

ANALISIS SENTIMEN ISI KOMENTAR PUBLIK INDONESIA TERHADAP VIDEO TRADER CRYPTO DI YOUTUBE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Saepuloh¹, Nana Suarna², Agus Bahtiar³, Puji Pramudya Marta⁴.

Program Studi Teknik Informatika¹²
Program Studi Sistem Informasi³⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
saepullohgrover32@gmail.com

(*) Corresponding Author : saepullohgrover32@gmail.com
Published : 30 Juni 2026

Abstract—This study aims to analyze Indonesian public sentiment toward comments on YouTube videos about crypto traders using the Multinomial Naïve Bayes algorithm and Support Vector Machine (SVM). The research data was obtained through scraping three relevant YouTube videos, resulting in 1,042 raw comments. The data was then subjected to preprocessing steps including case folding, cleaning, tokenizing, stopword removal, and stemming to produce more structured text. The sentiment labeling process was carried out using an Indonesian sentiment dictionary approach, referring to the Indonesian Sentiment Lexicon, which is openly available on the GitHub repository. Each comment was classified into three sentiment categories: positive, negative, and neutral, based on word polarity scores. The initial data distribution showed a class imbalance, so the Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the data in each sentiment class. After the balancing process, the data increased to 2,115 comments, with a balanced distribution of 705 comments for each sentiment class. The data was then represented numerically using the TF-IDF method and divided into training and testing data with an 80:20 ratio. The results showed that the Multinomial Naïve Bayes algorithm achieved an accuracy of 73.43%, while the Support Vector Machine (SVM) algorithm achieved a higher accuracy of 82.71%. This indicates that SVM has better capabilities in handling high-dimensional text data such as YouTube comments. Sentiment distribution analysis and word cloud visualizations indicate that public opinion towards crypto trader content on YouTube tends to be dominated by negative sentiment. This research contributes to the development of machine learning-based sentiment analysis and provides an overview of public perception of the crypto trader phenomenon on social media.

Keywords: sentiment analysis, Naïve Bayes, SVM, YouTube, trader crypto, SMOTE, TF-IDF.

Abstrak—Penelitian ini bertujuan untuk menganalisis sentimen publik Indonesia terhadap komentar pada video YouTube bertema trader *crypto* menggunakan algoritma *Multinomial Naïve Bayes* dan *Support Vector Machine (SVM)*. Data penelitian diperoleh melalui proses *scraping* dari tiga video YouTube yang relevan sehingga diperoleh sebanyak 1.042 komentar mentah. Data kemudian diproses melalui tahapan *preprocessing* yang meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming* untuk menghasilkan teks yang lebih terstruktur. Proses pelabelan sentimen dilakukan dengan pendekatan kamus sentimen bahasa Indonesia yang mengacu pada *Indonesian Sentiment Lexicon* yang tersedia secara terbuka pada *repository GitHub*. Setiap komentar diklasifikasikan ke dalam tiga kategori sentimen yaitu positif, negatif, dan netral berdasarkan skor polaritas kata. Distribusi awal data menunjukkan ketidakseimbangan kelas, sehingga diterapkan teknik *Synthetic Minority Oversampling Technique (SMOTE)* untuk menyeimbangkan jumlah data pada setiap kelas sentimen. Setelah proses penyeimbangan, jumlah data meningkat menjadi 2.115 komentar dengan distribusi seimbang sebanyak 705 komentar untuk masing-masing kelas sentimen. Data kemudian direpresentasikan ke dalam bentuk numerik menggunakan metode *TF-IDF* dan dibagi menjadi data pelatihan dan pengujian dengan proporsi 80:20. Hasil penelitian menunjukkan bahwa algoritma *Multinomial Naïve Bayes* menghasilkan akurasi sebesar 73,43%, sedangkan algoritma *Support Vector Machine (SVM)* memperoleh akurasi yang lebih tinggi yaitu 82,71%. Hal ini menunjukkan bahwa *SVM* memiliki kemampuan yang lebih baik dalam menangani data teks berdimensi tinggi seperti komentar YouTube. Analisis distribusi sentimen dan visualisasi word cloud menunjukkan bahwa opini publik terhadap konten trader *crypto* di YouTube cenderung didominasi oleh sentimen negatif.

Penelitian ini memberikan kontribusi dalam pengembangan analisis sentimen berbasis *machine learning* serta memberikan gambaran mengenai persepsi masyarakat terhadap fenomena trader crypto di media sosial.

Kata Kunci: *sentiment analysis, Naïve Bayes, SVM, YouTube, trader crypto, SMOTE, TF-IDF.*

INTRODUCTION

Perkembangan teknologi digital dalam satu dekade terakhir telah membawa perubahan signifikan terhadap cara masyarakat mengakses informasi dan melakukan aktivitas ekonomi, termasuk dalam bidang investasi aset kripto (*cryptocurrency*). Di Indonesia, fenomena peningkatan minat terhadap *cryptocurrency* didorong oleh semakin mudahnya akses *platform* perdagangan digital serta meningkatnya paparan melalui media sosial, terutama YouTube. *Platform* ini berperan sebagai ruang publik tempat opini, emosi, dan persepsi keuangan diproduksi serta dipertukarkan secara terbuka menurut peneliti [1].

YouTube menjadi salah satu media yang paling berpengaruh karena konten video dan ruang komentar menghasilkan interaksi dua arah antara kreator dan *audiens*. Studi menunjukkan bahwa komentar penonton dapat dipengaruhi oleh retorika, gaya penyampaian, hingga emosi dari pembuat konten, yang kemudian memicu efek "*cryptoemotional contagion*" atau penularan emosi dalam komunitas digital menurut (Meyer et al., 2023). Selain itu, peran *influencer* dalam memengaruhi keputusan ekonomi juga semakin kuat, terutama dalam konteks *crypto* trading yang sangat bergantung pada kepercayaan dan persepsi risiko menurut [2].

Di sisi lain, diskursus mengenai trader *crypto* tidak hanya bernuansa informatif tetapi juga sarat kritik sosial. Dalam konteks global, *crypto* dianggap sebagai media pembentukan imajinasi ekonomi baru dan kerap dikaitkan dengan narasi pemberdayaan finansial maupun spekulasi berisiko tinggi menurut (El Hajj & Farran, 2024; Cossu, 2022). Hal ini tercermin dalam komentar publik Indonesia, di mana sebagian penonton memberikan respons positif sebagai bentuk apresiasi terhadap edukasi keuangan, sedangkan sebagian lainnya menunjukkan *sentimen* negatif terkait afiliasi, penipuan, atau keraguan terhadap legalitas aktivitas trading [3], [4].

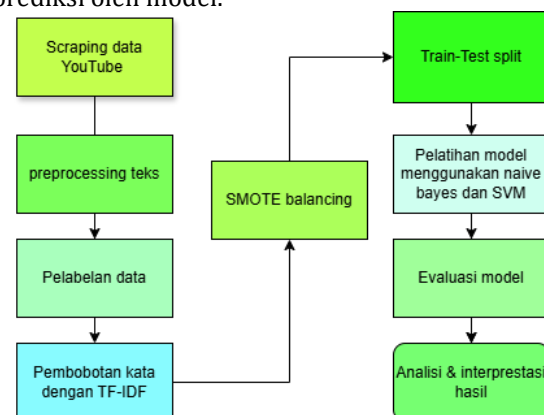
Analisis *sentimen* merupakan pendekatan utama yang digunakan untuk memetakan kecenderungan opini publik tersebut. Pendekatan ini telah terbukti efektif dalam mengidentifikasi pola emosi, persepsi risiko,

serta kecenderungan perilaku digital dalam konteks aset kripto menurut (Meyer et al., 2023). Algoritma *Naïve Bayes* biasa digunakan sebagai model *baseline* pada analisis *sentimen* karena kesederhanaan dan efektivitasnya pada teks berdimensi tinggi, sementara *Support Vector Machine (SVM)* digunakan sebagai model pembanding dengan performa lebih stabil pada teks pendek menurut [2].

Pada penelitian ini, sebanyak 1.042 komentar mentah yang berasal dari beberapa video trader *crypto* Indonesia di YouTube digunakan sebagai *dataset* utama. Sebelum dianalisis, komentar diproses melalui tahapan *preprocessing*, pelabelan *sentimen*, serta pembobotan teks menggunakan *TF-IDF*. Karena distribusi label tidak seimbang, metode *synthetic minority oversampling technique (SMOTE)* diterapkan untuk menyeimbangkan jumlah data pada setiap kelas *sentimen* agar model tidak bias.

MATERIALS AND METHODS

Alur penelitian ini menggambarkan tahapan sistematis dalam melakukan analisis sentimen atau klasifikasi teks berbasis data dari platform digital, khususnya YouTube. Penelitian diawali dengan proses pengumpulan data menggunakan teknik *scraping* untuk memperoleh komentar atau teks yang relevan. Data mentah yang diperoleh biasanya masih mengandung banyak noise, sehingga perlu dilakukan serangkaian tahapan *preprocessing* agar data menjadi lebih bersih dan siap dianalisis. Selain itu, pelabelan data juga menjadi bagian penting untuk menentukan kelas atau kategori yang akan diprediksi oleh model.



Gambar 1. Alur Penelitian

Berdasarkan gambar tersebut, proses dimulai dari *scraping* data YouTube untuk mengumpulkan

dataset. Selanjutnya dilakukan preprocessing teks, seperti pembersihan simbol, penghapusan stopwords, dan normalisasi kata. Setelah itu, data diberi label melalui tahap pelabelan data untuk menentukan kelas tertentu, misalnya sentimen positif atau negatif. Tahap berikutnya adalah pembobotan kata menggunakan TF-IDF, yang berfungsi mengubah teks menjadi representasi numerik agar dapat diproses oleh algoritma machine learning.

Untuk mengatasi ketidakseimbangan data, diterapkan metode SMOTE balancing sehingga distribusi kelas menjadi lebih seimbang. Data kemudian dibagi menjadi data latih dan data uji pada tahap train-test split. Selanjutnya dilakukan pelatihan model menggunakan algoritma seperti Naive Bayes dan Support Vector Machine (SVM). Setelah model dilatih, dilakukan evaluasi model menggunakan metrik tertentu untuk mengukur performanya. Tahap akhir adalah analisis dan interpretasi hasil, yang bertujuan untuk memahami kinerja model serta menarik kesimpulan dari penelitian yang dilakukan

RESULTS AND DISCUSSION

Model multinomial Naïve Bayes (NB) dilatih menggunakan hasil transformasi *TF-IDF*. Model ini menghitung probabilitas kata dalam setiap kelas *sentimen* dan menentukan kelas dengan probabilitas tertinggi sebagai hasil prediksi.

Tabel 1 Hasil Pelatihan *Naïve Bayes* (Metrik Per Kelas)

Label	Precision	Recall	F1-score	Jumlah data
Negatif (-1)	0.69	0.61	0.65	133
Netral (0)	0.78	0.83	0.80	133
Positif (1)	0.73	0.76	0.75	133
Rata-rata (accuracy)	0.73	-	-	73.43 %

Dari hasil di tabel 1 Dapat disimpulkan bahwa model *Naïve Bayes* berhasil mengklasifikasikan komentar dengan akurasi sebesar 73.43%. Nilai presisi dan *recall* yang relatif seimbang menunjukkan model cukup stabil, walaupun masih terdapat kesalahan

dalam membedakan komentar negatif yang bersifat sarkastik.

Pelatihan model SVM Model Support Vector Machine (SVM) digunakan sebagai pembanding karena memiliki kemampuan yang lebih tinggi dalam memisahkan data dengan margin optimal. Hasilnya seperti pada tabel 2 Berikut

Tabel 2 Pelatihan SVM (Metrik Per Kelas)

Label	Precision	Recall	F1-score	Jumlah data
Negatif (-1)	0.88	0.65	0.74	133
Netral (0)	0.84	0.92	0.88	133
Positif (1)	0.78	0.92	0.84	133
Rata-rata (accuracy)	0.83	-	-	82.71 %

Perbandingan kinerja model Perbandingan model seperti pada tabel 3 Berikut

Tabel 3 Perbandingan Ringkas Akurasi & Rata-Rata Metrik Model

Model	Akurasi	Precision rata-rata	Recall rata-rata	F1-score rata-rata
Naïve Bayes	73.43%	0.73	0.73	0.73
SVM	82.71%	0.83	0.83	0.82

Hasil di atas memperlihatkan bahwa: SVM memberikan peningkatan akurasi sebesar 9.28% dibandingkan *Naïve Bayes*.

SVM lebih unggul dalam mengenali pola *sentimen* kompleks, terutama ketika konteks komentar melibatkan ironi atau percampuran emosi.

Naïve Bayes tetap efisien untuk *baseline* model dengan waktu komputasi yang lebih cepat.

Confusion matrix Untuk memahami distribusi kesalahan klasifikasi, dibuat *confusion matrix* seperti pada tabel 4

Tabel 4 Confusion Matrix — *Naïve Bayes* (Nilai)

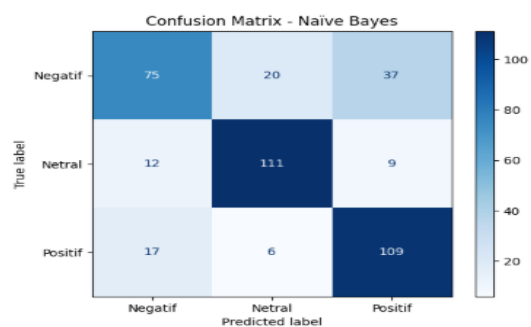
	Prediksi negatif	Prediksi netral	Prediksi positif
Negatif (-1)	81	33	19

Netral (0)	14	110	9
Positif (1)	10	21	102

Tabel 5 *Confusion Matrix* — SVM (Nilai)

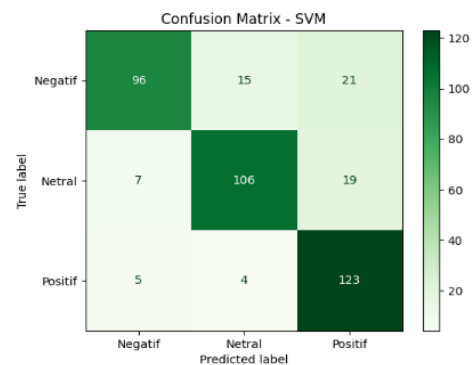
	Prediksi negatif	Prediksi netral	Prediksi positif
Negatif (-1)	86	31	16
Netral (0)	10	122	1
Positif (1)	8	14	111

Hasilnya seperti pada gambar 2 dan gambar 2 berikut.

Gambar 2 Visualisasi Model *Naïve Bayes*

Secara keseluruhan, diagonal *confusion matrix* yang dominan menunjukkan bahwa model *Naïve Bayes* memiliki performa klasifikasi yang baik untuk ketiga kelas *sentimen* tersebut. Namun, masih terdapat beberapa *misclassification* yang dapat disebabkan oleh faktor seperti kemiripan kosakata antar kelas, konteks kalimat yang ambigu, serta keterbatasan representasi fitur yang digunakan.

Analisis *confusion matrix* ini memberikan gambaran komprehensif mengenai kekuatan dan kelemahan model, sehingga dapat menjadi dasar untuk perbaikan, misalnya melalui penambahan data latih, peningkatan metode *preprocessing*, atau penggunaan algoritma pembandingan.

Gambar 3 Visualisasi Confusion Model *Naïve Bayes*

Pada gambar 3 *Confusion matrix* di atas menggambarkan performa model *Support Vector Machine (SVM)* dalam mengklasifikasikan *sentimen* negatif, netral, dan positif. Nilai diagonal menunjukkan prediksi yang benar, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi.

Kelas negatif *SVM* berhasil mengklasifikasikan 96 data sebagai negatif dengan benar. Kesalahan terjadi ketika 15 data diprediksi sebagai netral dan 21 menjadi positif. Namun secara umum, akurasi kelas ini cukup baik.

Kelas netral model mengidentifikasi 106 data secara tepat sebagai netral. Sebanyak 7 data keliru diprediksi sebagai negatif dan 19 sebagai positif. Nilai diagonal yang tinggi menunjukkan *SVM* cukup unggul dalam mengenali pola bahasa netral.

Kelas positif prediksi paling akurat tampak pada kelas positif, dengan 123 data diklasifikasikan dengan benar. Hanya sedikit kesalahan terjadi, yaitu 5 data diprediksi sebagai negatif dan 4 sebagai netral.

Secara keseluruhan, *confusion matrix* menunjukkan bahwa model *SVM* memiliki performa klasifikasi yang lebih stabil dan presisi tinggi pada ketiga kelas dibandingkan model lainnya, ditandai dengan dominasi nilai diagonal yang besar dan tingkat *misclassification* yang rendah.

CONCLUSION

Berdasarkan hasil penelitian mengenai analisis *sentimen* komentar publik Indonesia terhadap video trader *crypto* di YouTube menggunakan algoritma *Multinomial Naïve Bayes* dan *Support Vector Machine (SVM)*, maka dapat disimpulkan beberapa hal sebagai berikut.

Proses Proses pengolahan data komentar publik Indonesia pada video trader *cryptocurrency* di YouTube dapat dilakukan melalui serangkaian tahapan sehingga menghasilkan data yang siap digunakan dalam analisis *sentimen*. Proses tersebut

mampu mengubah data teks yang tidak terstruktur menjadi data yang lebih terorganisir dan dapat dianalisis menggunakan pendekatan *machine learning*.

Berdasarkan hasil perbandingan performa model, algoritma *Support Vector Machine (SVM)* menunjukkan kinerja yang lebih baik dibandingkan *Naïve Bayes* dalam mengklasifikasikan *sentimen* komentar publik. Hal ini ditunjukkan oleh nilai akurasi *SVM* yang lebih tinggi, sehingga *SVM* lebih efektif dalam menangani data teks berdimensi tinggi seperti komentar YouTube.

Hasil analisis sentimen menunjukkan bahwa kecenderungan opini publik terhadap konten trader *cryptocurrency* di YouTube didominasi oleh sentimen negatif, diikuti oleh sentimen netral dan positif. Hal ini mengindikasikan bahwa sebagian besar masyarakat memiliki persepsi yang cenderung kritis terhadap konten trader *crypto* yang beredar di media sosial..

REFERENCE

- [1] Moser, S., & Brauneis, and A., "Should you listen to crypto YouTubers? Finance Research Letters, 54, 103782," <https://doi.org/10.1016/j.frl.2023.103782>, 2023, doi: 10.1016/j.frl.2023.103782.
- [2] Meyer *et al.*, "High on Bitcoin: Evidence of emotional contagion in the YouTube crypto influencer space," *Journal of Business Research*, 164, 113850., 2023.
- [3] M. El Hajj and I. Farran, "The cryptocurrencies in emerging markets: Enhancing financial inclusion and economic empowerment," *Journal of Risk and Financial Management*, vol. 17, no. 10, p. 467, 2024, doi: 10.3390/jrfm17100467.
- [4] Cossu and A., "The unexpected consequences of a pandemic: Crypto-finance as cultural commons," *European Journal of Cultural Studies*, 26(4), 598–607., 2022.