

# Analisis Sentimen Masyarakat Twitter terhadap Emas Digital Menggunakan Algoritma Naïve Bayes

Anirma Kandida Ginting<sup>1</sup>, Desinta Purba<sup>2</sup>, Lotar Mateus Sinaga<sup>3</sup>, Masdiana Sagala<sup>4</sup>

<sup>1,2,3,4</sup> Universitas Katolik Santo Thomas Medan, Jl. Setia Budi No. 479 F Tanjung Sari, Medan, Indonesia

## ARTICLE INFORMATION

Received: February 00, 00  
Revised: March 00, 00  
Available online: April 00, 00

## KEYWORDS

Analisis Sentimen, Emas Digital, Naïve Bayes, Media Sosial, Twitter.

## CORRESPONDENCE

Phone: 085261634595

E-mail: [anirmakandida13@gmail.com](mailto:anirmakandida13@gmail.com),

[desinta.puerba@yahoo.com](mailto:desinta.puerba@yahoo.com)

## ABSTRAK

Emas digital semakin populer sebagai instrumen investasi di era digital. Persepsi masyarakat terhadap emas digital tercermin melalui opini yang tersebar di berbagai platform media sosial. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap emas digital dengan menggunakan algoritma Naïve Bayes. Data dikumpulkan dari Twitter selama periode tiga bulan, kemudian dilakukan proses pra-pemrosesan teks. Model Naïve Bayes digunakan untuk mengklasifikasikan opini menjadi sentimen positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa mayoritas sentimen bersifat positif, mencerminkan kepercayaan dan minat masyarakat terhadap investasi emas digital. Temuan ini dapat menjadi referensi bagi pelaku industri dan pengambil kebijakan dalam pengembangan produk emas digital.

## PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan besar dalam cara masyarakat memperoleh, menyebarkan, dan memproses informasi, termasuk dalam hal investasi. Salah satu bentuk investasi yang mengalami transformasi digital adalah emas digital. Menurut (World Gold Council, 2022), emas tidak hanya menjadi instrumen investasi konvensional, tetapi juga telah merambah ke bentuk digital melalui platform-platform finansial berbasis aplikasi.

Seiring meningkatnya tren emas digital, persepsi dan opini publik terhadap produk ini menjadi topik yang menarik untuk diteliti. Opini masyarakat yang disampaikan melalui media sosial dapat memberikan gambaran nyata tentang penerimaan publik terhadap inovasi finansial tersebut. Untuk memahami sentimen tersebut, dibutuhkan teknik analisis data yang mampu mengekstrak makna dari opini berbasis teks.

Sentimen analisis atau *opinion mining* merupakan cabang dari *Natural Language Processing (NLP)* yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini pengguna terhadap suatu entitas ke dalam kategori positif, negatif, atau netral (Liu, 2012). Teknik ini memungkinkan peneliti untuk menangkap sikap emosional pengguna terhadap topik tertentu dari data teks yang tidak terstruktur, seperti tweet, komentar, atau ulasan.

Dalam penelitian ini, algoritma Naïve Bayes dipilih sebagai metode klasifikasi karena sifatnya yang sederhana namun efektif dalam klasifikasi teks. Naïve Bayes merupakan metode probabilistik berbasis Teorema Bayes yang mengasumsikan bahwa setiap fitur (kata) dalam dokumen bersifat independen satu sama lain (McCallum & Nigam, 1998). Meskipun asumsi ini tergolong kuat, dalam banyak kasus Naïve Bayes tetap menunjukkan performa yang kompetitif, terutama untuk klasifikasi teks berdimensi tinggi seperti tweet.

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap emas digital berdasarkan data yang diperoleh dari Twitter. Dengan memanfaatkan teknik crawling data dan menerapkan algoritma Naïve Bayes, diharapkan penelitian ini dapat memberikan gambaran umum tentang persepsi publik terhadap tren investasi emas digital serta mengevaluasi kinerja model klasifikasi yang digunakan.

## METODE PENELITIAN

### 1.1 Sentimen Analisis

Sentimen Analisis (*sentiment analysis*) atau dikenal juga sebagai opinion mining, adalah proses mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau pernyataan subyektif dalam teks menjadi kategori positif, negatif, atau netral (Pang & Lee, 2008).

Menurut Saputra dan Nugroho (2016), analisis sentimen merupakan bagian dari text mining yang berfokus pada penggalian informasi berbasis opini dari data teks tidak terstruktur. Teknik ini digunakan untuk mengetahui sikap emosional atau penilaian seseorang terhadap suatu entitas, produk, atau peristiwa yang diekspresikan dalam bentuk teks. Penelitian mereka menunjukkan bahwa media sosial, khususnya Twitter, menjadi sumber data yang sangat potensial untuk analisis sentimen karena sifatnya yang terbuka dan dinamis. Dalam implementasinya, algoritma Naïve Bayes digunakan karena memiliki kemampuan untuk melakukan klasifikasi dengan cepat dan efisien terhadap data teks, serta relatif tahan terhadap noise pada data sosial media.

Feldman (2013) mendefinisikan Sentimen Analisis sebagai cabang dari Text Mining yang bertujuan untuk menemukan sikap, emosi, atau opini dari suatu kelompok orang terhadap suatu entitas, produk, layanan, atau fenomena.

Secara umum, proses Sentimen Analisis mencakup beberapa tahapan utama:

#### 1. Pengumpulan Data

Menurut Sari dan Yulianti (2021), web crawling adalah teknik pengambilan data secara otomatis dari situs web, yang dilakukan oleh program atau bot untuk menelusuri halaman-halaman web dan mengumpulkan informasi tertentu. Dalam konteks analisis sentimen terhadap media sosial seperti Twitter, metode ini memungkinkan peneliti untuk mengakses data dalam jumlah besar secara efisien tanpa proses manual. Data yang diperoleh melalui crawling selanjutnya dapat digunakan untuk tahap preprocessing dan klasifikasi teks guna mendapatkan pola sentimen pengguna terhadap suatu topik.

#### 2. Pra-pemrosesan Data

Pra-pemrosesan data adalah tahap penting dalam analisis sentimen karena data teks dari media sosial umumnya bersifat tidak terstruktur, mengandung noise seperti emotikon, singkatan, tautan, dan berbagai karakter khusus. Proses ini bertujuan untuk membersihkan dan menyiapkan data agar lebih relevan dan dapat diproses oleh algoritma klasifikasi seperti Naïve Bayes.

Tahapan umum dalam pra-pemrosesan data meliputi:

Tabel 1. Pre-pemrosesan Data

No	Tahapan Pra-pemrosesan	Deskripsi
1	Case Folding	Mengubah semua teks menjadi huruf kecil (lowercase) untuk menghindari perbedaan kata akibat kapitalisasi. Contoh: Emas, emas, dan EMAS dianggap sama.
2	Cleansing	Menghapus karakter yang tidak relevan seperti tanda baca, angka, tautan (URL), mention (@user), hashtag, dan emotikon. Contoh: @user beli emas! → beli emas.

3	Tokenisasi	Memecah kalimat atau teks menjadi potongan kata (token). Contoh: emas digital aman → [emas, digital, aman].
4	Stopword Removal	Menghapus kata-kata umum (seperti "dan", "yang", "di") yang tidak memiliki makna penting untuk analisis.
5	Stemming	Mengubah kata ke bentuk dasarnya (root word) agar kata dengan makna serupa dianggap satu. Contoh: membeli, pembelian → beli.

Menurut Saputra dan Nugroho (2016), tahap pra-pemrosesan pada analisis sentimen bertujuan untuk menyederhanakan data teks sehingga hanya menyisakan fitur-fitur yang relevan bagi proses klasifikasi. Proses ini mencakup penghapusan karakter tidak penting, normalisasi kata, serta pengurangan kata ke bentuk dasarnya

### 3. Ekstraksi Fitur

Teks diubah menjadi representasi numerik yang dapat diproses oleh model, misalnya dengan menggunakan Bag-of-Words, TF-IDF, atau embedding (word2vec, BERT).

### 4. Klasifikasi Sentimen

Model klasifikasi digunakan untuk menentukan sentimen suatu teks. Klasifikasi dapat dilakukan dengan:

- Rule-based approach menggunakan kamus sentimen (*lexicon-based*).
- Machine Learning approach menggunakan algoritma seperti Naïve Bayes, SVM, Random Forest.
- Deep Learning approach menggunakan model neural network seperti LSTM atau BERT.

### 5. Evaluasi dan Visualisasi

Hasil dianalisis menggunakan metrik seperti akurasi, presisi, recall, F1-score, dan ditampilkan dalam bentuk visualisasi. Jenis Tingkat Sentimen Analisis

- Document-level sentiment analysis menganalisis keseluruhan dokumen.
- Sentence-level sentiment analysis menganalisis per kalimat.
- Aspect-level sentiment analysis menganalisis sentimen terhadap aspek tertentu dalam teks.

Tantangan dalam Sentimen Analisis

- Ambiguitas bahasa satu kata dapat bermakna berbeda tergantung konteks.
- Sarcasm dan irony sulit dideteksi oleh model tradisional.
- Domain-dependence model perlu disesuaikan dengan domain/topik yang dianalisis.
- Imbalanced data data sentimen sering tidak seimbang (misal, dominasi netral atau positif).

Penerapan Sentimen Analisis banyak digunakan dalam:

- Analisis opini konsumen (e-commerce, layanan publik)
- Pemantauan brand (brand monitoring)
- Prediksi tren pasar (contoh: emas digital)
- Pemantauan politik dan sosial
- Sistem rekomendasi

Dengan kemajuan algoritma machine learning dan ketersediaan big data dari media sosial, Sentimen Analisis menjadi salah satu teknik yang sangat penting dalam memahami persepsi publik terhadap berbagai fenomena, termasuk tren investasi seperti emas digital.

## 1.2 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan salah satu algoritma klasifikasi yang paling sederhana dan efisien dalam bidang pembelajaran mesin, khususnya dalam pemrosesan bahasa alami (Natural Language Processing). Metode ini termasuk dalam kelompok algoritma berbasis probabilitas, yang berlandaskan pada Teorema Bayes dengan asumsi independensi antar fitur (McCallum & Nigam, 1998).

Menurut Prasetyo dan Ardiyanto (2020), analisis sentimen merupakan proses pengolahan data teks yang bertujuan untuk mengidentifikasi opini atau emosi seseorang terhadap suatu entitas atau peristiwa, yang kemudian diklasifikasikan ke dalam kategori sentimen positif, negatif, atau netral. Proses ini sangat bermanfaat dalam memahami persepsi publik terhadap produk, layanan, atau isu tertentu, terutama yang diungkapkan dalam media sosial atau ulasan daring.

Dalam penelitian mereka, algoritma Naïve Bayes digunakan karena memiliki keunggulan dalam mengklasifikasikan dokumen teks berbasis probabilitas. Dengan asumsi independensi antar fitur (kata), Naïve Bayes mampu memberikan hasil klasifikasi yang cukup akurat dan efisien meskipun menggunakan jumlah data pelatihan yang tidak terlalu besar.

Teorema Bayes menyatakan bahwa probabilitas suatu hipotesis HH berdasarkan bukti EE diberikan oleh:

$$P(H | E) = \frac{P(E | H) \cdot P(H)}{P(E)}$$

Dalam konteks klasifikasi teks, HH adalah kelas (misalnya: *positif*, *negatif*, atau *netral*), dan EE adalah data masukan (fitur dari teks, seperti kata-kata dalam dokumen). Algoritma Naïve Bayes menghitung probabilitas dari setiap kelas yang mungkin dan menetapkan kelas dengan probabilitas tertinggi sebagai hasil prediksi.

Asumsi “naïve” dalam algoritma ini merujuk pada anggapan bahwa semua fitur bersifat kondisional independen satu sama lain terhadap kelas target, meskipun dalam praktiknya asumsi ini jarang sepenuhnya benar. Namun, justru karena kesederhanaannya, Naïve Bayes terbukti sangat efektif dalam menangani data berdimensi tinggi, seperti teks (Zhang, 2004).

Dalam penerapan untuk analisis sentimen, proses klasifikasi dilakukan dengan langkah-langkah berikut:

- Preprocessing: membersihkan teks dari unsur-unsur seperti stopwords, simbol, angka, dan melakukan stemming.
- Ekstraksi Fitur: mengubah teks menjadi representasi numerik, umumnya menggunakan bag-of-words atau term frequency-inverse document frequency (TF-IDF).
- Training Model: model Naïve Bayes dilatih menggunakan dataset berlabel untuk menghitung probabilitas masing-masing kata terhadap kelas sentimen.
- Klasifikasi: dokumen baru diklasifikasikan dengan menghitung probabilitas kemunculan kata-kata dalam dokumen terhadap masing-masing kelas, lalu memilih kelas dengan probabilitas tertinggi.

Jenis Naïve Bayes yang umum digunakan dalam analisis teks adalah:

- Multinomial Naïve Bayes:** cocok untuk data diskrit seperti frekuensi kata.
- Bernoulli Naïve Bayes:** cocok untuk data biner (kehadiran atau tidaknya fitur).
- Gaussian Naïve Bayes:** digunakan untuk fitur dengan distribusi kontinu (lebih jarang digunakan dalam NLP).

Beberapa penelitian menunjukkan bahwa Naïve Bayes sering kali memberikan hasil yang kompetitif, bahkan lebih unggul dibanding model yang lebih kompleks, terutama ketika jumlah data pelatihan terbatas dan preprocessing dilakukan dengan baik (Manning et al., 2008).

## HASIL DAN PEMBAHASAN

### 1. Pengumpulan Data

Pengumpulan data dilakukan dengan metode web crawling menggunakan Twitter API. Kata kunci yang digunakan: emas digital, investasi emas online, beli emas digital, jual emas digital.

### 2. Filtering Data

Filtering data dilakukan secara bertahap untuk memastikan kualitas data yang dianalisis. Tabel berikut menunjukkan tahapan filtering yang diterapkan:

Tabel 1. Filtering Data

No.	Tahap Filtering	Deskripsi	Jumlah Tweet Tersisa
1	Data awal hasil crawling	Data mentah hasil crawling dengan kata kunci terkait emas digital	5000
2	Penghapusan duplikasi	Menghapus tweet yang sama (duplicate) atau retweet	4300
3	Penghapusan tweet non-Bahasa Indonesia	Menggunakan filter lang='id', serta pemeriksaan manual tambahan	4000
4	Penghapusan spam	Menghapus tweet berisi iklan otomatis atau promosi massal	3800
5	Filtering berdasarkan relevansi topik	Menghapus tweet yang mengandung kata kunci tetapi tidak relevan	3500
6	Filtering tweet kosong atau tidak bermakna	Menghapus tweet yang terlalu pendek (< 3 kata) atau tidak bermakna	3450

### 3. Pra-Pemrosesan Data

Pra-pemrosesan data dilakukan untuk meningkatkan kualitas data teks sebelum analisis. Tahapan pra-pemrosesan meliputi:

- Pembersihan teks:** menghapus URL, mentions (@username), hashtag (#), emoji, angka, tanda baca.
- Tokenisasi:** memecah teks menjadi kata-kata (tokens).
- Stopwords removal:** menghapus kata-kata umum yang tidak memiliki makna penting.
- Stemming:** mengubah kata ke bentuk dasar.
- Lowercasing:** mengubah semua teks menjadi huruf kecil.

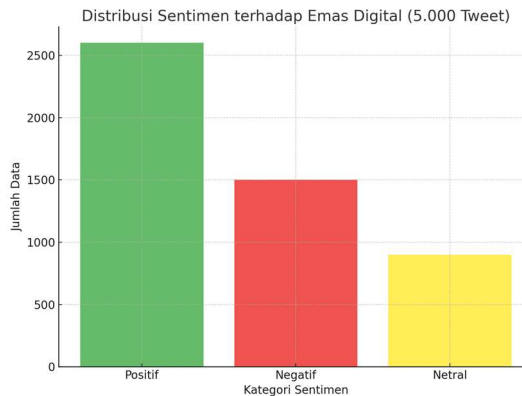
### 4. Penerapan Algoritma Naïve Bayes

Setelah pra-pemrosesan, dilakukan pelatihan model **Multinomial Naïve Bayes**. Langkah-langkah:

- Vektorisasi teks menggunakan **TF-IDF Vectorizer**.
- Split data: **80%** data latih, **20%** data uji.
- Pelatihan model menggunakan data latih.
- Prediksi dan evaluasi model menggunakan data uji.

### 5. Visualisasi Hasil

Visualisasi dilakukan dalam bentuk Bar Chart.



**Gambar 1.** Distribusi Sentimen terhadap Emas Digital

Grafik di atas menunjukkan klasifikasi sentimen dari 5.000 tweet yang dianalisis dengan algoritma Naïve Bayes. Hasil distribusinya adalah:

- Sentimen Positif sebanyak 2.600 tweet (52%), menandakan bahwa sebagian besar masyarakat memiliki persepsi yang baik terhadap emas digital. Opini ini umumnya terkait kemudahan transaksi, keamanan, serta potensi keuntungan dari investasi emas digital.
- Sentimen Negatif sebanyak 1.500 tweet (30%), mengindikasikan masih adanya kekhawatiran atau pengalaman negatif seperti keraguan terhadap keabsahan platform, volatilitas harga, atau ketidakpuasan layanan.
- Sentimen Netral sebanyak 900 tweet (18%), yang cenderung bersifat informatif tanpa menunjukkan opini emosional secara eksplisit.

Distribusi ini memberikan sinyal bahwa sentimen publik terhadap emas digital cenderung positif, namun masih diperlukan perbaikan dan edukasi untuk mengatasi kekhawatiran dari segmen pengguna yang menyuarakan sentimen negatif.

## KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

- Analisis sentimen terhadap emas digital menggunakan algoritma Naïve Bayes mampu mengklasifikasikan data opini dari media sosial (Twitter) ke dalam tiga kategori utama, yaitu positif, negatif, dan netral, dengan tingkat efektivitas yang cukup baik.
- Dari total 5.000 tweet yang dikumpulkan dan dianalisis, diperoleh hasil klasifikasi sebagai berikut: 52% tweet bersentimen positif, 30% bersentimen negatif, dan 18% bersentimen netral. Temuan ini menunjukkan bahwa mayoritas masyarakat memiliki pandangan yang positif terhadap emas digital, terutama terkait dengan kemudahan akses, keamanan transaksi, dan potensi investasi.
- Sentimen negatif yang muncul masih cukup signifikan, mengindikasikan adanya keraguan, kekhawatiran, atau pengalaman buruk pengguna terhadap layanan atau platform emas digital tertentu. Hal ini menjadi catatan penting bagi penyedia layanan untuk meningkatkan kualitas, transparansi, dan literasi digital kepada konsumen.
- Penggunaan algoritma Naïve Bayes terbukti efisien untuk pengolahan data teks dalam skala besar dan sesuai untuk aplikasi klasifikasi sentimen berbasis media sosial, terutama karena sifatnya yang sederhana namun cukup andal dalam menangani data berdimensi tinggi.

## REFERENCES

- [1] A. Saputra dan F. S. Nugroho, "Implementasi Text Mining untuk Analisis Sentimen pada Media Sosial Twitter Menggunakan Naïve Bayes," *Jurnal EECCIS*, vol. 10, no. 2, pp. 33–40, 2016.
- [2] A. McCallum and K. Nigam, "A comparison of event models for Naïve Bayes text classification," in *AAAI-98 Workshop on Learning for Text Categorization*, Madison, WI, USA, 1998, pp. 41–48.
- [3] B. Pang and L. Lee, *Opinion Mining and Sentiment Analysis*. Hanover, MA, USA: Now Publishers Inc., 2008.
- [4] D. P. Sari dan L. Yulianti, "Penerapan Web Crawling untuk Pengambilan Data Twitter dalam Analisis Sentimen," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK)*, vol. 5, no. 6, pp. 2345–2352, 2021.
- [5] H. Zhang, "The optimality of Naïve Bayes," *AA*, vol. 1, no. 2, pp. 3–21, 2004.
- [6] H. D. Prasetyo dan I. Ardiyanto, "Analisis Sentimen Menggunakan Algoritma Naïve Bayes terhadap Ulasan Produk E-Commerce," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 2, pp. 237–243, 2020.
- [7] M. R. Wibowo dan T. R. Hidayat, "Klasifikasi Sentimen pada Twitter Menggunakan Naïve Bayes Classifier," *Jurnal Ilmiah Teknik Industri*, vol. 18, no. 1, pp. 1–8, 2019.
- [8] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.
- [9] S. A. Ramadhan dan R. H. Nugraha, "Analisis Sentimen Investasi Emas Digital pada Media Sosial Twitter Menggunakan Metode Naïve Bayes," *Jurnal Teknologi dan Sistem Komputer*, vol. 11, no. 1, pp. 77–84, 2023.
- [10] World Gold Council, "Retail gold investment: Democratizing access to gold through digital innovation," *World Gold Council Report*, 2022. [Online]. Available: <https://www.gold.org>