pergerakan harga aset kripto. Dengan demikian, para investor dan pelaku pasar dapat membuat keputusan yang lebih terinformasi dan akurat.

DAFTAR PUSTAKA

- [1] S. E. Hardiyanti, "Analysis of Volatility and Factors Affecting Cryptocurrency Prices: A Case Study of Bitcoin and Ethereum," *MSR J.*, vol. 3, no. 2, hal. 2024, 2024.
- [2] P. Alipour dan S. E. Charandabi, "Analyzing the Interaction between Tweet Sentiments and Price Volatility of Cryptocurrencies," *Eur. J. Bus. Manag. Res.*, vol. 8, no. 2, hal. 211–215, 2023, doi: 10.24018/ejbmr.2023.8.2.1865.
- [3] A. S. Rusydiana, "Cryptocurrency: A Sentiment Analysis from Twitter Data," *Islam. Financ. Technol.*, vol. 1, no. 1, 2024, doi: 10.58968/ift.v1i1.359.
- [4] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," 2019, [Daring]. Tersedia pada: <http://arxiv.org/abs/1908.10063>
- [5] A. Khan dan I. Ihsan, "Crypto Currency Price Prediction through Tweets Using NLP," *J.*

- Comput. Biomed. Informatics*, vol. 7, no. 1, hal. 1, 2024, [Daring]. Tersedia pada: <https://doi.org/10.56979/701/2024>
- [6] S. Ahmed, "Enhancing Data Security and Transparency: The Role of Blockchain in Decentralized Systems," *Int. J. Adv. Eng. Manag. Sci.*, vol. 11, no. 1, hal. 167–176, 2025, doi: 10.22161/ijaems.111.12.
- [7] D. Hardiansyah, R. A. Aziz, dan M. S. Hasibuan, "The Classification Method is Used for Sentiment Analysis in My Telkomsel," *Int. J. Artif. Intell. Res.*, vol. 8, no. 2, hal. 169, 2024, doi: 10.29099/ijair.v8i2.1229.
- [8] S. K. Rongali, "Natural Language Processing (NLP) in Artificial Intelligence," *World J. Adv. Res. Rev.*, vol. 25, no. 1, hal. 2515–2519, 2025, doi: 10.30574/wjarr.2025.25.1.0277.
- [9] M. Masadeh *et al.*, "Investigating the Impact of Preprocessing Techniques and Representation Models on Arabic Text Classification using Machine Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 1, hal. 1115–1123, 2024, doi: 10.14569/IJACSA.2024.01501110.
- [10] E. Hruska dan F. Liu, "Machine learning: An overview," *Quantum Chem. Age Mach. Learn.*, vol. 9, no. 5, hal. 135–151, 2022, doi: 10.1016/B978-0-323-90049-2.00024-X.
- [11] Marc Zenus Labuguen, "Enhancing LSTM performance in sentiment analysis through advanced data preprocessing and model optimization techniques," *World J. Adv. Res. Rev.*, vol. 25, no. 1, hal. 2433–2443, 2025, doi: 10.30574/wjarr.2025.25.1.0098.
- [12] A. A. Purnama dan Y. R. Sipayung, "Sentiment Analysis of Public Service Using Naïve Bayes Classifier," *J. Inf. Syst. Informatics*, vol. 7, no. 3, hal. 2439–2457, 2025, doi: 10.51519/journalisi.v7i3.1207.
- [13] I. Makarchuk, I. Granovska, dan I. Makarchuk, "Cryptocurrencies from a behavioural finance perspective," *Univ. Econ. Bull.*, vol. 18, no. 4, hal. 17–23, 2023, doi: 10.69587/ueb/4.2023.17.