

Model Klasifikasi Multinomial Naïve Bayes Untuk Analisis Sentiment Terkait Non-Fungible Token

Rayhan Yuda Lesmana¹, Roni Andarsyah²

Program Studi D4 Teknik Informatika, Politeknik Pos Indonesia
Jalan Sariasih No. 54, Sarijadi, Sukasari, Kota Bandung, Jawa Barat 40151¹²

email: ¹rayhanyuda13@gmail.com, ²roniandarsyah@poltekpos.ac.id

Abstrak

Belakangan ini perkembangan ilmu pengetahuan dan teknologi semakin maju. Seiring berjalannya waktu telah dikenal teknologi yang disebut Blockchain. NFT (*Non-Fungible Token*) adalah aset digital yang sedang populer di dunia crypto merupakan salah satu teknologi berbasis blockchain. Laporan ini menyajikan hasil penelitian yang dilakukan terkait “Model Klasifikasi Multinomial Naïve Bayes untuk Analisis Sentimen terkait NFT” dengan menggunakan 7060 data twitter berdasarkan tagar #NFT. Kemudian dilakukan preprocessing, proses labeling menggunakan Vader menunjukkan bahwa sentiment pengguna twitter terkait NFT cenderung lebih banyak ditemukan sentiment negatif sebanyak 3840 tweet sedangkan sentiment positif sebanyak 3220 tweet yang divisualisasikan menggunakan wordcloud. Setelah diklasifikasikan menggunakan metode MNB data mendapatkan akurasi sebesar 84% setelah di evaluasi performansi model yang didapatkan yaitu precision 84% recall 81% dan f1-score 83%. Untuk menganalisis kinerja model di ukur dengan menggunakan confusion matrix mendapatkan hasil TruePositive (830), TrueNegative (661), FalsePositive (122), FalseNegative (152). Sehingga model Multinomial Naïve Bayes mampu bekerja dengan baik dalam menganalisis sentiment.

Kata Kunci: *Non-Fungible token, Multinomial Naïve bayes, Tweet.*

Abstract

Recently, the development of science and technology has been increasingly advanced. Over time, a technology called Blockchain has become known. NFT (Non-Fungible Token) is a digital asset currently popular in the crypto world. It is a blockchain-based technology. This report presents the results of research conducted on the “Naive Bayes Multinomial Classification Model for Sentiment Analysis related to NFT” using 7060 Twitter data based on the hashtag #NFT. Then preprocessing was carried out, and the labeling process using Vader showed that the sentiment of Twitter users related to NFT tended to be found in 3840 negative sentiments, while 3220 positive sentiments were visualized using a word cloud. After being classified using the MNB method, the data obtained an accuracy of 84%. After evaluating the performance of the model, the results obtained were 84% precision, 81% recall, and 83% f1-score. To analyze the performance of the model, it is measured using a confusion matrix to get TruePositive (830), TrueNegative (661), FalsePositive (122), and FalseNegative (152) results so that the Multinomial Naïve Bayes model can work well in analyzing sentiment.

Keywords: *Non-fungible token, Multinomial Naïve Bayes, Tweet.*

1. PENDAHULUAN

Non-Fungible Token adalah salah satu produk blockchain jenis dari cryptocurrency [1], nft dapat mengesahkan aset digital yang sifatnya unik [2] yang diturunkan oleh smart contracts of Ethereum [3]. Nft bersifat unik dan tidak dapat dipertukarkan like-for-like, sehingga cocok untuk mengidentifikasi sesuatu

atau seseorang dengan cara yang unik, dengan menggunakan nft seorang pembuat konten dapat dengan mudah untuk membuktikan kepemilikan aset digital dalam bentuk video, gambar atau seni [4] dll. Kedalam smart contract dan disimpan di blockchain [5]. Teknologi blockchain memungkinkan asosiasi bukti kepemilikan ke aset digital hanya

merupakan bukti kepemilikan, bukan hak cipta; yaitu, pemilik tidak memiliki akses eksklusif ke konten nft. Nft telah tumbuh secara eksponensial pada tahun 2021. Harga nft yang fluktuatif dan popularitasnya yang tiba-tiba telah membuat banyak orang membuat nft dan pada gilirannya mempromosikan dan menjualnya untuk mendapatkan keuntungan. Dampak paling signifikan dari nft tentang bagaimana mereka mengubah dunia seni [2].

Pada media social twitter pembahasan mengenai nft sedang sangat tinggi banyak orang – orang ingin mengetahui tentang nft. Twitter yang merupakan salah satu media social dengan pengguna aktif terbanyak sampai dengan 6,43 juta orang pada tahun 2019 [6]. Dengan media social twitter dapat mendapatkan berbagai macam informasi, karena dengan banyaknya tweet yang dipublikasikan melalui twitter, tweet ini mungkin berisi balasan dari pengguna tentang objek, dan objek tersebut dapat berupa cara terkait komunitas, produk, dan layanan lain nya. dengan seiring waktu dan banyak nya data tweet, data tersebut akan terus bertambah dan jika itu dibiarkan akan menjadi sampah data [7].

Karena nft merupakan fenomena baru-baru ini sehingga penelitian pada data terkait nft masih terbatas. Sebagian besar penelitian sebelumnya mengenai topik nft adalah pada analisis hubungan antara blockchain, mata uang kripto dan nft. Penelitian yang terbaru membahas bagaimana memahami korelasi antara asar nft dengan enthereum dan bitcoin [8]. Analisis cryptocurrency dan dasar-dasar teknis blockchain, khususnya token non-fungible dan "cryptocollectibles", dan perubahan yang dapat ditimbulkan oleh inovasi ini di pasar seni dan industri kreatif [9]. Berdasarkan penelitian sebelumnya karena dengan banyak keingin tahunan orang – orang mengenai nft dan penelitian pada data terkait nft masih terbatas, sehingga penelitian ini bertujuan untuk mengeksplorasi bagaimana sentiment pangguna twitter sebagai sarana untuk membangun persepsi publik terhadap nft dapat diukur melalui analisis sentiment dan mengkasifikasikan data tweet melalui machine learning.

Pada penelitian ini, algoritma yang digunakan untuk klasifikasi adalah Multinomial Naive Bayes karena lebih sederhana dan konvensional dibandingkan algoritma lainnya yang lebih kompleks,

sehingga waktu komputasi yang dibutuhkan dalam proses klasifikasi tentunya akan lebih singkat dengan naive bayes [10]. Pada penelitian sebelumnya yang menerapkan metode NBC, bahwa algoritma Naive Bayes Classifier dapat mengklasifikasikan suatu opini dalam bentuk tweet menjadi dua kelas yaitu positive dan negative dengan akurat [11]. Pada penelitian sebelumnya untuk menganalisis sentiment positive dan negative pada ulasan film yang berjumlah banyak sehingga penanganan dan pengorganisasian dokumen secara manual memakan waktu dan bukan solusi yang layak mengingat jumlah dokumen sangat sangat besar. Sehingga diusulkan penggunaan metode MNB untuk pengklasifikasian, Pengklasifikasi MNB adalah jenis pengklasifikasi NB dan sering digunakan sebagai dasar untuk teks klasifikasi tetapi di sini diterapkan untuk Analisis Sentimen (SA). Multinomial dianggap sebagai pendekatan pemodelan dominan dan lebih efisien daripada model Bernoulli multivariat yang memperkenalkan bahasa pemodelan dalam pencarian informasi. Klasifikasi tekstual NB paling sering digunakan untuk mengkategorikan teks karena cepat dan mudah diterapkan. Untuk memilih pengklasifikasi Naive Bayes (NBC) akan lebih diinginkan karena kecepatannya yang tinggi. metode klasifikasi Naive Bayes merupakan salah satu algoritma yang banyak digunakan pada proses analisis sentimen. Dengan begitu, Pengklasifikasi Multinomial Naive Bayes algoritma lebih cocok untuk diterapkan dalam penelitian ini. Selain tingkat akurasi yang tinggi dan pemrosesan data yang cepat dari metode lainnya.

2. LANDASAN TEORI

2.1 Non-fungible token

Non-Fungible Token adalah salah satu produk blockchain jenis dari cryptocurrency [1], nft dapat mengesahkan asset digital yang sifatnya unik [2] yang diturunkan oleh smart contracts of Ethereum [3]. Nft bersifat unik dan tidak dapat dipertukarkan like-for-like, sehingga cocok untuk mengidentifikasi sesuatu atau seseorang dengan cara yang unik, dengan menggunakan nft seorang pembuat konten dapat dengan mudah untuk membuktikan kepemilikan asset digital dalam bentuk vidio, gambar atau seni [4] dll. Kedalam smart contract dan disimpan di blockhain [5].

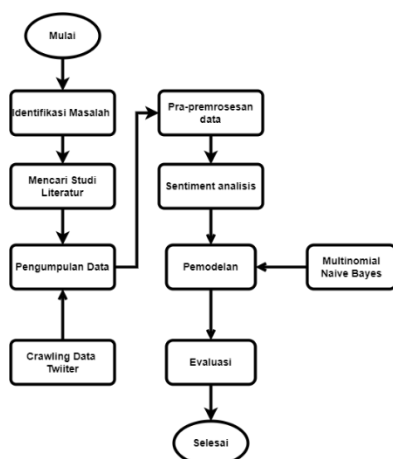
2.2 Sentiment analisis

Analisis Sentiment dapat didefinisikan sebagai ilmu yang menganalisis pendapat, emosi, sikap, dan perasaan seseorang dalam sebuah text tertulis. Manfaat utama dari analisis sentimen adalah untuk secara efektif mengidentifikasi dan mengklasifikasikan sikap dan sentimen (positive, negative) pengguna dalam teks untuk memastikan sikap user terhadap produk, subjek, atau layanan.

2.3 Multinomial Naïve bayes

Naïve bayes merupakan algoritma pembelajaran sederhana yang menggunakan aturan bayes untuk pengklasifikasian atau memprediksi probabilitas suatu class data. Kelas dengan probabilitas tertinggi dianggap sebagai kelas yang paling mungkin. Pengklasifikasi multinomial Naive Bayes cocok untuk klasifikasi dengan fitur diskrit (misalnya, jumlah kata untuk klasifikasi teks). Distribusi multinomial biasanya membutuhkan jumlah fitur integer. Frekuensi istilah sering dinormalisasi dengan membagi frekuensi istilah mentah dengan panjang dokumen. Setelah normalisasi, frekuensi istilah dapat digunakan untuk menghitung perkiraan kemungkinan maksimum berdasarkan pelatihan data untuk memperkirakan probabilitas bersyarat.

3. METODE PENELITIAN



Gambar 1. Alur Metodologi Penelitian

4. HASIL DAN PEMBAHASAN

A. Pengumpulan data

Untuk mengumpulkan dataset histori tweet terkait nft yang digunakan pada penelitian ini menggunakan Teknik *crawling* dengan Bahasa

pemrograman python menggunakan tools jupyter notebook. Tweet yang diambil dengan kata kunci “NFT”, dengan atribut `create_at`, `user_name`, `tweet`. Tanggal pengambilan data dimulai pada tanggal 2022-06-19 sampai dengan 2022-06-27, dengan menghasilkan 7060 data.

	date	user_name	text
0	2022-06-19 23:59:59+00:00	boringoldguy	10ix "Sitting Down Under" by @daddyson_nft sho...
1	2022-06-19 23:59:59+00:00	wlch_town_wff	A bit more than 4k #Witches left. in/nThe spe...
2	2022-06-19 23:59:58+00:00	ka_balagtas3316	@mindzeye @CozomoMedici Check WEYU's Ecosystem...
3	2022-06-19 23:59:57+00:00	slo_m1995	My new monkey collection 'monkey 10/n/n #nft...
4	2022-06-19 23:59:56+00:00	DommmiqueDesdu1	@2Cold_XRPL @seokings_nft @ClassyXoge @TREASUR...
...
7055	2022-06-27 23:34:58+00:00	lauraverse_00	@shapes_nft send me dm for promo I...
7056	2022-06-27 23:34:58+00:00	oussama_menouer	@AItty_@AItty_@DuMbGums_NFT Dumgums ARE...
7057	2022-06-27 23:34:58+00:00	NadMthri	@HakunaBanana6 @p1u_nft @durga1996 @gorio_yk @...
7058	2022-06-27 23:34:57+00:00	ArbysMMMMMMMM	#NewNFTProfilePic NFT by @DegenToonz #MutantMo...
7059	2022-06-27 23:34:57+00:00	guangyutatsngs1	Perhaps, some emotions, sincere and warm abloc...

Gambar 2. Data Hasil Crawling

B. Preprocessing data

Pembersihan data tweets yang telah didapat dilakukan dalam beberapa tahapan, ini bertujuan untuk memastikan kualitas data yang akan digunakan saat proses analisis data menjadi lebih baik, proses ini menyeragamkan format data kedalam satu set data. Berikut tahapan – tahapan preprocessing data yang di gunakan pada penelitian ini:

1) Casefolding

Case folding merupakan proses dalam text preprocessing yang dilakukan untuk menyeragamkan karakter pada data. Proses case folding adalah proses mengubah seluruh huruf menjadi huruf kecil. Pada proses ini karakter-karakter `A`-'Z' yang terdapat pada data diubah kedalam karakter 'a'-'z'. Untuk Casefolding menggunakan library `re` yang berguna untuk mencari string atau teks dengan menggunakan pola.

2) StopWord

Stopwords adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopwords dalam bahasa Inggris adalah “i”, “me”, “my”, “our”. Untuk Stopwords menggunakan library `nlTK` dengan mengimport library `stopwords` di `nlTK`, yang di dalam nya sudah terdapat kata – kata kurang penting.

3) Lemmatization

Lemmatization adalah perubahan kata dengan makna yang sama menjadi satu bentuk karena komputer akan membacanya sebagai data yang berbeda sehingga proses analisis akan menjadi lebih sulit. Contohnya, “eat”, “ate”, “eaten”, “eating” merupakan bentuk dari “eat”, sehingga akan diubah menjadi “eat”. Untuk lemmatization menggunakan libray `nlTK` dengan

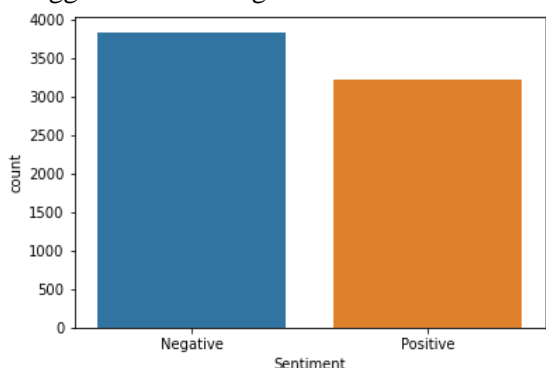
mengimport library wordnet dan wordnetlemmatizer.

4) SpellingCorrections

mendeteksi bagian dari teks yang tidak sesuai dengan kaidah ejaan yang benar dan menghasilkan rekomendasi sesuai kaidah ejaan yang benar. Untuk spelling corrections menggunakan library spellchecker.

C. Sentiment analisis

Setelah dataset di bersihkan selanjutnya melakukan sentiment analisis terhadap teks tweet agar dapat mengetahui sentiment negative atau positive yang terkandung didalam tweet tersebut, dengan menggunakan library vader data tweet akan di berikan polarity score, jika compound score yang didapat lebih besar sama dengan 0.05 akan mendapatkan label positive selain dari itu akan mendapatkan label negative. Setelah didapatkan semua sentiment pada data tweet kemudia di visualisasikan data yang ada dengan menggunakan bar diagram.



Gambar3. Hasil sentiment analisis

Dari hasil sentiment analisis didapatkan hasil sentiment negative (3841) lebih besar dari pad sentiment positive (3219).

A. Pemodelan

Sebelum melakukan penerapan model, data yang sudah di dapatkan sentiment nya kemudian merubah label yang sebelumnya string menjadi numerik dengan ketentuan positive=1 dan negative=0.

date	user_name	text	class_text	stopword	lemmatized	text	Negative	Positive	Compound	Sentiment	Polarity
2022-06-16 22:25:54+03:00	komperku	50% barang yang masuk ke dalam negeri...	50% barang yang masuk ke dalam negeri...	50% barang yang masuk ke dalam negeri...	50% barang yang masuk ke dalam negeri...	50% barang yang masuk ke dalam negeri...	0.000	0.000	0.000	Negative	0
2022-06-16 23:59:58+03:00	whol_mak_ah	0.000	0.000	-0.000	Negative	0
2022-06-16 23:59:58+03:00	ku_habibac315	0.000	0.000	0.000	Negative	0
2022-06-16 23:59:57+03:00	ku_m380	0.000	0.000	0.000	Negative	0
2022-06-16 23:59:54+03:00	komnoco0001	0.000	0.200	0.200	Positive	1

Gambar4. Hasil polarity

Selanjutnya data akan di split menjadi dua data training dan testing, pada penelitian ini data

yang akan di split menggunakan komposisi data training sebanyak 75% (5295) dari dataset yang ada dan data testing sebanyak 25% (1765). Kemudian setelah data di split akan dilakukan vektorisasi data untuk mengubah atau mengkonversikan teks ke dalam bentuk vector atau jumlah token. Proses ini diperlukan teks yang akan di encode sebagai bilangan bulat untuk pemodelan teks prediktif yang nanti nya akan di gunakan sebagai pembelajaran algoritma machine learning. Kemudian dilakukan penerapan model dengan menggunakan data training dengan algoritma Multinomial Naïve Bayes.

D. Evaluasi

Setelah dilakukan pengujian model, selanjutnya mengevaluasi performa model dapat ditampilkan dengan classification report Didapatkan nilai evaluasi sebagai berikut:

- 1) Accuracy (rasio prediksi benar): 0.84.
- 2) Precision (rasio prediksi benar positive dibandingkan dengan keseluruhan hasil yang diprediksi positive), “Berapa persen tweet yang benar positive dari keseluruhan tweet yang diprediksi positive?”: 0.84.
- 3) Recall (rasio prediksi benar positive dibandingkan dengan keseluruhan data yang benar positive), “Berapa persen tweet yang diprediksi positive dibandingkan keseluruhan tweet yang sebenarnya negative?”: 0.81.
- 4) F1-score (perbandingan rata-rata presisi dan recall yang dibobotkan): 0.83.

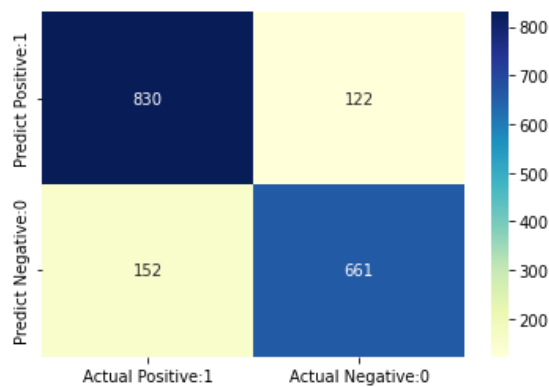
	precision	recall	f1-score	support
0	0.85	0.87	0.86	952
1	0.84	0.81	0.83	813
accuracy			0.84	1765
macro avg	0.84	0.84	0.84	1765
weighted avg	0.84	0.84	0.84	1765

Gambar5. Hasil classification_report

Selanjutnya mengukur kinerja algoritma multinomial naïve bayes dengan confusion matrix dengan menggunakan library sklearn. Dengan confusion matrix dapat menampilkan informasi yang membandingkan hasil klasifikasi yang dilakukan oleh system dan hasil klasifikasi yang seharusnya. Dalam confusion matrix terdapat 4 istilah sebagai representasi hasil proses klasifikasi.

- 1.True Positive (Jumlah data positive yang terklasifikasi benar oleh sistem)
- 2.True Negative (Jumlah data negative yang terklasifikasi benar oleh sistem)

- 3.False Positive (Jumlah data negative tetapi terklasifikasi salah oleh sistem)
- 4.False Negative (Jumlah data positive tetapi terklasifikasi salah oleh sistem)



Gambar6. Hasil confusion matrix

Setelah dilakukan perhitungan dengan menggunakan confusion matrix didapatkan hasil. TruePositive (830), TrueNegative (661), FalsePositive (122), FalseNegative (152).

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan terkait Model Klasifikasi Multinomial Naïve Bayes untuk Analisis Sentimen terkait NFT dengan menggunakan 7060 data tweet pengguna twitter berdasarkan tagar #NFT kemudian di lakukan preprocessing diantaranya case folding, stopword removal, lemmatization, spelling correction dan proses labeling menggunakan Vader menunjukkan bahwa sentiment pengguna twitter terkait NFT cenderung lebih banyak ditemukan sentiment negatif sebanyak 3840 tweet sedangkan sentiment positif sebanyak 3220 tweet. Setelah diklasifikasikan menggunakan metode multinomial naïve bayes data mendapatkan akurasi sebesar 84% setelah di evaluasi performansi model yang didapatkan yaitu precision 84% recall 81% dan f1-score 83%. Untuk kinerja model multinomial naïve bayes yang di ukur dengan menggunakan confusion matrix mendapatkan hasil TruePositive (830), TrueNegative (661), FalsePositive (122), FalseNegative (152). Sehingga model multinomial naïve bayes mampu bekerja dengan baik dalam menganalisis sentiment.

Untuk meningkatkan akurasi pada model pada penelitian selanjutnya klasifikasi sentiment akan ditambah pengklasifikasian

sentiment netral dan ditambahkan model TF-IDF untuk ekstraksi fitur pada kata dan dilakukan perbandingan metode Naïve bayes dengan SVM untuk membandingkan performansi klasifikasi yang paling baik dari model tersebut.

6. DAFTAR PUSTAKA

- [1] joshua fairfield, "tokenized: the law of non-fungible tokens and unique digital property □," 2021. [online]. available: https://www.espn.com/mlb/story/_/id/29723239/mike-trout-
- [2] a. kapoor, d. guhathakurta, m. mathur, r. yadav, m. gupta, and p. kumaraguru, "tweetboost: influence of social media on nft valuation," jan. 2022.
- [3] g. wood, "ethereum: a secure decentralised generalised transaction ledger," 2014.
- [4] m. franceschet et al., "crypto art: a decentralized view," leonardo, vol. 54, no. 4, pp. 402–405, aug. 2021, doi: 10.1162/leon_a_02003.
- [5] j. van slooten, "predictive value of tweet sentiment on the bored ape yacht club’s trading volume and floor price," 2022, doi: 10.13140/rg.2.2.11790.84808.
- [6] p. adam and f. ferdiansyah, "sentiment analisis terhadap cryptocurrency berdasarkan comment dan reply pada platform twitter," 2020.
- [7] r. parlika, s. i. pradika, a. m. hakim, and k. r. n m, "analisis sentimen twitter terhadap bitcoin dan cryptocurrency berbasis python textblob," jurnal ilmiah teknologi informasi dan robotika, vol. 2, no. 2, pp. 33–37, dec. 2020, doi: 10.33005/jifti.v2i2.22.
- [8] l. ante, "the non-fungible token (nft) market and its relationship with bitcoin and ethereum," ssm electronic journal, 2021, doi: 10.2139/ssrn.3861106.
- [9] s. chevet, "blockchain technology and non-fungible tokens: reshaping value chains in creative industries," ssm electronic journal, 2018, doi: 10.2139/ssrn.3212662.
- [10] m. z. naf’an, a. a. bimantara, a. lasasati, e. m. risondang, and n. a. s. nugraha, "sentiment analysis of cyberbullying on instagram user comments," journal of data science and its applications, vol. 2, no. 1, pp. 88–98, apr. 2019, doi: 10.21108/jdsa.2019.2.20.
- [11] p. chinnsamy, v. suresh, k. ramprathap, b. j. a. jebamani, k. srinivas rao, and m. shiva kranthi, "covid-19 vaccine sentiment analysis using public opinions on twitter," materials today: proceedings, apr. 2022, doi: 10.1016/j.matpr.2022.04.809.