

ANALISIS SENTIMEN PENGGUNA TWITTER TENTANG NFT (NON FUNGIBLE TOKEN) DENGAN METODE NAIVE BAYES CLASSIFIER

Muhammad Iqbal Zakasih¹, Widiyanto Tri Handoko²

^{1,2}Program Studi Teknik Informatika, Universitas Stikubank Semarang,

Jln. Tri Lomba Juang No.1, Semarang 50241

¹ zakasihzakasih@gmail.com, ² wthandoko@edu.unisbank.ac.id²

Abstract

NFT became a hot topic of conversation on Twitter after a teenager managed to earn billions of rupiah by selling selfie photos that were used as NFT. This becomes the pros and cons among the public which of course raises personal opinions. Sentiment analysis is currently widely used by the public as a material to see public opinion or opinion about various things. By using sentiment analysis, we can classify whether the data is positive opinion negative opinion. This study discusses sentiment analysis to measure the accuracy of twitter users' opinions on NFT with the Naïve Bayes Classifier method which is useful for measuring the accuracy value in this study. Many methods can be used to classify these opinions, but this study uses the Naïve Bayes Classifier because this method is widely used by other researchers and produces a high accuracy score. The results of this study are the percentage of accuracy value of 92.33%, recall value of 94%, precision value of 92.12%, and f1-score value of 92.45%.

Keywords : sentiment analysis, classification, opinion, NFT, naive bayes classifier, accuracy

Abstrak

NFT menjadi topik perbincangan hangat di twitter setelah seorang remaja berhasil mendapatkan miliaran rupiah dengan menjual foto selfie yang dijadikan NFT. Hal tersebut menjadi pro dan kontra di kalangan masyarakat yang tentunya menimbulkan opini pribadi. Analisis sentimen saat ini banyak digunakan oleh masyarakat sebagai bahan untuk melihat pendapat atau opini masyarakat tentang berbagai macam hal. Dengan menggunakan analisis sentimen, kita dapat mengklasifikasikan apakah data opini positif opini negatif. Penelitian ini membahas tentang analisis sentimen untuk mengukur tingkat akurasi opini pengguna twitter terhadap NFT dengan metode Naïve Bayes Classifier yang berguna untuk mengukur nilai akurasi dalam penelitian ini. Banyak metode yang dapat digunakan untuk mengklasifikasikan pendapat tersebut, namun penelitian ini menggunakan Naïve Bayes Classifier karena metode ini banyak digunakan oleh peneliti lain dan menghasilkan skor akurasi tinggi. Hasil penelitian ini berupa persentase nilai akurasi sebesar 92,33%, nilai recall sebesar 94%, nilai precision sebesar 92,12%, dan nilai f1 – score sebesar 92,45%.

Kata kunci : analisis sentimen, klasifikasi, pendapat, NFT, naive bayes classifier, akurasi

1. PENDAHULUAN

NFT pertama kali diperkenalkan pada tahun 2014 oleh sebuah platform bernama Counterparty, dengan NFT pertama yang dibuat berjudul "Quantum", kini karya tersebut bernilai 7 juta dollar Amerika. Semenjak 2017 hingga

2020 semakin banyak individu atau korporasi yang tertarik terjun pada bisnis NFT ini, beberapa diantaranya adalah Nike dan NBA. Nike sebagai contoh, meluncurkan NFT berupa sepatu sneakersnya dalam bentuk digital dan NBA menggunakan NFT untuk membuat kartu-kartu pemain basketnya dalam bentuk digital yang

nantinya dapat dikoleksi layaknya kartu basket konvensional. Melihat trendnya maka NFT diprediksi dapat diimplementasikan ke dalam berbagai jenis hal lain, termasuk di antaranya informasi terekam apapun dalam bentuk digital[1].

NFT menjadi topik perbincangan hangat di twitter setelah seorang remaja berhasil mendapatkan miliaran rupiah dengan menjual foto selfie yang dijadikan NFT. Hal tersebut menjadi pro dan kontra di kalangan masyarakat yang tentunya menimbulkan opini pribadi. NFT sederhananya merupakan sekumpulan data yang tersimpan dalam buku besar digital yang dikenal sebagai rantai blok atau *blockchain*. NFT kurang lebih seperti halnya mata uang digital seperti bitcoin, NFT berjalan pada platform blockchain. Perbedaannya, jika pada mata uang kripto seperti bitcoin, coin tersebut, yang sebenarnya adalah berupa kumpulan kode, dapat dipecah menjadi banyak bagian, sedangkan NFT bersifat non-fungible. Yang artinya tidak dapat dipecah layaknya koin. Karena pada NFT, kumpulan kode tersebut dapat diembed dengan arsip digital sehingga membuat kumpulan kode pada NFT menjadi unik satu dengan lainnya. Hal inilah yang menjadi pembeda paling dasar antara NFT dengan mata uang kripto konvensional[1].

Twitter merupakan situs jejaring sosial yang memungkinkan penggunaannya untuk bertukar informasi dalam bentuk teks dengan batasan karakter yang telah ditentukan. Kehadiran twitter membuat para pengguna bisa bebas menyuarakan pendapatnya. Selain digunakan untuk sarana menyuarakan opini atau pendapat, twitter juga bisa digunakan sebagai media untuk membaca berita dengan syarat, pengguna yang ingin memperoleh *update* informasi atau berita dari pengguna lain harus menjadi pengikut atau *followers* akun tersebut terlebih dahulu[2]. Media sosial twitter sebagai sarana informasi banyak dimanfaatkan oleh pengguna sebagai media promosi untuk menginformasikan kepada khalayak umum. Salah satunya adalah sebagai media promosi karya seni digital seperti NFT (*Non Fungible Token*).

Para kreator NFT di Indonesia pada umumnya juga banyak yang menggunakan media sosial twitter sebagai media promosi. Salah satunya adalah Ghozali yang mempromosikan karya NFT-nya yang diberi nama Ghozali Everyday. Uniknya lagi NFT tersebut berisi kumpulan foto *selfie*-nya selama beberapa tahun terakhir, tak ayal publik twitter banyak yang memberikan opini mereka

tentang foto *selfie* Ghozali yang dijadikan NFT. Banyaknya komentar publik yang menyampaikan opininya bisa dimanfaatkan untuk mencari sebuah informasi. Tetapi dalam pemanfaatannya dibutuhkan sebuah analisis yang tepat, sehingga bisa didapatkan informasi untuk membantu banyak pihak. Salah satu metode yang bisa digunakan untuk menganalisis opini publik dari twitter adalah dengan menggunakan metode analisis sentimen.

Analisis sentimen ialah sebuah proses menemukan opini *user* mengenai beberapa topik yang disampaikan pengguna[3]. Pendapat yang disampaikan pengguna ada yang bersentimen positif dan negatif. Sentimen positif dan negatif dari suatu *tweet* dapat ditentukan dan dikerjakan secara manual, namun dilihat dari banyaknya jumlah pengguna, maka opini yang dihasilkan juga semakin banyak, dibutuhkan juga waktu dan usaha yang semakin banyak pula. Oleh karena itu dibutuhkan alat atau mesin yang bisa menganalisis *tweet* dan mengklasifikasikan *tweet* yang banyak tersebut ke dalam sentimen positif atau negatif secara otomatis, karena menganalisis *tweet* secara manual kurang efektif lagi. Metode yang bisa dipergunakan untuk analisis sentimen adalah metode *Naive Bayes Classifier*. Metode tersebut merupakan metode pengklasifikasian statistik yang dapat dipergunakan untuk memprediksi probabilitas keanggotaan suatu kelas, selain itu metode ini mempunyai kecepatan serta akurasi tinggi saat diaplikasikan pada data yang banyak.

2. TINJAUAN PUSTAKA DAN TEORI

2.1. Tinjauan Pustaka

Penelitian yang dilakukan oleh Yuda Septian Nugroho memuat masalah tentang penumpukan data mahasiswa yang berpengaruh pada pencarian informasi terhadap data tersebut. Sistem ini digunakan untuk menentukan klasifikasi kelulusan mahasiswa universitas Dian Nuswantoro Fakultas Ilmu Komputer Angkatan 2009 serta mengukur akurasi dengan *Naive Bayes*. Hasil nilai akurasi terhadap kelulusan sebesar 82,08%[4].

Penelitian yang dilakukan Ghulam Asrofi Buntoro yang membandingkan 2 metode pengklasifikasian untuk mencari akurasi tertinggi antara dua metode untuk mengukur akurasi tentang opini masyarakat terhadap calon gubernur DKI Jakarta pada tahun 2017 dengan 300 dataset. Hasil penelitian metode *Naive Bayes* mendapat akurasi

lebih tinggi dengan nilai akurasi 95%, presisi 95% dan recall 95%[5].

Penelitian yang dilakukan Dewi Simanjuntak dan Anita Sindar memuat masalah tentang keterbatasan jumlah tenaga medis bisa terbantu oleh aplikasi ini tanpa ada maksud untuk menggantikan Pakar. Sistem ini digunakan dengan tujuan untuk deteksi awal penyakit gizi buruk pada balita dengan mempergunakan algoritma *Naive Bayes Classifier*. Hasil dari penelitian ini menunjukkan hasil klasifikasi *naive bayes* merupakan jenis penyakit gizi buruk yang diderita balita serta dapat dipergunakan sebagai informasi deteksi awal gizi buruk. Hasil diagnosa penderita gizi buruk memiliki gejala nomer G3, G6, G12, G13 diperoleh klasifikasi perkalian tertinggi yaitu nilai 0,0013168617, jenis penyakit Kwashiorkor[6].

Penelitian yang dilakukan Bambang Satrio Gandhi, Dyah Ayu Megawaty, dan Debby Alita memuat masalah tentang penentuan peringkat kelas yang masih secara manual dan kurang efektif. Hasil dari penelitian ini, penentuan peringkat kelas menggunakan *Naive Bayes Classifier* dengan probabilitas didapatkan hasil akurasi 66.94% mempergunakan Rapidminer. Hasil pengujian didapat dengan hasil yang baik dengan *Blackbox* sebagai fungsional dan ISO 25010 sebagai pengujian *usability* dan *performance*[7].

Penelitian yang dilakukan Ni Made Ayu Juli Astari memuat masalah tentang Permasalahan Virus corona. Penelitian ini bertujuan untuk melakukan analisis dokumen untuk mengetahui sentimen positif atau negatif dari masyarakat. Data yang dipergunakan berupa tweet yang diperoleh dari Twitter tentang dampak virus Corona. Data yang telah diperoleh dipecah menjadi dua bagian data *training* dan data *test*. Penelitian ini mempergunakan Metode *Naive Bayes Classifier*. Hasil penelitian dapat disimpulkan bahwa metode Naive Bayes dapat mengklasifikasikan dokumen *tweet* dengan nilai akurasi 67% dan *error rate* sebesar 33%[8].

2.2. NFT (Non Fungible Token)

NFT adalah bagian dari Blockchain Ethereum. Blockchain Ethereum adalah salah satu jaringan blockchain diantara kurang lebih puluhan atau ratusan blockchain yang sekarang aktif beroperasi. Blockchain sendiri adalah buku besar tetap dan saling terhubung dan berbagi untuk melakukan pencatatan transaksi aset dalam

jaringan bisnis. Aset dalam ini dapat berwujud maupun tidak berwujud seperti hak intelektual, paten, hak cipta, dan merek. Dapat dikatakan seluruh entitas yang memiliki nilai dapat dilacak dan diperjualbelikan dalam jaringan blockchain. Sehingga dalam Bahasa sederhananya, informasi apapun dapat dimasukkan ke dalam jaringan blockchain.[9] Lebih jauh, pada konteks NFT, blockchain tersebut tidak hanya kumpulan kode namun berupa token individual yang didalamnya dapat diembedd informasi tambahan berupa file digital atau arsip digital sehingga kemudian arsip digital tersebut memiliki nilai yang dapat diperjual belikan. Tidak hanya sekedar kumpulan kode[1].

2.3. Naive Bayes Classifier

Pada algoritma *Naive Bayes*, bagian dokumen digambarkan dengan atribut $x_1, x_2, x_3, \dots, x_n$ dimana x_1 adalah kata pertama, x_2 adalah kata kedua, x_3 adalah kata ketiga, hingga x_n [10]. Sementara itu V merupakan kumpulan kelas *tweet*. Saat klasifikasi, algoritma mencari kemungkinan tertinggi dari keseluruhan kelas yang di tes, berikut persamaannya

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \dots\dots\dots (2.1)$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ bernilai konstan untuk semua kategori V_j , sehingga persamaan dapat ditulis seperti dibawah ini

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \dots\dots\dots (2.2)$$

Persamaan tersebut masih bisa disederhanakan menjadi sebagai berikut

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^n P(x_i | V_j) P(V_j) \dots\dots\dots (2.3)$$

Keterangan:

- V_j =kategori tweet $j = 1, 2, \dots, n$
- j_1 =kategori tweet bersentimen positif
- j_2 =kategori tweet bersentimen negatif
- $P(x_1 | V_j)$ =probabilitas x_1 kategori V_j
- $P(V_j)$ =probabilitas V_j

Untuk $P(V_j)$ dan $P(x_i|V_j)$ dihitung pada saat di mana persamaannya dapat dilihat pada persamaan 2.4 dan 2.5.

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \dots\dots\dots (2.4)$$

$$P(x_i|V_j) = \frac{nk+1}{n+|kosakata|} \dots\dots\dots (2.5)$$

Keterangan:

- $|docs\ j|$ =jumlah dokumen kategori j
- $|contoh|$ =jumlah dokumen semua kategori
- nk = jumlah frekuensi kemunculan setiap kata
- n =jumlah frekuensi setiap kategori V_j
- $|kosakata|$ =jumlah semua kata semua kategori

2.3. Twitter dan Twitter API

Twitter merupakan situs berbasis *web* yang dipegang serta dijalankan oleh perusahaan Twitter Inc yang menyediakan media sosial berbentuk *microblogging* yang memungkinkan pengguna membaca dan menyampaikan pesan yang sering disebut dengan *tweet*. Media sosial twitter biasa dipergunakan untuk mengirim pendapat atau opini dari pengguna mengenai sebuah layanan, produk, atau hal-hal lainnya. Pesan atau *tweet* yang ditulis oleh pengguna memiliki batasan hingga 140 karakter pada beranda *user*[11].

Tweet bisa ditujukan secara umum, tetapi pengirim bisa membatasi pesan yang diterima hanya daftar teman saja atau pengikut (*followers*). Pengguna dapat menjalankan twitter melalui laman *web* twitter atau melalui aplikasi yang tersedia pada ponsel. Pengguna dapat menggunakan simbol tagar (#) untuk menuliskan *tweet* berdasarkan dari topik yang diinginkan. Sementara itu untuk menyebut pengguna lain, dapat mempergunakan simbol *mention* (@). Fitur lain yang disediakan oleh twitter adalah sebagai berikut.

1. *Home*
 Pada menu *home*, pengguna dapat melihat *tweet* yang dikirim oleh pengguna yang sudah diikuti (*follow*).
2. *Profile*
 Pada menu *profile*, pengguna lain dapat melihat data diri beserta *tweet* yang telah ditulis.
3. *Follow*
Follow menjadi salah satu fitur utama pada media sosial twitter. Fitur ini memungkinkan pengguna untuk menjalin pertemanan atau kontak dengan pengguna lain. Setiap *tweet* yang diunggah dari seorang pengikut dapat dilihat di beranda pengikut tersebut.
4. *Retweet*
Retweet menjadi fitur yang berguna untuk membagikan ulang *tweet* dari pengguna lain, sehingga dapat ditampilkan di beranda pribadi.
5. *Trending*
Trending menjadi fitur yang menyajikan topik yang sedang populer yang sedang banyak diperbincangkan khalayak twitter.
6. *Mentions*
 Fitur *mentions* berfungsi agar pengguna dapat menandai orang yang terlibat dalam percakapan.
7. *Favorite*
 Fitur ini memiliki fungsi untuk menandai *tweet* sebagai favorit agar tidak hilang dari halaman.
8. *Direct Message*
 Fitur ini memiliki fungsi seperti SMS atau merupakan fitur untuk mengirim pesan secara pribadi antar pengguna.
9. *Hashtag*
 Fitur *hashtag* atau tagar (#) memiliki fungsi agar pengguna dapat mencari topik serupa yang dikirimkan oleh pengguna lain.
10. *List*
 Pengguna twitter dapat menyesuaikan *tweet* yang akan ditampilkan di *timeline*.

Untuk mendapatkan data *tweet* dari twitter ada beberapa cara, salah satunya adalah dengan memanfaatkan Twitter API[12]. Twitter API (*Application Programming Interface*) adalah fungsi yang dapat dipergunakan oleh *developer software* untuk mengerjakan data ketika ingin membuat *software*. Twitter API terdiri dari dua jenis, yaitu Twitter *Search* API serta Twitter *Streaming* API. Perbedaan dari kedua API tersebut adalah Twitter *Search* API digunakan untuk pencarian *tweet* yang sudah berlalu, sedangkan Twitter *Streaming* API digunakan untuk pencarian *tweet* yang akan datang. Untuk dapat mengakses Twitter API, pengguna harus terlebih dahulu mendaftar ke situs Twitter *Developer*. Setelah berhasil mendaftar, pengguna akan diberikan kode unik dari twitter yang berfungsi sebagai kunci untuk mengakses berbagai macam *tweet* dengan menggunakan suatu *keyword* tertentu[13].

2.4. Confusion Matrix

Confusion matrix adalah tabel pencatatan hasil kinerja klasifikasi dengan kuantitas *confusion matrix* bisa diringkas menjadi nilai akurasi dan laju *error*. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, maka bisa diketahui akurasi hasil prediksi, serta dengan mengetahui jumlah data yang diklasifikasikan secara salah, maka bisa diketahui laju *error* dari prediksi yang telah dilangsungkan. Dua kuantitas tersebut dipergunakan sebagai matriks kinerja klasifikasi[14].

Berikut beberapa ukuran yang bias dipergunakan untuk mengevaluasi model klasifikasi[15]

1. Akurasi : menyatakan persentase dari data uji yang diklasifikasikan dengan benar. Untuk menghitung akurasi digunakan rumus

$$Akurasi = \frac{\text{jumlah data diprediksi benar}}{\text{jumlah data yang diprediksi}}$$

atau

$$Akurasi = \frac{TP + TN}{(TP + TN + FP + FN)}$$

2. *Error rate* : menyatakan persentase data uji yang diklasifikasikan dengan tidak benar. Untuk menghitung *error rate* digunakan rumus

$$Error\ Rate = \frac{FP + FN}{FP + FN + TP + TN}$$

3. *Recall* : menyatakan persentase data positif yang diklasifikasikan dengan benar. Untuk menghitung *recall* digunakan rumus

$$Recall = \frac{TP}{TP + FN}$$

4. *Specificity* : menyatakan persentase data negatif yang diklasifikasikan dengan benar. Untuk menghitung *specificity* digunakan rumus

$$Specificity = \frac{TN}{FP + TN}$$

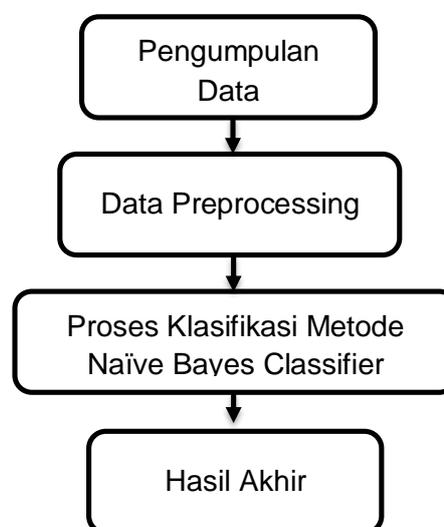
5. *Precision* : beberapa persentase data yang dilabeli positif adalah memang benar positif. Untuk menghitung *precision* digunakan rumus

$$Precision = \frac{TP}{TP + FP}$$

3. METODOLOGI PENELITIAN

3.1. Skema Alur Penelitian

Pada penelitian ini mempunyai tahapan alur penelitian yang akan dilakukan, alur disusun dengan struktur dan divisualisasikan dalam bentuk gambar dengan tahapan yang berurutan. Berikut merupakan alur penelitian.

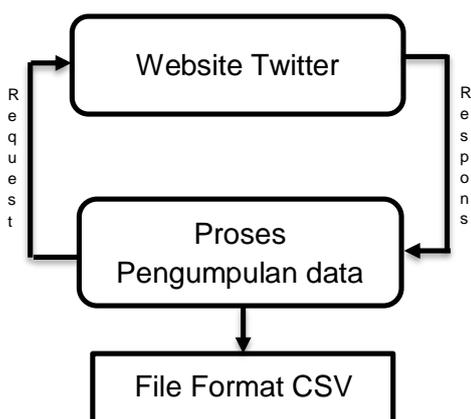


Gambar 1. Tahapan Penelitian

3.2. Pengumpulan Data

Penelitian ini menggunakan data *tweet* yang terdapat pada laman *twitter.com*. Data yang dipergunakan berupa teks yang diambil dengan menggunakan twitter API. Tahap awal untuk mengambil data dari twitter adalah dengan mendapatkan *key number*, *secret number*, *acces token*, dan *acces token secret* dari Twitter API terlebih dahulu. Untuk bisa mengakses twitter API harus melakukan pengajuan ke situs <https://developer.twitter.com/en/apps>.

Setelah mendapatkan akses twitter API, tahap selanjutnya adalah proses *crawling* data dengan kata kunci utama NFT.



Gambar 2. Proses Pengumpulan Data

Data yang diperoleh adalah data twitter dengan kata kunci NFT.

3.3. Data Preprocessing

Data tweet yang diperoleh akan diproses pada tahap preprocessing yang terdiri dari :

1. Case Folding

Case *folding* yaitu proses penyeragaman ukuran huruf pada data tweet menjadi lowercase atau huruf kecil. Sampel penggunaan case folding bisa dilihat pada tabel 1.

TABEL 1. SAMPEL CASE FOLDING

Teks Input	Teks Output
Kini banyak orang yang membuat NFT	kini banyak orang yang membuat nft

untuk mencoba peruntungan seperti Ghozali Everyday	untuk mencoba peruntungan seperti ghozali everyday
--	--

2. Cleansing

Cleansing merupakan proses pembersihan atau penghapusan kata yang tidak penting. Contoh kata yang tidak penting seperti “@”, “#”, “rt”, dan lain sebagainya.

3. Stemming

Stemming merupakan proses perubahan kata yang memiliki imbuhan dikembalikan ke kata dasar yang sesuai standar bahasa Indonesia. *Stemming* menjadi teknik pemrosesan teks dasar yang sering digunakan untuk mengembangkan kemampuan lebih lanjut dalam pengklasifikasian teks. Sampel proses *Stemming* bisa dilihat pada tabel 2.

TABEL 2. SAMPEL STEMMING

Teks Input	Teks Output
Kini banyak orang yang membuat NFT untuk mencoba peruntungan seperti Ghozali Everyday	kini banyak orang yang buat nft untuk coba untung seperti ghozali everyday

4. Tokenizing

Tokenizing yaitu proses pembagian data teks menjadi beberapa bagian kata. Sampel proses *Tokenizing* bisa dilihat pada tabel 3.

TABEL 3. SAMPEL TOKENIZING

Teks Input	Teks Output
kini banyak orang yang buat nft untuk coba untung	kini, banyak, orang, yang, buat, nft, untuk, coba, untung, seperti,

seperti ghozali everyday	ghozali, everyday
--------------------------	-------------------

3.4. Proses Klasifikasi

Setelah data selesai di proses melalui tahap *preprocessing* langkah selanjutnya yaitu proses klasifikasi data dengan metode *Naive Bayes Classifier*.

4. HASIL DAN PEMBAHASAN

Pada penelitian yang mempergunakan data dari twitter dengan memanfaatkan library *tweepy* dengan kata kunci "NFT" menggunakan program berikut :

Data tweet sebanyak 300 dipisah menjadi dua bagian dengan jumlah 154 data positif dan 146 data negatif disimpan dengan format *csv*. Selanjutnya data dilakukan *preprocessing*, dilatih serta diuji sehingga akan didapatkan nilai akurasi, *recall*, *precision*, dan *f1-score*. Berikut merupakan tabel yang berisi sampel data yang telah melalui tahap *preprocessing*.

TABEL 4. SAMPEL DATA TRAINING

Data Training	Label
NFT menjadi aset digital yang tidak dapat ditukar	Negatif
bodo amat lah dengan data, data e-KTP aja diupload di NFT	Negatif
karya NFT kamu bagus	Positif
ghozali meraup keuntungan dari NFT	Positif

Selanjutnya, data *preprocessing* dan diberi label secara manual akan diuji untuk mengukur akurasi. Pengetesan akurasi pada proses klasifikasi dilangsungkan agar dapat diketahui tingkat akurasi yang didapatkan dengan metode *Naive Bayes Classifier*. Teknik uji yang dipergunakan dalam pengetesan adalah *confusion matrix* yang diperlukan sebagai pembandingan data prediksi dengan data asli yang sebelumnya telah diinput. Berikut merupakan representatif kelas data dan kelas prediksi.

TABEL 5. TABEL CONFUSION MATRIX

Aktual	Prediktif	
	Negatif	Positif
Negatif	TN	FP
Positif	FN	TP

Pada tabel 5 dijelaskan bahwa *confusion matrix* merupakan hasil dari prediksi menggunakan metode *Naive Bayes Classifier* dengan mengukur performa tiap kelas dengan menghitung *precision*, *recall*, dan *f1-score*. Selanjutnya dilakukan penginputan data sesuai tabel 2.

TABEL 6. HASIL CONFUSION MATRIX

Aktual	Prediktif	
	Negatif	Positif
Negatif	123	0
Positif	23	154

Berdasarkan tabel 6, jumlah data yang didapatkan dan diprediksi oleh sistem sebanyak 154 data *true positive*, 123 data *true negative*, 0 data *false positive*, dan 23 data *false negative*.

Setelah didapatkan data yang telah diprediksi kemudian dilakukan pengujian akurasi dengan *Naive Bayes Classifier* dengan bantuan *software jupyter notebook*. Hasil pengujian bisa dilihat pada gambar 3.

```

test time: 0.000s
accuracy: 0.923
          precision  recall  f1-score  support
negatif    0.83     1.00     0.91     10
positif    1.00     0.88     0.93     16

accuracy                0.92     26
macro avg    0.92     0.94     0.92     26
weighted avg 0.94     0.92     0.92     26
    
```

Gambar 3. Hasil Pengujian *Jupyter Notebook*

Pengujian akurasi dengan bantuan *software jupyter notebook* nilai akurasi yang didapatkan sebesar 92.3%. Pada kelas negatif menunjukkan *precision* sebesar 83%, *recall* sebesar 100%, dan *f1-score* sebesar 91%. Sedangkan pada kelas

positif menunjukkan *precision* sebesar 100%, *recall* sebesar 88%, dan *f1-score* sebesar 93%.

Hasil perhitungan manual *confusion matrix* dapat dilihat pada tabel 7.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{(TP+TN+FP+FN)} \\
 &= \frac{154+123}{(154+123+0+23)} \\
 &= 0,9233 \times 100\% \\
 &= 92,33 \\
 \text{Recall Positif} &= \frac{TP}{TP+FN} = \frac{154}{154+23} \\
 &= 0,88 \times 100\% \\
 &= 88\% \\
 \text{Recall Negatif} &= \frac{TN}{TN+FP} = \frac{123}{123+0} \\
 &= 1 \times 100\% \\
 &= 100\% \\
 \text{Precision Positif} &= \frac{TP}{TP+FP} = \frac{154}{154+0} \\
 &= 1 \times 100\% \\
 &= 100\% \\
 \text{Precision Negatif} &= \frac{TN}{TN+FN} = \frac{123}{123+23} \\
 &= 0,8424 \times 100\% \\
 &= 84,24\% \\
 \text{f1 score Positif} &= 2 \times \frac{\text{recall positif} * \text{precision positif}}{\text{recall positif} + \text{precision positif}} \\
 &= 2 \times \frac{0,88 * 1}{0,88 + 1} = 0,9361 \times 100\% \\
 &= 93,61\%
 \end{aligned}$$

$$\begin{aligned}
 \text{f1 score Negatif} &= 2 \times \frac{\text{recall negatif} * \text{precision negatif}}{\text{recall negatif} + \text{precision negatif}}
 \end{aligned}$$

$$\begin{aligned}
 &= 2 \times \frac{1 * 0,84}{1 + 0,84} = 0,9130 \times 100\% \\
 &= 91,30\%
 \end{aligned}$$

TABEL 7. HASIL PENGUJIAN

	Nilai
Akurasi	92,33%
Recall Positif	88%
Recall Negatif	100%
Precision Positif	100%
Precision Negatif	84,24%
f1-score Positif	93,61%
f1-score Negatif	91,30%

Hasil dari tabel 3 dapat disederhanakan seperti pada tabel 5.

TABEL 8. HASIL PENGUJIAN KESELURUHAN

	Nilai
Akurasi	92,33%
Recall	94%
Precision	92,12%
f1-score	92,45%

Untuk penelitian ini dengan klasifikasi data *tweet* dengan kata kunci "NFT" menggunakan metode *Naive Bayes Classifier* mendapatkan nilai akurasi sebesar 92,33%, nilai *recall* sebesar 94%, nilai *precision* sebesar 92,12%, dan nilai *f1 - score* sebesar 92,45%.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Penelitian ini berhasil menerapkan algoritma *Naive Bayes Classifier* untuk menganalisis sentimen pengguna twitter tentang NFT. Tingkat keakuratan dari metode *Naive Bayes Classifier* dalam melakukan klasifikasi *tweet* ke dalam *tweet* bersentimen positif atau negatif dengan kata kunci NFT menunjukkan angka sebesar 92,33% untuk nilai akurasi, angka sebesar 94%

untuk nilai *recall*, angka sebesar 92,12% untuk nilai *precision*, dan angka sebesar 92,45% untuk nilai *f1 – score*. Hasil tersebut menunjukkan bahwa metode *Naive Bayes Classifier* mampu mengklasifikasikan data *tweet* NFT dengan sangat baik (*excellent classification*), karena mampu menghasilkan tingkat akurasi yang tinggi pada periode ini.

5.2. Saran

Untuk mendapatkan hasil yang lebih baik kedepannya, perlu ada beberapa saran yang bisa diterapkan pada penelitian selanjutnya.

1. Dapat dibandingkan dengan metode-metode lain, seperti *Support Vector Machine*, *K-Nearest Neighbor* dan sebagainya untuk memperoleh tingkat akurasi dari setiap metode.
2. Menggunakan data dari sumber lain selain twitter, bisa melalui media sosial seperti instagram atau facebook, bisa melalui *e-commerce* seperti tokopedia atau shopee, atau *platform* lainnya.

DAFTAR PUSTAKA

- [1] F. G. Becker *et al.*, "NFT (NON-FUNGIBLE TOKEN): MASA DEPAN ARSIP DIGITAL? ATAU HANYA SEKEDAR BUBBLE?," *Syria Stud.*, vol. 7, no. 1, pp. 37–72, 2015.
- [2] B. Kurniawan, M. A. Fauzi, and A. W. Widodo, "Klasifikasi Berita Twitter Menggunakan Metode Improved Naïve Bayes," vol. 1, no. 10, pp. 1193–1200, 2017.
- [3] D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," pp. 8–9, 2018.
- [4] N. Y. Septian, "Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro," *J. Semant.* 2013, pp. 1–11, 2009.
- [5] G. A. Buntoro, "Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine," *J. Din. Inform.*, vol. 5, no. 2, pp. 1–13, 2016.
- [6] A. S. R. Sinaga and D. Simanjuntak, "Sistem Pakar Deteksi Gizi Buruk Balita Dengan Metode Naïve Bayes Classifier," *J. Inkofar*, vol. 1, no. 2, pp. 54–60, 2020, doi: 10.46846/jurnalinkofar.v1i2.110.
- [7] B. S. Gandhi, D. A. Megawaty, and D. Alita, "Aplikasi Monitoring dan Penentuan Peringkat Kelas Menggunakan Naive Bayes Classifier," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 1, pp. 54–63, 2021, doi: 10.33365/jatika.v2i1.722.
- [8] N. M. A. J. Astarti, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 27–29, 2020, doi: 10.30864/jsi.v15i1.332.
- [9] L. Ante, "The non-fungible token (NFT) market and its relationship with Bitcoin and Ethereum," *SSRN Electron. J.*, pp. 216–224, 2021, doi: 10.2139/ssrn.3861106.
- [10] Mahardika Tania Nitami. Herny Februariyanti, "Analisis Sentimen Ulasan Ekpedisi J&T Ekspres Menggunakan Algoritma Naive Bayes," *Anal. Sentimen Ulas. Ekpedisi J&T Ekspres Menggunakan Algorith. Naive Bayes*, vol. 5, no. 1, pp. 20–29, 2022.
- [11] A. E. Holton, "Weller et al., Twitter and Society," *J. Media Innov.*, vol. 1, no. 1, pp. 134–137, 2014, doi: 10.5617/jmi.v1i1.825.
- [12] D. Rustiana and N. Rahayu, "Analisis sentimen pasar otomotif mobil," *J. SIMETRIS*, vol. 8, no. 1, pp. 113–120, 2017.
- [13] S. S. Sohail, M. M. Khan, and M. Arsalan, "Crawling Twitter data through API: A technical / legal perspective".
- [14] H. M. Al-Lauroussi, "Implementasi algoritma Naive Bayes Sebagai Proses Seleksi Penerima Beasiswa Libyan Embassy Berbasis WEB," p. 99, 2015.
- [15] M. Riyyan and H. Firdaus, "PERBANDINGAN ALGORITME NAÏVE BAYES DAN KNN TERHADAP DATA PENERIMAAN BEASISWA (Studi Kasus Lembaga Beasiswa Baznas Jabar)," vol. 5, no. 1, pp. 1–10, 2022.