

## ANALISIS SENTIMEN PENGGUNA TOKOCRYPTO SEBAGAI SARANA INVESTASI DI MEDIA SOSIAL X MENGGUNAKAN METODE KNN

Muhammad Raihan Irzi<sup>1</sup>, Alvino Octaviano<sup>2</sup>

<sup>1</sup>Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia, 15417  
e-mail: <sup>1</sup>muhammadrai323@gmail.com

<sup>2</sup>Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia, 15417  
e-mail: <sup>2</sup>dosen00397@unpam.ac.id

### Abstract

*The advancement of digital technology has driven a significant transformation in the investment landscape, particularly through digital assets such as cryptocurrency. Tokocrypto, as one of the local crypto exchanges in Indonesia, has become the focus of social media users, especially on X (formerly Twitter), who often express their opinions about the platform. To understand user sentiment, a sentiment analysis was conducted using 1,386 tweets collected with the keywords “investasi tokocrypto” and “cex tokocrypto”, which were then manually filtered to remove irrelevant data, resulting in a final dataset of 600 tweets. The data was collected through a crawling method and cleaned through preprocessing steps including case folding, normalization, tokenizing, stopword removal, and stemming. After labeling, neutral sentiments were removed, leaving 417 tweets categorized into positive and negative sentiments. The classification process was carried out using the K-Nearest Neighbor (K-NN) method, evaluated across various training and testing ratios and different K values. The best-performing model was achieved with K=7 and an 80:20 ratio, producing the highest performance with accuracy = 74%, precision = 75%, F1-score = 78%, and recall = 81%. The results of this study demonstrate the effectiveness of the K-Nearest Neighbor method in sentiment analysis and provide insights into user perceptions of Tokocrypto as an investment platform on social media platform X.*

**Keywords:** Sentiment Analysis, Crypto Investment, Tokocrypto, Twitter, K-Nearest Neighbor(K-NN)

### Abstrak

Perkembangan teknologi digital telah mendorong transformasi besar dalam dunia investasi, khususnya melalui aset digital seperti cryptocurrency. Tokocrypto, sebagai salah satu bursa kripto lokal di Indonesia, menjadi objek perhatian pengguna media sosial, terutama X (dulu Twitter), yang kerap mengungkapkan opini mereka terkait platform tersebut. Untuk memahami sentimen dari para pengguna, dilakukan analisis sentimen dengan 1386 data *tweets* yang berhasil dikumpulkan menggunakan kata kunci “investasi tokocrypto, cex tokocrypto” lalu di filter manual untuk menghilangkan data yang tidak relevan dan menyisakan 600 dataset. Data dikumpulkan melalui metode *crawling*, kemudian dibersihkan melalui *preprocessing* yang mencakup *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Setelah pelabelan, data netral dihapus sehingga menyisakan 417 *tweets* yang diklasifikasikan sebagai sentimen positif dan negatif. Proses klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) dengan evaluasi pada berbagai rasio data latih dan uji, serta nilai K. Model terbaik diperoleh adalah nilai K=7 dengan rasio 80:20, karena menghasilkan performa terbaik dengan akurasi = 74%, presisi = 75%, *F1-score* tertinggi yaitu 78%, dan nilai K ini memberikan *Recall* = 81% . Hasil dari penelitian ini menunjukkan

efektivitas dari metode *K-Nearest Neighbor* dalam menganalisis sentimen, sekaligus memberikan wawasan mengenai persepsi pengguna terhadap Tokocrypto sebagai sarana investasi di media sosial *Twitter*.

Kata Kunci: Analisis Sentimen, Investasi Crypto, Tokocrypto, *Twitter*, *K-Nearest Neighbor* (*K-NN*)

## 1. PENDAHULUAN

Perkembangan teknologi digital telah membuat perubahan besar pada dunia keuangan global, terutama dalam penampilan aset digital seperti *cryptocurrency*. *Cryptocurrency* adalah mata uang digital menggunakan teknologi *blockchain* sebagai sistem perekaman terdesentralisasi, tanpa partisipasi lembaga keuangan tradisional. Bukan hanya menjadi alat pembayaran saja, *Cryptocurrency* juga di gunakan sebagai instrumen investasi digital, dengan ramainya berita tentang Bitcoin yang menyentuh angka lebih dari 100 ribu dollar membuat banyak orang-orang yang mulai berinvestasi di kripto karena dianggap berpotensi.

Investasi adalah suatu keinginan tentang bagaimana cara menggunakan sebagian dana yang ada atau sumber daya yang dimiliki untuk memperoleh keuntungan yang besar di masa depan [1] Seiring perkembangan teknologi, bentuk investasi pun ikut bertransformasi, salah satunya melalui aset digital seperti *cryptocurrency*. Dan untuk memperjual belikan *cryptocurrency* secara legal tersebut dibutuhkan bursa kripto contohnya bursa kripto global yaitu Binance, Bybit, Gateio dan lainnya. Di Indonesia pun sudah memiliki beberapa bursa kripto lokal, salah satunya adalah Tokocrypto

TokoCrypto adalah sebuah platform perdagangan aset kripto (*cryptocurrency*) yang digunakan untuk membeli, menjual, dan menukar berbagai jenis mata uang digital seperti Bitcoin, Ethereum, dan banyak koin kripto lainnya. Tokocrypto didirikan pada tahun 2018 dan merupakan salah satu platform perdagangan kripto terbesar di Indonesia, serta telah diawasi oleh BAPPEBTI untuk menjamin keamanannya. Meskipun begitu masih banyak pertanyaan-pertanyaan yang muncul dari berbagai macam media sosial.

Media sosial telah menjadi sarana utama masyarakat modern dalam berkomunikasi, berbagi informasi, dan membentuk komunitas [2]. Salah satu platform yang berperan besar adalah *Twitter*, yang kini berganti nama menjadi X. Setelah diakuisisi oleh Elon Musk pada tahun 2022, platform ini mengalami berbagai perubahan baik dari sisi manajemen, kebijakan konten, maupun model bisnis.

Inilah mengapa peran analisis sentimen itu penting. Analisis sentimen adalah proses sistematis untuk mengidentifikasi dan opini publik. Analisis sentimen dipandang sebagai alat bisnis intelijen yang diinginkan, karena kemampuannya dapat mengekstrak

opini publik tentang topik tertentu, produk atau jasa yang tertanam dalam teks-teks yang tidak terstruktur [3]

Dalam konteks penelitian ini, analisis sentimen digunakan untuk mengeksplorasi sikap atau kesadaran pengguna Tokocrypto sebagai sarana investasi. Memerlukan algoritma yang mendukung sentimen ini, maka digunakan algoritma *K-Nearest Neighbor* (KNN), metode *K-Nearest Neighbor* (KNN) merupakan metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain, metode KNN (*K-Nearest Neighbor*) merupakan metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi. [4]

## 2. PENELITIAN YANG TERKAIT

Berikut adalah jurnal yang terkait dengan Analisis Sentimen Pengguna Tokocrypto Sebagai Sarana Investasi di Media Sosial X Menggunakan Metode KNN.:

- Pada penelitiannya dengan judul “Analisis Sentimen Masyarakat Terhadap Exchange Tokocrypto Pada Twitter Menggunakan Metode LSTM”. Tujuan penelitian ini adalah untuk menganalisis opini masyarakat terhadap Tokocrypto melalui data Twitter serta menguji performa metode LSTM dalam mengklasifikasikan sentimen tersebut Dengan hasil Dari 5000 tweets yang ada mengenai keyword tokocryptodalam bahasa Indonesia, ditemukan sebanyak 2022 kata positif, 1632 kata negatif, dan 1012 kata netral [5].
- Pada penelitiannya dengan judul “Analisis Sentimen Tokocrypto pada Twitter menggunakan Metode Long Short-Term Memory”. Penelitian ini bertujuan untuk Menganalisis sentimen pengguna Twitter terhadap Tokocrypto menggunakan metode Bidirectional LSTM untuk mengetahui opini publik (positif atau negatif) terhadap layanan Tokocrypto. Dengan hasil Model Bidirectional LSTM berhasil mengklasifikasikan sentimen dengan baik, ditunjukkan oleh nilai akurasi sebesar 95%, f1-score sebesar 0.95, precision sebesar 0.98, dan recall sebesar 0.92 [6].
- Pada penelitiannya dengan judul “Analisis Sentimen Pada Ulasan Aplikasi Indodax Di Google Play Store Menggunakan Metode Support Vector Machine”. Penelitian ini bertujuan untuk menganalisis opini pengguna terhadap aplikasi Indodax di Google Play Store serta menguji efektivitas metode Support Vector Machine dalam klasifikasi sentimen. Dengan

hasil Hasil penelitian menunjukkan bahwa klasifikasi sentimen ulasan pengguna aplikasi Indodax menggunakan metode Support Vector Machine (SVM) menghasilkan akurasi tertinggi sebesar 85% pada rasio data 80:20, dengan temuan utama berupa sentimen negatif terkait biaya penarikan tinggi dan lambatnya server saat transaksi, serta sentimen positif yang menyoroti kemudahan penggunaan dan kecepatan layanan aplikasi [7].

### 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dikarenakan data yang digunakan untuk analisis berupa teks yang diubah menjadi data numerik dan dianalisis secara statistik. Pendekatan ini digunakan untuk mengukur dan mengklasifikasikan sentimen-sentimen pengguna Tokocrypto di media sosial *X* atau *Twitter* menjadi kategori positif dan negatif. Dengan algoritma *K-Nearest Neighbor*, penelitian ini bersifat eksperimen, dengan melakukan tahapan-tahapan seperti *preprocessing*, pembentukan fitur TF-IDF, dan evaluasi metrik seperti akurasi, *recall*, dan *Precision*.

Penelitian ini menggunakan data berupa kumpulan *tweets* berbahasa Indonesia yang diperoleh dari pendapat para pengguna *Twitter*. Data *tweets* ini didapatkan melalui program *crawling* dengan menggunakan *Tweet Harvest*. Dalam proses *crawling*, data *tweets* yang diambil merupakan data yang berisi kata-kata “Investasi di Tokocrypto, cex Tokocrypto” data di ambil dari tanggal 4 September 2023 hingga 1 Februari 2024. Setelah terkumpul, data akan melalui tahapan *preprocessing* kemudian akan diklasifikasikan ke dalam dua kategori yaitu, sentimen positif dan sentimen negatif.

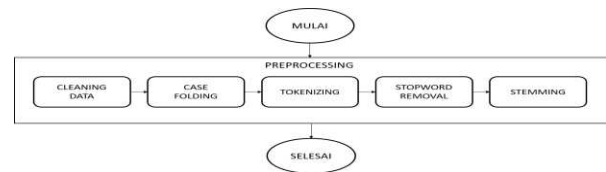
### 4. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data yang telah dikumpulkan menggunakan *crawling* data pada media sosial twitter dengan total data mentah 1386, lalu dataset tersebut difilter secara manual untuk membuang data-data yang tidak relevan dan menyisakan 600 dataset yang siap digunakan untuk mengimplementasikan metode penelitian ini. Lalu dataset tersebut akan melalui berbagai tahap seperti *preprocessing* data, pelabelan otomatis menggunakan *lexicon* Bahasa Indonesia, mengidentifikasi koin yang paling banyak dibicarakan, pembentukan fitur (TF-IDF), klasifikasi algoritma KNN.

#### a. Preprocessing Data

*Text preprocessing* adalah sebuah proses untuk mengolah data teks sehingga menjadi lebih terstruktur, sehingga data siap untuk dianalisis lebih lanjut[8]. Tahap awal yang dilakukan yaitu melakukan pembersihan data dimana data yang telah didapatkan

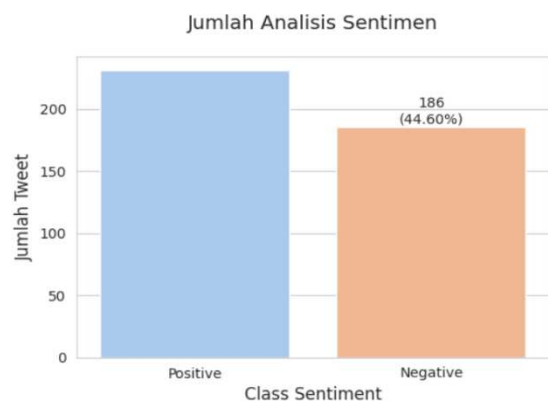
digabung menjadi satu *dataset*. Pada pembersihan data akan dilakukan yaitu *Case Folding*, *Cleaning Data*, *Tokenizing*, *Stopword Removal*, *Stemming*.



Gambar 6. 1 Tahapan *Preprocessing*

#### b. Pelabellan dengan *Lexicon*

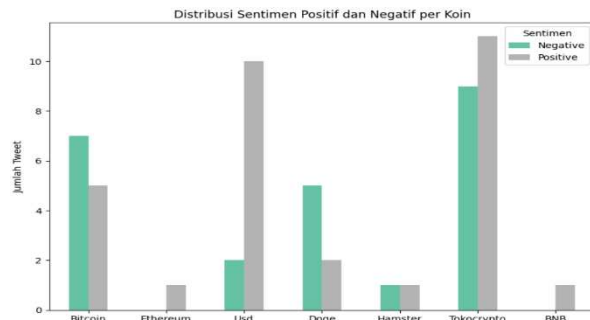
*Lexicon* adalah sebuah metode yang dapat digunakan untuk menganalisis sebuah sentimen dengan pendekatan berbasis kamus. Berikut hasil pelabelan dengan *Lexicon*. *Lexicon* yang digunakan dalam penelitian ini adalah *InSet Lexicon* karena sudah teruji cukup baik untuk analisis sentimen data berbahasa Indonesia. *InSet Lexicon* (*Indonesia Sentiment lexicon*) terdiri dari 3.609 kata positif dan 6.609 kata negatif berbahasa Indonesia yang telah memiliki bobot nilai atau *polarity score* pada setiap katanya dengan kisaran bobot antara -5 sampai +5. Setelah *dataset* diberikan label secara otomatis menggunakan kamus *lexicon*, data *tweets* dengan sentimen netral dihapus sehingga hanya menyisakan data dengan sentimen positif dan negatif yang menyisakan data sebanyak 417 data. Berikut *bar chart* untuk total label sentimen yang didapat:



Gambar 6. 2 Setelah Melalui Tahap Pelabelan

c. Identifikasi Koin Yang Paling Banyak Dibicarakan  
Dataset yang telah melalui tahap *pre processing* di analisis manual untuk mengidentifikasi *tweet* yang mengandung kata kunci yang berkaitan dengan setiap produk yang dibicarakan. Kata kunci ini ditentukan berdasarkan nama-nama produk/koin yang dibicarakan serta variasi penulisannya, misalnya “bitcoin”, “btc”, “tko”, “ethereum” dan sebagainya. Jika tidak di temukan kata kunci yang cocok, maka *tweet* tersebut akan

dikategorikan sebagai “Tidak Terkait” dan dihapus dari tahap ini. Lalu sentimen positif dan negatif dianalisis terpisah untuk setiap produk/koin, sehingga dapat diketahui distribusi sentimen yang mengarah ke setiap produk yang dibicarakan.



**Gambar 6.3** Distribusi Sentiment Pada Tiap Koin.

#### d. Pembentukan Fitur (TF-IDF)

Pembentukan fitur TF-IDF (Term Frequency – Inverse Document Frequency) adalah salah satu teknik dalam text mining / NLP untuk mengubah teks menjadi representasi numerik (vektor) agar bisa diolah oleh algoritma machine learning. Tujuannya adalah memberi bobot pada setiap kata dalam dokumen sehingga kata yang sering muncul di dokumen tertentu tapi jarang muncul di seluruh koleksi dokumen dianggap lebih penting.

$$TF\ t, d = \frac{\text{frekuensi kemunculan term } t \text{ pada dokumen } d}{\text{total dokumen } d}$$

*Document Frequency*, merupakan total jumlah dokumen yang berisi perhitungan IDF. Perhitungan IDF dapat dilakukan dengan persamaan berikut:

$$idf\ t = \frac{1}{df\ t}$$

$$idf\ t = \ln \frac{N}{df\ t}$$

**Gambar 6.4** Rumus Pembentukan TF-IDF

#### e. Klasifikasi Algoritma KNN

Pendekatan KNN ini melakukan klasifikasi terhadap suatu objek berdasarkan pada kesamaan atau jarak terdekat dengan objek-objek dalam data latih (*training*) [9]. Pada tahap ini, algoritma *K-Nearest Neighbor* (KNN) digunakan untuk mengklasifikasikan sentimen *tweet* terkait sentimen pengguna Tokocrypto. Data dibagi menjadi data latih dan data uji dengan empat rasio berbeda yaitu 60:40, 70:30, 80:20, 90:10. Proses klasifikasi dilakukan dengan menggunakan nilai k sebesar 3, 5, 7 dan 9 untuk membandingkan mana yang lebih unggul dari keempat rasio yang berbeda-beda tersebut. Berikut proses klasifikasi algoritma *K-Nearest Neighbor* (K-NN) dengan menggunakan *python*.

**Tabel 6.1** Tabel Nilai K Terbaik

Rasio	Nilai K	Accuracy	Precision	Recall	F1-Score
60:40	9	0.72	0.72	0.81	0.76
70:30	3	0.69	0.74	0.72	0.70
80:20	7	0.74	0.75	0.81	0.78
90:10	3	0.71	0.69	0.87	0.77

60:40	9	0.72	0.72	0.81	0.76
70:30	3	0.69	0.74	0.72	0.70
80:20	7	0.74	0.75	0.81	0.78
90:10	3	0.71	0.69	0.87	0.77

Berdasarkan pengujian yang telah dilakukan menggunakan data uji dengan rasio 80:20 dan nilai k=7, di mana data uji terdiri dari 80 *tweets* (20% dari dataset), dan data latih yang terdiri dari 337 *tweets* (80% dari dataset) yang digunakan untuk membangun model klasifikasi dan menghasilkan performa terbaik dari semua nilai k yang telah di uji dengan hasil akurasi = 74%, presisi = 75%, *F1-score* tertinggi yaitu 78%, dan nilai K ini memberikan *Recall* = 81% .

## 5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan mengenai analisis sentimen terkait Tokocrypto sebagai sarana investasi menggunakan metode *K-Nearest Neighbor* (KNN) yang telah selesai dilaksanakan, maka diperoleh hasil analisis yang memberikan gambaran mengenai persepsi masyarakat. Menurut hasil analisis sentimen ini , diperoleh kesimpulan sebagai berikut.

- Berdasarkan hasil penelitian ini, jumlah sentimen positif terkait Tokocrypto sebagai sarana investasi memperoleh hasil persentase sebanyak 55.40% atau sebanyak 231 data, sementara persentase dari sentimen negatif yang di dapat yaitu sebanyak 44.60% atau sebanyak 186 data dari total dataset. Hal ini menunjukkan bahwa mayoritas pengguna cenderung memberikan respon positif terhadap Tokocrypto.
- Analisis mendalam mengenai produk/koin yang paling sering di bicarakan menunjukkan data dengan sentimen positif terbanyak yaitu dari produk *native* Tokocrypto yaitu TKO dengan pujian terkait keamanan serta sebagai koin dengan peluang investasi yang menjanjikan. Sedangkan komentar dengan sentimen negatif ditunjukkan ke koin Bitcoin yang menyoroti kekhawatiran terhadap adopsi bitcoin di Indonesia, dan volatilitas pasar. Sementara itu sentimen positifnya yang tertinggi serta negatifnya terendah ada di USD yang dominan membicarakan tentang ke stabilannya dan dapat diandalkan sebagai instrumen transaksi dan keamanannya di tengah volatilitas pasar kripto.
- Berdasarkan hasil penelitian ini, hasil pengujian menggunakan metode *K-Nearest Neighbor* (KNN) untuk analisis sentimen terhadap Tokocrypto sebagai sarana investasi di media sosial *Twitter (X)* yang menunjukkan kinerja model yang cukup baik dengan



nilai  $k=7$  pada rasio 80:20 dengan akurasi tertinggi yaitu 74%, dan presisi yang didapat adalah 75%, *recall* 81%, dan *F1-score* nya sebesar 78%. Hasil dari penelitian ini menunjukkan bahwa model *K-Nearest Neighbor* mampu melakukan klasifikasi sentimen dengan performa yang cukup memadai terkait Tokocrypto sebagai sarana investasi.

#### DAFTAR PUSTAKA

- [1] E. Suyanti and N. U. Hadi, "Analisis Motivasi Dan Pengetahuan Investasi Terhadap Minat Investasi Mahasiswa di Pasar Modal," *Edunomic Jurnal Pendidikan Ekonomi*, vol. 7, no. 2, p. 108, Sep. 2019, doi: 10.33603/EJPE.V7I2.2352.
- [2] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 479, Apr. 2023, doi: 10.30865/JURIKOM.V10I2.5945.
- [3] S. S. Salim and J. Mayary, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP DOMPET ELEKTRONIK DENGAN METODE LEXICON BASED DAN K – NEAREST NEIGHBOR," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 1, pp. 1–17, Apr. 2020, doi: 10.35760/IK.2020.V25I1.2411.
- [4] I. Yolanda, H. Fahmi, and [ Penerapan, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT.Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 3, no. 1.1, pp. 9–15, 2020, doi: 10.9767/JIKOMSI.V3I1.1.83.
- [5] G. A. Sandag and J. Waworundeng, "Public Sentiment Analysis Against Tokocrypto Exchange on Twitter Using LSTM Method," *CogITO Smart Journal*, vol. 8, no. 2, pp. 411–421, Dec. 2022, doi: 10.31154/COGITO.V8I2.418.411-421.
- [6] C. Rio, A. Toruan, N. Yudistra, and R. S. Perdana, "Analisis Sentimen Tokocrypto pada Twitter menggunakan Metode Long Short-Term Memory," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 2, pp. 719–726, Mar. 2023, Accessed: May 27, 2025. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12293>
- [7] A. Kusuma, E. Ermatita, and H. N. Irmanda, "Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine," *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, vol. 3, no. 2, pp. 773–782, Aug. 2022, Accessed: May 27, 2025. [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/2214>
- [8] H. Sibyan and N. Hasanah, "ANALISIS SENTIMEN ULASAN PADA WISATA DIENG DENGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)," *Jurnal Penelitian dan Pengabdian Kepada Masyarakat UNSIQ*, vol. 9, no. 1, pp. 38–47, Jan. 2022, doi: 10.32699/PPKM.V9I1.2218.
- [9] D. Ardiyansyah and N. Oktafiani, "PERBANDINGAN METODE PENGUKURAN JARAK PADA K-NEAREST NEIGHBOUR DALAM KLASIFIKASI DATA TEKS CARDIOVASKULAR," *Journal of Information Systems Management and Digital Business*, vol. 1, no. 2, pp. 116–122, Jan. 2024, doi: 10.59407/JISMDB.V1I2.260.