

Analisis Sentimen Pengguna Media Sosial X terhadap Perubahan Harga *Bitcoin*: Pendekatan *Machine Learning*

J E Savero^{*1}, V H Pranatawijaya², E Christian³

¹⁻³Program Studi Teknik Informatika, Universitas Palangka Raya

E-mail: joshuaevan@mhs.eng.upr.ac.id^{*1}, viktorhp@it.upr.ac.id², efrans@it.upr.ac.id³

Abstrak. Media sosial X menjadi gudang data yang dapat dimanfaatkan untuk memperoleh wawasan mengenai sentimen publik dan potensi yang berdampak pada harga *cryptocurrency*. Dalam beberapa tahun terakhir, *bitcoin* menjadi pusat perhatian sebagai bentuk investasi yang menarik bagi para pelaku pasar. Bitcoin (BTC) sering kali ditandai dengan tingkat volatilitas yang tinggi dan harganya menunjukkan kenaikan dan penurunan yang ekstrem dalam jangka waktu yang singkat. Dengan menganalisis *tweet* pengguna media sosial X, penelitian ini bertujuan untuk meneliti hubungan antara sentimen yang diungkapkan oleh pengguna media sosial X dan perubahan harga *bitcoin*. Data set yang digunakan dalam penelitian ini yaitu dataset pelatihan model yang terdapat di laman Kaggle dan dataset pengujian yang dikumpulkan dari *tweet* media sosial X berdasarkan tanggal terjadinya *golden cross* dan *death cross*. Data set akan melalui teknik *preprocessing data*, klasifikasi sentimen positif, negatif, dan netral menggunakan VADER. Pembangunan model menggunakan algoritma *naïve bayes* dan *support vector machine*. Hasil penelitian ini memperoleh model *support vector machine* memiliki kinerja terbaik terhadap keakuratan model dalam klasifikasi sentimen dengan *accuracy* sebesar 95.92%, ketepatan model dalam memprediksi nilai positif dengan tingkat *precision* sebesar 95.89%, tingkat usaha dalam menemukan informasi kembali dengan tingkat *recall* sebesar 95.92%, dan presentasi nilai bobot dari nilai *precision* dengan nilai *recall* pada *f1-score* sebesar 95.89%. Akan tetapi, dalam memprediksi sentimen lima dataset pengujian yang diberikan menggunakan model yang telah dilatih ditemukan algoritma *naïve bayes* memiliki persentase lebih tinggi yaitu 80% dalam memperoleh hasil yang sesuai antara sentimen positif untuk kondisi *golden cross* dan sentimen negatif untuk kondisi *death cross*.

Kata kunci: analisis sentimen; *bitcoin*; *machine learning*

Abstract. *Social media X is a data warehouse that can be utilized to gain insight into public sentiment and its potential impact on cryptocurrency prices. In recent years, Bitcoin has become the center of attention as an attractive form of investment for market players. Bitcoin (BTC) is often characterized by high levels of volatility and its price exhibits extreme rises and falls over short periods of time. By analyzing the tweets of social media user X, this study aims to examine the relationship between the sentiment expressed by social media user X and changes in Bitcoin prices. The dataset used in this research is the model training dataset found on the Kaggle page and the testing dataset collected from X's social media tweets based on the dates of the golden cross and death cross. The dataset will go through data preprocessing techniques, classifying positive, negative and neutral sentiment using VADER. Model construction uses the Naïve Bayes algorithm and Support Vector Machine. The results of this research show that the Support Vector Machine model has the best performance regarding model accuracy in sentiment classification with an accuracy of 95.92%, model accuracy in predicting positive values with a precision level of 95.89%, level of effort in finding information again with a recall rate of 95.92%, and presentation of the weighted value of the precision value with the recall value on*

the *f1*-score of 95.89%. However, in predicting the sentiment of the five test datasets provided using the trained model, it was found that the Naïve Bayes algorithm had a higher percentage, namely 80%, in obtaining results that matched positive sentiment for the golden cross condition and negative sentiment for the death cross condition.

Keywords: sentiment analysis; bitcoin; machine learning

1. Pendahuluan

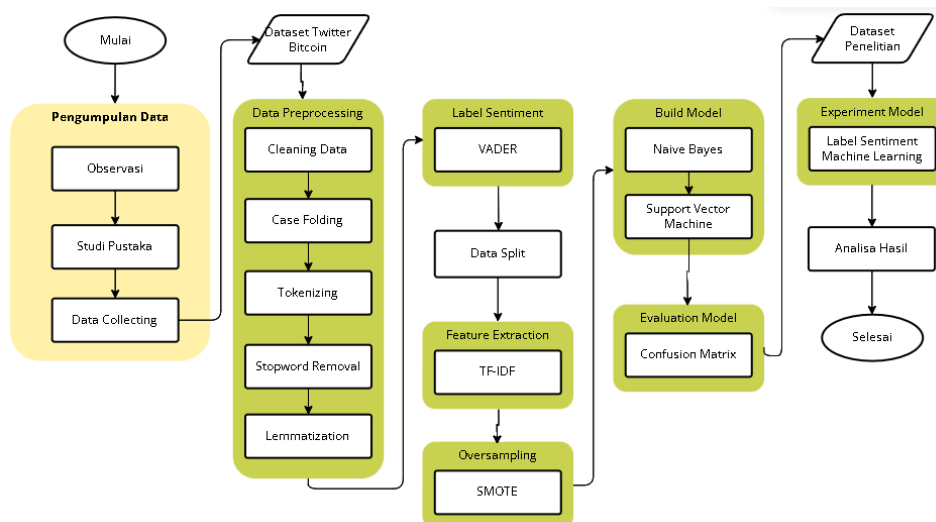
Dalam perkembangan era digital saat ini, data berperang penting. Hal ini dikarenakan volume data yang terus meningkat setiap hari, termasuk dari sumber data media sosial seperti media sosial X yang menjadi semakin populer untuk mengekspresikan opini dan sentimen publik mengenai berbagai topik, termasuk pasar keuangan. Platform ini menyediakan gudang data yang dapat dimanfaatkan untuk mendapatkan wawasan mengenai sentimen publik dan potensi yang berdampak terhadap harga saham [1]. Dalam beberapa tahun terakhir, *bitcoin* telah menjadi pusat perhatian sebagai bentuk investasi yang menarik bagi para pelaku pasar. Sebagai salah satu *cryptocurrency* paling terkemuka, *bitcoin* berbeda dari perbankan *online* tradisional lainnya dengan menggunakan skema konsensus terdesentralisasi untuk memverifikasi kebenaran dan keaslian transfer mata uang antar pengguna [2] serta ketiadaan pengaturan oleh otoritas pusat. Bitcoin (BTC) sering kali ditandai dengan tingkat volatilitas yang tinggi dan harganya menunjukkan kenaikan dan penurunan yang ekstrem dalam jangka waktu yang singkat [3] sehingga menimbulkan kekhawatiran dan ketidakpastian bagi para investor. Dalam konteks ini, analisis sentimen media sosial X menjadi sebuah alat yang penting dalam memahami opini dan sentimen publik yang berkembang seiring waktu serta dampaknya pada harga *bitcoin*. Dengan menganalisis *tweet* pengguna media sosial X, penelitian ini bertujuan untuk meneliti hubungan antara sentimen yang diungkapkan oleh pengguna media sosial X dan perubahan harga *bitcoin*, sehingga memberikan wawasan yang lebih dalam kepada investor dan pelaku pasar untuk membuat keputusan yang lebih baik dalam mengelola investasi mereka.

Terdapat beberapa penelitian yang telah dipublikasi pada analisis sentimen menggunakan algoritma *machine learning* yang beragam. Dalam penelitian oleh [4] melakukan klasifikasi sentimen pada media sosial X yang memuat kata kunci *#bitcoin*, *#cardano*, *#ethereum*, *#litcoin*, *#ripple*. Pada pengujian sistem model *k-NN*, *decision tree*, *naïve bayes*, *neural network* dan *support vector machine* diperoleh SVM menghasilkan akurasi terbesar yaitu 82%. Penelitian menemukan bahwa promosi *cryptocurrency* yang positif melalui media sosial X dapat mempengaruhi persepsi masyarakat secara positif. Selain itu, perusahaan yang berinvestasi pada *cryptocurrency* dapat memanfaatkan platform tersebut sebagai alat promosi yang efektif. Penelitian lainnya oleh [5] melakukan klasifikasi sentimen berita perusahaan Apple Inc. dalam rentang waktu 3 tahun dimulai 1 Februari 2013 sampai 2 April 2016. Pada pengujian sistem model *random forest*, *support vector machine* dan *naïve bayes* diperoleh *random forest* menghasilkan akurasi terbesar yaitu 88%. Dalam penelitian ini, ditemukan tren saham dapat diprediksi menggunakan artikel berita dan riwayat sebelumnya, di mana berita positif memiliki dampak positif bagi pasar, sementara berita negatif cenderung mempengaruhi penurunan harga dalam tren. Selain itu, penelitian oleh [6] melakukan klasifikasi sentimen postingan media sosial X berdasarkan kata kunci “\$btc, #bitcoin” rentang waktu 1 Januari sampai 31 Mei 2018 dan data harga *bitcoin* rentang waktu 8 Februari sampai 31 Maret 2018. Pada hasil pengujian sistem menggunakan CNN-LSTM, CNN, LSTM, *naïve bayes* dan *support vector machine* diperoleh CNN-LSTM menghasilkan akurasi terbesar yaitu 88,7%. Dalam penelitian ini, peneliti menerapkan korelasi *Pearson* dan menemukan bahwa sentimen hari ini (*d*) memiliki efek positif terhadap *return bitcoin* di masa depan pada hari berikutnya (*d+1*).

Dalam pengembangan *machine learning* analisis sentimen, penulis menggunakan pelabelan sentimen *vader lexicon*. Lalu, menggunakan model *naïve bayes* dan *support vector machine*. Penelitian ini bertujuan untuk meneliti hubungan antara sentimen yang diungkapkan oleh pengguna media sosial X dan perubahan harga *bitcoin*. Dengan memahami bagaimana sentimen ini terbentuk dan berubah seiring waktu, investor dan pelaku pasar dapat membuat keputusan yang lebih baik dalam mengelola investasi mereka.

2. Metode

Terdapat beberapa metode yang digunakan yaitu pengumpulan data, data *preprocessing*, label sentimen, data *split*, *feature extraction*, *oversampling*, *build model*, *evaluation model*, *experiment model* dan analisis hasil. Lebih lengkapnya dapat dilihat pada Gambar 1.



Gambar 1. Alur Metode Penelitian

Pada tahap awal penulis mengamati tren pasar di *TradingView* untuk menemukan *death cross* dan *golden cross* pada harga *bitcoin*. *Golden cross* terjadi ketika garis rata-rata pergerakan jangka pendek melintasi garis rata-rata pergerakan jangka panjang dari bawah ke atas. Fenomena ini menunjukkan bahwa lebih banyak investor yang ingin membeli daripada menjual, menandakan potensi kenaikan harga dan terindikasi aman untuk membeli. Sebaliknya, *death cross* terjadi ketika garis rata-rata pergerakan jangka pendek turun di bawah garis rata-rata pergerakan jangka panjang. Ini menunjukkan tren penjualan yang sedang berlangsung, menandakan bahwa lebih banyak investor yang menjual dibandingkan membeli, dan menjadi peringatan untuk menjual karena harga kemungkinan besar akan turun. Rata-rata pergerakan 50 hari ditandai sebagai Rata-rata Pergerakan Eksponensial (EMA) jangka pendek dan rata-rata pergerakan 200 hari sebagai Rata-rata Pergerakan Eksponensial (EMA) jangka panjang [7].

Lalu, pengumpulan data berupa jurnal dari *website* yang berhubungan dengan topik penelitian. Selanjutnya, pengumpulan data uji menggunakan *plugin instant data scraper* dari Microsoft Edge untuk melakukan *scraping* data *tweet* media sosial X berdasarkan terjadinya *golden cross* dan *death cross*. *Web scraping* adalah proses mengekstrak informasi teks yang berharga dan menarik dari halaman web, sehingga dapat diandalkan untuk mengumpulkan data dalam jumlah besar dengan cepat dan efisien [8]. Sedangkan, data latih menggunakan dataset *Bitcoin Sentiment Analysis* yang telah disediakan dari laman Kaggle.

Setelah pengumpulan data, selanjutnya yaitu *data preprocessing* merupakan tahap mempersiapkan teks yang dapat diolah lebih lanjut [9]. Penelitian ini menggunakan lima tahap *preprocessing* yaitu pembersihan data, *case folding*, *tokenizing*, *stopwords*, dan *lemmatization* [10]. Tahap awal melibatkan pembersihan teks dengan menghapus karakter tidak relevan seperti nama pengguna, spasi kosong, karakter khusus, dan URL. Lalu, *case folding* untuk mengonversikan setiap teks menjadi huruf kecil. Selanjutnya *tokenizing*, yaitu mengonversikan setiap kalimat menjadi per kata, dan dimuat dalam *stopwords* untuk membersihkan kata yang tidak memiliki makna. Terakhir, *lemmatization* menghilangkan imbuhan kata menjadi kata dasar.

Lalu, proses pemberian kelas sentimen untuk setiap data setelah melalui tahap *preprocessing*. Penulis menggunakan *Vader Lexicon* untuk menilai kalimat berdasarkan *compound* dari kata-kata yang terdapat di data [11]. Dan data tersebut akan dibagi untuk menunjang pembangunan model yaitu 80% sebagai data latih dan 20% sebagai data uji. Data sampel terbagi menjadi variabel X yaitu data *preprocessing* yang akan dikonversi ke *vectorizer* dan variabel Y yaitu data *sentiment vader*. *Feature*

extraction adalah tahap mengonversikan teks menjadi bentuk numerik (*vectorizer*) supaya dapat dikenali algoritma *machine learning* dalam mengklasifikasi teks. Dalam tahap ini, peneliti melakukan pembobotan kata berdasarkan frekuensi kemunculan kata di dalam teks menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) [12].

Selanjutnya, *oversampling* dilakukan sebagai tahap menangani ketidakseimbangan kelas sentimen terhadap dataset yang terjadi ketika jumlah sampel dalam satu kelas lebih besar daripada jumlah sampel dalam kelas lainnya sehingga perlu menciptakan lebih banyak sampel bagi kelas yang masih kurang [13]. *Oversampling* dilakukan dengan menggunakan SMOTE. Setelah data seimbang, maka data akan digunakan pada pembangunan model dengan mengimplementasikan algoritma *supervised learning* yaitu *naïve bayes* dan *support vector machine*. Model yang telah dibangun akan dilakukan tahap *evaluation model* untuk menilai kinerja dari setiap model *machine learning* dalam melakukan klasifikasi teks ke dalam kelas sentimen tertentu. *Evaluation model* menggunakan *confusion matrix* supaya dapat menghitung parameter *accuracy*, *precision*, *recall* dan *F1-score* [14].


Selanjutnya, tahap *experiment model* untuk mengetahui bagaimana cara model *machine learning* melakukan klasifikasi sentimen terhadap data yang belum pernah dilihat sebelumnya. Algoritma *machine learning* yang digunakan untuk melakukan klasifikasi adalah algoritma yang telah dilatih sebelumnya. Hal ini dilakukan untuk mengetahui kelas sentimen yang dihasilkan pada setiap dataset yang dimasukkan. Tahap terakhir yaitu analisis hasil dilakukan untuk melakukan perbandingan harga *bitcoin* berdasarkan terjadinya *golden cross* ataupun *death cross* melalui *TradingView* terhadap hasil sentimen yang dihasilkan pada *experiment model*. Hal ini dilakukan untuk mengetahui apakah hasil sentimen pada dataset yang dimasukkan memberi pengaruh terhadap perubahan harga *bitcoin*.

3. Hasil dan Pembahasan

3.1 Scraping Data

Pada tahap pertama adalah pengumpulan data pelatihan dari dataset laman Kaggle yang disediakan oleh Furkan Golsen. Data ini berisikan data *tweet* tentang *bitcoin* berjumlah 48.583 data. Data dimuat menjadi tiga kolom yaitu *user_name*, *date*, dan *text*. Untuk selengkapnya dapat dilihat pada Tabel 1.

Tabel 1. Hasil Data set Laman Kaggle

No	User_Name	Date	Text
1	DeSota Wilson	10/02/2021 23:59:00	Blue Ridge Bank shares halted by NYSE after #bitcoin ATM announcement https://t.co/xaaZmaJKiV @MyBlueRidgeBankâ€¹ https://t.co/sgBxMkP1SI
2	Alex Kirchmaier ATSE #FactsSuperspreader	2/10/2021 23:54:06	This network is secured by 9 508 nodes as of today. Soon, the biggest bears will recognise: #BTC in too big to fail... https://t.co/1XovDA8rKw
3	WSB Crypto Mod 	2/10/2021 23:32:14	Institutions are buying #Bitcoin Over \$100 million worth of \$BTC, got moved to cold storage just today! Don't... https://t.co/FLAxRkyQMg
4	Roman Wiligut	2/10/2021 17:36:25	The meritocratic model of government shows one of the best, as does technocracy. #BTC #Bitcoin #Crypto #Blockchain
5	Humayun MHA	2/10/2021 16:17:31	Watching the Bitcoin price on 1-min candles is my new

Pada tahap kedua adalah pengumpulan data pengujian menggunakan plugin *Instant Data Scraper* dari *tweet* pengguna media sosial X pada postingan @Bitcoin berdasarkan terjadinya *golden cross* dan *death cross* dari 1 Mei 2023 sampai 31 Maret 2024. Salah satu contoh dataset *golden cross* yang terjadi pada tanggal 29 Mei 2023 dapat dilihat lebih jelas pada Tabel 2.

Tabel 2. Hasil Data Set Tweet Laman Media Sosial X

No	User Name	Date	Text
1	@lopp	May 31, 2023	Don't dox me, bro!
2	@marshallmixing	May 31, 2023	People who have driven on 280 before I preferred it back in the day when it was all painted the same bland color (looked like it was made from Silly Putty), but the addition of the dinos is pretty trick.
3	@scratchedpaint	May 31, 2023	Did they "read" the blueprint before building that
4	@ChrisBTC2000	May 31, 2023	When reach 3000
5	@Amusuk8	Jun 5, 2023	

3.2 Data Preprocessing

Setelah pengumpulan dataset untuk pelatihan model, kemudian dilakukan tahapan *preprocessing* untuk membersihkan dan memproses data teks agar siap untuk dianalisis. Berikut tahapan *preprocessing* ditampilkan di Tabel 3.

Tabel 3. Hasil Data Preprocessing

Data Awal	Data Mentah	Data Preprocessing
Data Cleansing	Blue Ridge Bank shares halted by NYSE after #bitcoin ATM announcement https://t.co/xaazmaJKiV x@MyBlueRidgeBank... https://t.co/sgBxMkP1SI	Blue Ridge Bank shares halted by NYSE after bitcoin ATM announcement MyBlueRidgeBank
Case Folding	Blue Ridge Bank shares halted by NYSE after bitcoin ATM announcement MyBlueRidgeBank	blue ridge bank shares halted by nyse after bitcoin atm announcement myblueridgebank
Tokenizing	blue ridge bank shares halted by nyse after bitcoin atm announcement myblueridgebank	[blue, ridge, bank, shares, halted, by, nyse, after, bitcoin, atm, announcement, myblueridgebank]
Stopword Removal	[blue, ridge, bank, shares, halted, by, nyse, after, bitcoin, atm, announcement, myblueridgebank]	[blue, ridge, bank, shares, halted, nyse, bitcoin, atm, announcement, myblueridgebank]
Lemmatization	[blue, ridge, bank, shares, halted, nyse, bitcoin, atm, announcement, myblueridgebank]	blue ridge bank share halt nyse bitcoin atm announcement myblueridgebank

3.3 Label Sentimen

Dalam tahap ini, pelabelan sentimen dilakukan menggunakan *Vader Lexicon*. Pada proses ini dihasilkan 22.363 data positif, 5.620 data negatif, dan 20.253 data netral dari total 48.583 data. Berikut data yang telah diberi label dapat dilihat pada Tabel 4.

Tabel 4. Pelabelan Vader

Teks	Compound	Sentimen
<i>blue ridge bank share halt nyse bitcoin atm announcement myblueridgebank</i>	0.296	Positif
<i>network secure node today soon big bear recognise btc big fail</i>	-0.2732	Negatif
<i>institution buy bitcoin million worth btc get move cold storage today</i>	0.2263	Positif
<i>meritocratic model government show one best technocracy btc bitcoin crypto blockchain</i>	0.4404	Positif
<i>watch bitcoin price min candle new drug choice bitcoin btc</i>	0.0	Netral

3.4 Data Split

Pada tahap data *split* penulis menggunakan *library train_test_split* untuk membagi data *train* sebesar 80 % untuk identifikasi pola dalam data dan data tes sebesar 20% untuk evaluasi model dalam memprediksi. Hasil pembagian data ditampilkan pada Tabel 5.

Tabel 5. Pembagian Data

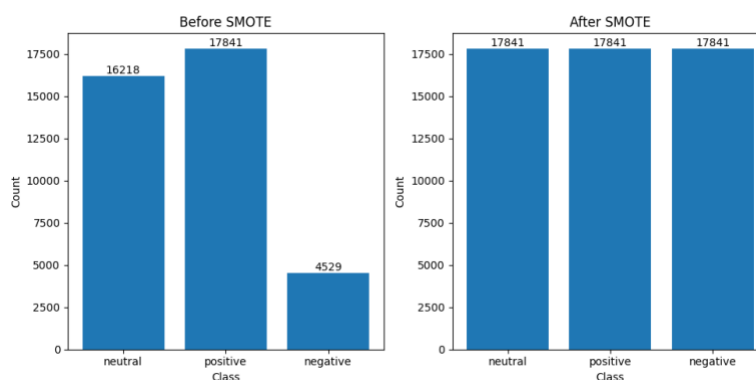
Metode	Sentimen	Data Latih (80%)	Data Uji (20%)
<i>Vader Lexicon</i>	Positif	17841	4522
	Negatif	4529	1091
	Netral	16218	4035

3.5 Feature Extraction

Tahap *feature extraction* menggunakan *TfidfVectorizer* untuk mengubah data teks menjadi bentuk vektor melalui pembobotan kata terhadap data dokumen teks sehingga dapat dikenali oleh model.

3.6 Oversampling

Oversampling menggunakan SMOTE supaya membuat kelas sintesis pada kelas minoritas menggunakan sampel dari kelas mayoritas sehingga data menjadi seimbang. Tampilan distribusi data *oversampling* ditampilkan di Gambar 2.



Gambar 2. Hasil *Oversampling*

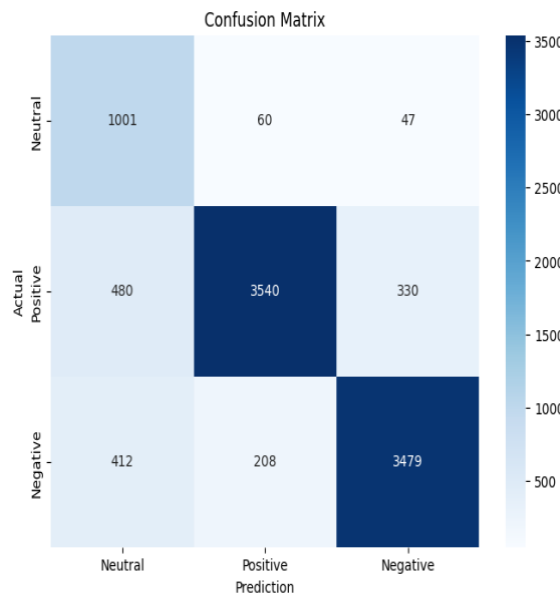
3.7 Build Model

Model yang digunakan dalam analisis sentimen yaitu *naïve bayes* dan *support vector machine* menggunakan data yang sudah dibagi antara data *training* dan data testing akan divisualisasikan dalam *confusion matrix*. Penjelasan lebih lengkapnya di bawah ini.

3.7.1 Naïve Bayes

Algoritma *naïve bayes* menerapkan metode teorema *naïve bayes* dalam memprediksi probabilitas suatu kelas [9]. Pembangunan model *naïve bayes* dengan mengimpor modul *multinomial naïve bayes* dari

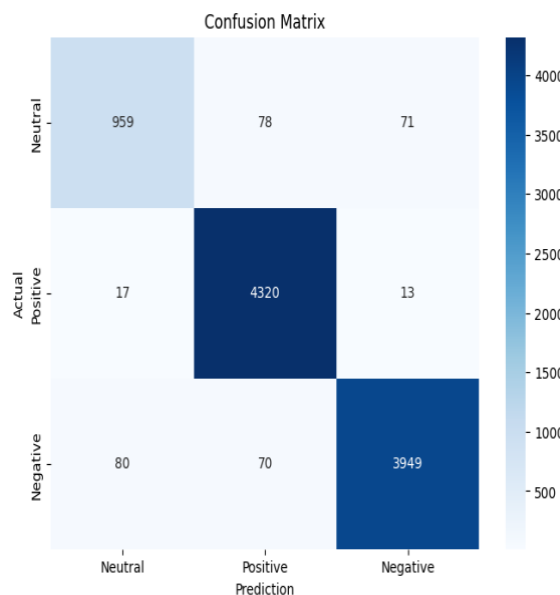
scikit learn. Model ini memperoleh hasil *accuracy* 83.92%, *precision* 87.14%, *recall* 83.92%, dan *f1-score* 84.75%. Tampilan *confusion matrix naïve bayes* ditampilkan pada Gambar 3.



Gambar 3. Confusion Matrix Naïve Bayes

3.7.2 Support Vector Machine

Algoritma *support vector machine* adalah representasi dari data sebagai titik dalam ruang vektor yang dipetakan untuk membagi data dalam ruang yang selebar mungkin dan akan dipetakan dalam ruang yang sama untuk klasifikasi berdasarkan titik data dipetakan [15]. Pembangunan model *support vector machine* dengan mengimpor modul *Linear SVC* dan *SVC* dari *scikit learn*. Model ini memperoleh hasil *accuracy* 95.92%, *precision* 95.89%, *recall* 95.92%, dan *f1-score* 95.89%. Tampilan *confusion matrix support vector machine* ditampilkan pada Gambar 4.



Gambar 4. Confusion Support Vector Machine

3.8 Experiment Model

Experiment model adalah proses pengujian model yang telah dilatih. Data set yang digunakan adalah data *tweet* yang dikumpulkan dari postingan *bitcoin* akun media sosial X milik @Bitcoin. Dalam hasil pengujian model yang dilakukan menggunakan lima dataset, model *naïve bayes* mampu menganalisis empat dataset dengan benar sehingga memperoleh persentase 80% dan model *support vector machine* mampu menganalisis tiga dataset dengan benar sehingga memperoleh persentase 60%, Berikut adalah rangkaian proses pengujian model yang dimuat dalam Tabel 6.

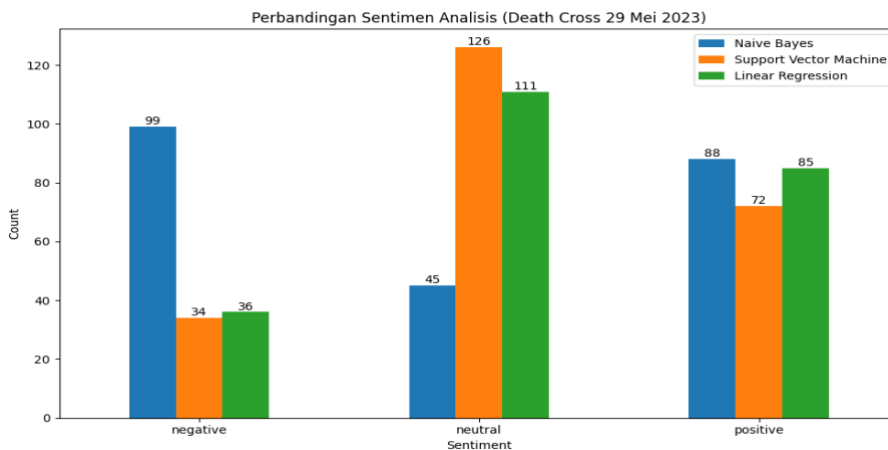
Tabel 6. Hasil Data Pengujian Model

Algoritma	Tanggal	Status	Total	Positif	Netral	Negatif	Hasil
Naïve Bayes	27 Mei 2023 – 1 Juni 2023	29 Mei 2023 (Death Cross)	232 data	84 data (36.21%)	48 data (20.69%)	100 data (43.10%)	Sesuai
	16 Juni 2023 – 22 Juni 2023	19 Juni 2023 (Golden Cross)	741 data	306 data (41.30%)	169 data (22.81%)	266 data (35.90%)	Sesuai
	28 Juli 2023 – 3 Agustus 2023	31 Juli 2023 (Death Cross)	390 data	152 data (38.97%)	87 data (22.31%)	151 data (38.72%)	Tidak Sesuai
	13 Oktober 2023 – 19 Oktober 2023	16 Oktober 2023 (Golden Cross)	1201 data	538 data (44.80%)	243 data (20.23%)	420 data (38.72%)	Sesuai
	26 Januari 2024 – 1 Februari 2024	29 Januari 2024 (Golden Cross)	2430 data	1237 data (50.91%)	498 data (20.49%)	695 data (28.60%)	Sesuai
Support Vector Machine	27 Mei 2023 – 1 Juni 2023	29 Mei 2023 (Death Cross)	232 data	72 data (31.17%)	125 data (54.11%)	34 data (14.72%)	Tidak Sesuai
	16 Juni 2023 – 22 Juni 2023	19 Juni 2023 (Golden Cross)	741 data	258 data (34.82%)	416 data (56.14%)	67 data (9.04%)	Sesuai
	28 Juli 2023 – 3 Agustus 2023	31 Juli 2023 (Death Cross)	390 data	134 data (34.36%)	207 data (53.08%)	49 data (12.56%)	Tidak Sesuai
	13 Oktober 2023 – 19 Oktober 2023	16 Oktober 2023 (Golden Cross)	1201 data	451 data (37.55%)	608 data (50.62%)	142 data (11.82%)	Sesuai
	26 Januari 2024 – 1 Februari 2024	29 Januari 2024 (Golden Cross)	2430 data	1015 data (41.77%)	1205 data (49.59%)	210 data (8.64%)	Sesuai

Pada data pertama, karena terjadi *death cross* pada tanggal 29 Mei 2023 yang ditampilkan di Gambar 5, data *tweet* mulai dikumpulkan dari tanggal 27 Mei 2023 sampai 1 Juni 2023. Data yang terkumpul sebanyak 232 data. Dalam pengujian model *naïve bayes* ditemukan 84 data positif, 48 data netral, dan 100 data negatif. Pada pengujian model *support vector machine* ditemukan 72 data positif, 125 data netral, dan 34 data negatif. Hasil sentimen analisis dari model dimuat pada Gambar 6.



Gambar 5. Death Cross 29 Mei 2023

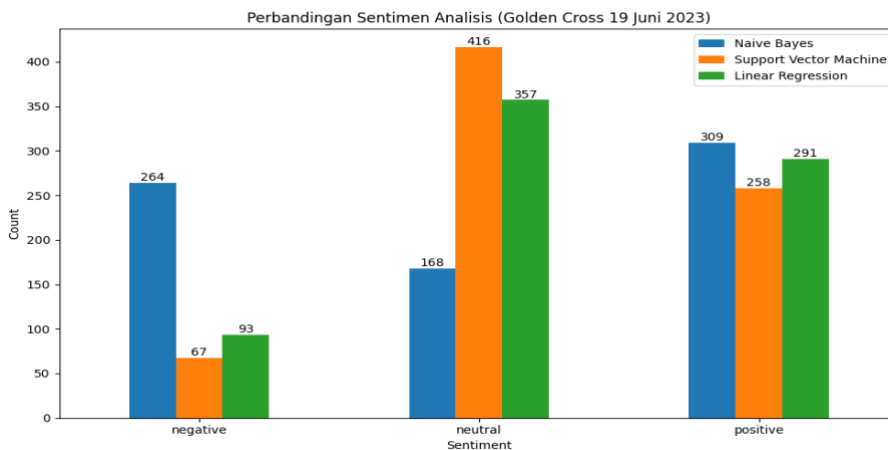


Gambar 6. Sentimen Analisis *Death Cross* 29 Mei 2023

Pada data kedua, terjadi *golden cross* pada tanggal 19 Juni 2023 yang ditampilkan di Gambar 7 sehingga data tweet mulai dikumpulkan dari tanggal 16 Juni 2023 sampai 22 Juni 2023. Data yang terkumpul sebanyak 741 data. Dalam pengujian model *naïve bayes* ditemukan 306 data positif, 169 data netral, dan 266 data negatif. Pada pengujian model *support vector machine* ditemukan 258 data positif, 416 data netral, dan 67 data negatif. Hasil sentimen analisis dari model dimuat pada Gambar 8.



Gambar 7. *Golden Cross* 19 Juni 2023

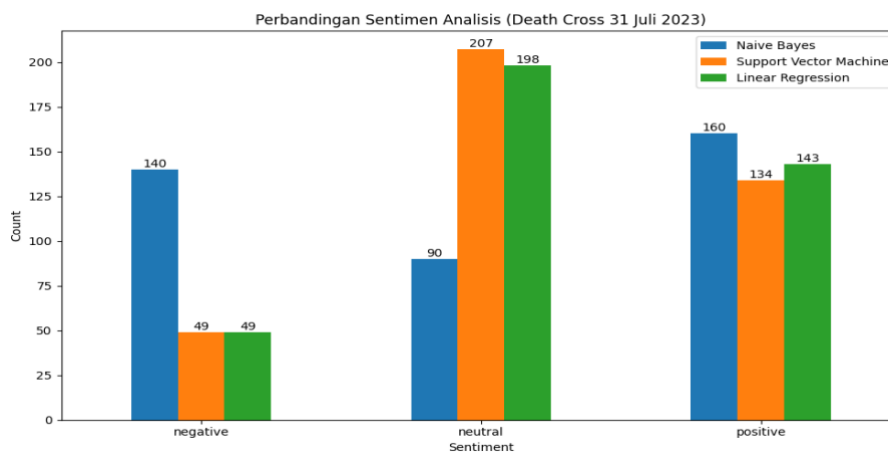


Gambar 8. Sentimen Analisis *Golden Cross* 19 Juni 2023

Pada data ketiga, terjadi *death cross* pada tanggal 31 Juli 2023 yang ditampilkan di Gambar 9 sehingga data tweet mulai dikumpulkan dari tanggal 28 Juli 2023 sampai 3 Agustus 2023. Data yang terkumpul sebanyak 390 data. Dalam pengujian model *naïve bayes* ditemukan 152 data positif, 87 data netral, dan 151 data negatif. Pada pengujian model *support vector machine* ditemukan 134 data positif, 207 data netral, dan 49 data negatif. Hasil sentimen analisis dari model dimuat pada Gambar 10.



Gambar 9. Death Cross 31 Juli 2023

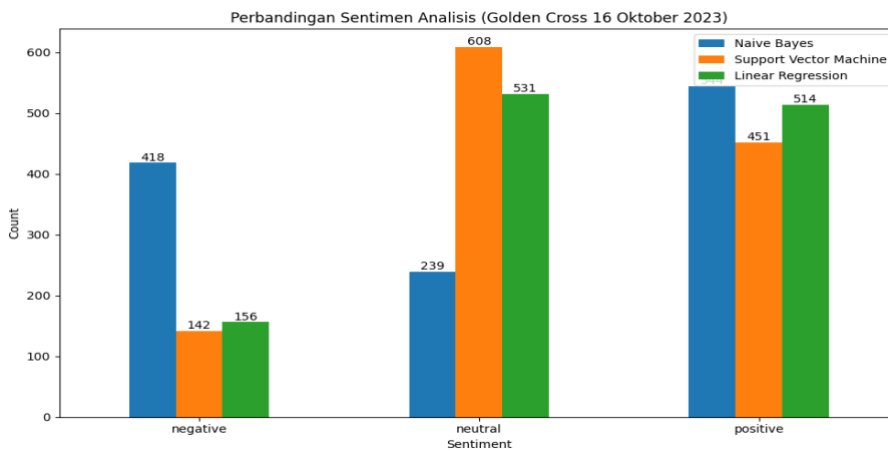


Gambar 10. Sentimen Analisis *Death Cross* 31 Juli 2023

Pada data keempat, terjadi *golden cross* pada tanggal 16 Oktober 2023 yang ditampilkan di Gambar 11 sehingga data tweet mulai dikumpulkan dari tanggal 13 Oktober 2023 sampai 19 Oktober 2023. Data yang terkumpul sebanyak 1201 data. Dalam pengujian model *naïve bayes* ditemukan 538 data positif, 243 data netral, dan 420 data negatif. Pada pengujian model *support vector machine* ditemukan 451 data positif, 608 data netral, dan 142 data negatif. Hasil sentimen analisis dari model dimuat pada Gambar 12.



Gambar 11. Golden Cross 16 Oktober 2023

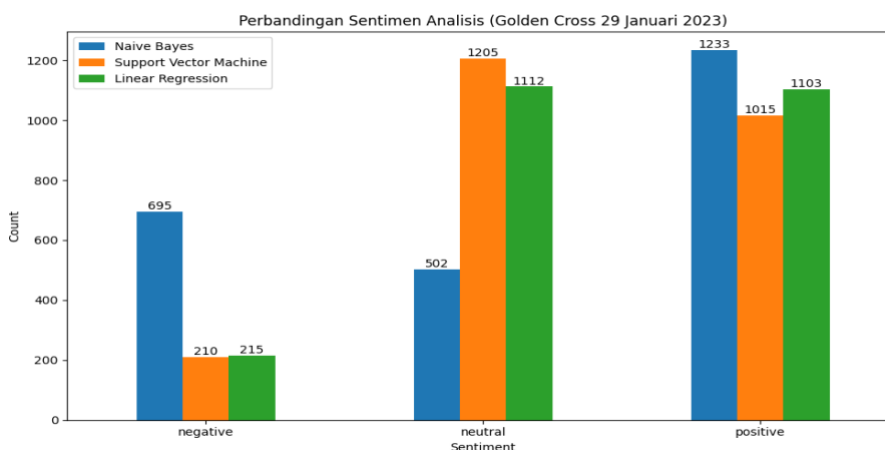


Gambar 12. Sentimen Analisis Golden Cross 16 Oktober 2023

Pada data kelima, terjadi *golden cross* pada tanggal 29 Januari 2024 yang ditampilkan di Gambar 13 sehingga data tweet mulai dikumpulkan dari tanggal 26 Januari 2023 sampai 1 Februari 2024. Data yang terkumpul sebanyak 2430 data. Dalam pengujian model *naïve bayes* ditemukan 1237 data positif, 498 data netral, dan 695 data negatif. Pada pengujian model *support vector machine* ditemukan 1015 data positif, 1205 data netral, dan 210 data negatif. Hasil sentimen analisis dari model dimuat pada Gambar 14.



Gambar 13. Golden Cross 29 Januari 2024



Gambar 14. Sentimen Analisis Golden Cross 29 Januari 2024

3.9 Analisis Hasil

Dalam analisis hasil ini peneliti melakukan perbandingan harga *bitcoin* pada status *death cross* dan *golden cross* terhadap sentimen yang dihasilkan dari tiap data uji yang terkumpul dari *tweet* postingan akun *bitcoin*. Peneliti menemukan sentimen netral memiliki persentase terbanyak, sehingga mendominasi sentimen positif maupun negatif. Hal ini menunjukkan bahwa mayoritas pengguna media sosial X cenderung mencari informasi tentang *bitcoin* tanpa mengekspresikan perasaan positif atau negatif yang kuat. Selanjutnya, ditemukan ketika persentase sentimen positif lebih besar daripada sentimen negatif pada saat status *golden cross*, maka *bitcoin* akan mengalami kenaikan harga dikarenakan optimisme para pengguna terhadap pasar. Selanjutnya ketika terjadinya *death cross*, ditemukan sentimen positif tetap mendominasi daripada sentimen negatif sehingga menunjukkan para pengguna tetap optimis atau berharap harga *bitcoin* akan naik walaupun pada kenyataannya *bitcoin* mengalami penurunan harga sesuai dengan status yang terjadi yakni *death cross*. Hal ini dapat diartikan bahwa komunitas pengguna *bitcoin* di media sosial X memiliki pandangan jangka panjang terhadap harga *bitcoin* ke depannya. Hal ini diperoleh dari hasil proses *experiment model* yang telah dilakukan menggunakan tiga model antara lain *naïve bayes*, *support vector machine*, dan *logistic regression* berdasarkan data pengujian yang dikumpulkan dari *tweet* postingan akun *bitcoin*.

4. Kesimpulan

Penelitian ini menganalisis data *tweet* dari akun *@bitcoin* yang diambil berdasarkan tanggal terjadinya *golden cross* atau *death cross* pada 1 Mei 2023 – 31 Maret 2024. Dalam evaluasi sentimen pada model *naïve bayes* dan *support vector machine* diperoleh jika model *support vector machine* menjadi algoritma terbaik terhadap keakuratan model dalam klasifikasi sentimen dengan *accuracy* sebesar 95.92%, ketepatan model dalam memprediksi nilai positif dengan tingkat akurat 95.89%, tingkat usaha dalam menemukan informasi kembali sebesar 95.92%, dan presentasi nilai bobot dari nilai *precision* dengan nilai *recall* sebesar 95.89%. Akan tetapi, dalam memprediksi sentimen lima dataset pengujian yang diberikan menggunakan model yang telah dilatih ditemukan algoritma *naïve bayes* memiliki persentase lebih tinggi yaitu 80% dibandingkan algoritma *support vector machine* yaitu 60% dalam memperoleh hasil yang sesuai antara sentimen positif untuk kondisi *golden cross* dan sentimen negatif untuk kondisi *death cross*.

Dalam pengujian model yang dilakukan dalam menganalisis terhadap dataset *tweet* berdasarkan terjadinya *golden cross* ataupun *death cross* diperoleh hasil sentimen netral lebih mendominasi daripada sentimen positif ataupun negatif. Sehingga dapat disimpulkan jika pengguna media sosial X hanya ingin mengetahui informasi tentang *bitcoin*. Pada saat terjadinya *golden cross* memiliki hasil yang sesuai yaitu sentimen positif pengguna Twitter memiliki persentase lebih tinggi dibandingkan sentimen negatif. Akan tetapi, ketika terjadi *death cross*, sentimen yang dihasilkan tetap lebih dominan sentimen positif. Sehingga dapat disimpulkan jika para pengguna *tweet* memiliki keinginan supaya harga *bitcoin* tetap naik walaupun dalam status *death cross*.

Referensi

- [1] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, 2020, doi: 10.1007/s10462-019-09794-5.
- [2] S. Aggarwal, R. Chaudhary, G. S. Aujla, N. Kumar, K. K. R. Choo, and A. Y. Zomaya, "Blockchain for smart communities: Applications, challenges and opportunities," *J. Netw. Comput. Appl.*, vol. 144, no. February, pp. 13–48, 2019, doi: 10.1016/j.jnca.2019.06.018.
- [3] K. H. Al-Yahyaee, M. U. Rehman, W. Mensi, and I. M. W. Al-Jarrah, "Can uncertainty indices predict Bitcoin prices? A revisited analysis using partial and multivariate wavelet approaches," *North Am. J. Econ. Financ.*, vol. 49, no. March, pp. 47–56, 2019, doi: 10.1016/j.najef.2019.03.019.
- [4] S. Rouhani and E. Abedin, "Crypto-currencies narrated on tweets: a sentiment analysis approach," *Int. J. Ethics Syst.*, vol. 36, no. 1, pp. 58–72, 2020, doi: 10.1108/IJOES-12-2018-0185.
- [5] K. Joshi, B. H. N, and J. Rao, "Stock Trend Prediction Using News Sentiment Analysis," *Int. J. Comput. Sci. Inf. Technol.*, vol. 8, no. 3, pp. 67–76, 2016, doi: 10.5121/ijcsit.2016.8306.
- [6] I. Hamraoui and A. Boubaker, "Impact of Twitter sentiment on stock price returns," *Soc. Netw. Anal. Min.*, vol. 12, no. 1, pp. 60–71, 2022, doi: 10.1007/s13278-021-00856-7.
- [7] E. W. Kim, N. E. Syaheerah, A. Shukor, N. R. Ismail, and S. A. Halim, "Golden cross as Buying Indicator for Stock Investment in Bursa Malaysia," *Teren. Int. Financ. Econ. J.*, vol. 2, no. 2, pp. 94–102, 2012, [Online]. Available: <http://thestar.com.my>
- [8] S. C. Pranoto, "Web Scraping On Forecast Report Form Of Geomagnet," *[CEPAT] J. Comput. Eng. Progress, Appl. Technol.*, vol. 1, no. 02, p. 16, 2022, doi: 10.25124/cepat.v1i02.5226.
- [9] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 205–215, 2022, doi: 10.23887/janapati.v11i3.49450.
- [10] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
- [11] M. Ghiassi and S. Lee, "A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach," *Expert Syst. Appl.*, vol. 106, pp. 197–216, 2018, doi: 10.1016/j.eswa.2018.04.006.
- [12] S. Rizkia, E. Budi Setiawan, and D. Puspendari, "Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF," *e-Proceeding Eng.*, vol. 6, no. Agustus, pp. 9683–9693, 2019.
- [13] R. Maulana, P. A. Rahayuningsih, W. Irmayani, D. Saputra, and W. E. Jayanti, "Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain," in *Journal of Physics: Conference Series*, Nov. 2020, vol. 1641, no. 1. doi: 10.1088/1742-6596/1641/1/012060.
- [14] A. Kulkarni, D. Chong, and F. A. Batarseh, *Foundations of data imbalance and solutions for a data democracy*. Elsevier Inc., 2020. doi: 10.1016/B978-0-12-818366-3.00005-8.
- [15] Burhanudin, Y. Musa, and Y. Wihardi, "Klasifikasi Komentar Spam Pada Youtube Menggunakan Metode Naive Bayes, Support Vector Machine, Dan K-Nearest Neighbors," *J. Inform. dan Komput.*, vol. 3, no. 2, pp. 54–59, 2108.