Analisis Sentimen Terhadap Pariwisata di Masa *Covid-19* Menggunakan Naïve Bayes

Daniel Arsa[#1], Indra Weni[#2], Agil Fahreza[#3]

[#]*Program Studi Sistem Informasi, Universitas Jambi*
Jl. Jambi - Muara Bulian Mendalo Darat, Indonesia
[1]danielarsa@unja.ac.id
[2]indra_weni@unja.ac.id
[3]agil_fahreza@gmail.com

Abstract*— The tourism industry is a sector that has the potential to be developed as a source of state income. The tourism sector is the second largest source of state revenue after taxes. The arrival of tourists in an area has an impact on local resident who have provided prosperity and prosperity in the vicinity. The Covid-19 pandemic that has occurred in the world has had a very broad impact, including on the tourism sector and the creative economy. The decline in foreign tourist arrivals resulted in huge losses. This creates a public response to government policies. The community's response to tourism can be seen in social media. One of the most popular social media is Twitter. Obtained as many as 3000 tweet data that will be classified using the Naïve Bayes algorithm. Naive Bayes is a text mining technique to build a simple classifier model but has high accuracy in classifying. With the use of the Naive Bayes algorithm in this study, the results obtained are 62% accuracy values with an average value of 62% precision, 62% recall value, and 62% F1-score value.

***Keywords*— Sentiment analysis, Covid-19, Naïve Bayes, tourism, text mining**

Abstrak*— Industri pariwisata merupakan sektor yang potensial untuk dikembangkan sebagai salah satu sumber pendapatan negara. Sektor pariwisata menjadi sumber pendapatan negara kedua terbesar setelah pajak. Kedatangan wisatawan pada suatu daerah memiliki dampak bagi penduduk setempat yang telah memberikan kemakmuran dan kesejahteraan di sekitarnya. Pandemi Covid-19 yang terjadi di dunia memberikan dampak sangat luas, termasuk sektor pariwisata dan ekonomi kreatif. Menurunnya kunjungan wisatawan mancanegara, mengakibatkan kerugian yang sangat besar. Hal ini menimbulkan respon masyarakat atas kebijakan pemerintah. Respon masyarakat terhadap pariwisata dapat dilihat dari media sosial. Salah satu media sosial yang banyak penggunanya adalah Twitter. Didapatkan sebanyak 3000 data *tweet* yang akan diklasifikasi dengan menggunakan algoritme Naïve Bayes. Naive Bayes adalah teknik *text mining* untuk membangun model pengklasifikasi yang sederhana tetapi memiliki keakuratan yang tinggi dalam mengklasifikasi. Dengan digunakannya algoritme Naive Bayes pada penelitian ini, maka didapatkan hasil nilai akurasi sebesar 62% dengan nilai rata-rata pada *precision* sebesar 62%, nilai *recall* sebesar 62%, dan nilai *F1-score* sebesar 62%.

***Kata Kunci*— Analisis sentimen, Covid-19, Naïve Bayes, pariwisata, penambangan kata**

I. PENDAHULUAN

Coronavirus disease 2019 (Covid-19) adalah virus yang disebabkan oleh *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-CoV-2) yang pertama kali ditemukan di Wuhan, Cina. Virus ini menular dari manusia ke manusia dan telah menyebar secara luas di seluruh dunia. *World Health Organization* (WHO) menetapkan bahwa virus Covid-19 menjadi kedaruratan kesehatan masyarakat yang meresahkan dunia. Penambahan jumlah kasus Covid-19 berlangsung cukup cepat dan terjadi penyebaran antar negara [1]. Dalam data terbaru WHO (*World Health Organization*) per tanggal 17 Januari 2021, jumlah kasus terkonfirmasi positif Covid-19 mencapai 92.542.904 kasus, sedangkan di Indonesia jumlah kasus terpapar Covid-19 mencapai 907.929 kasus, kemungkinan akan bertambah setiap harinya. Penyebaran virus Covid-19 ini menyebabkan turunnya roda penggerak ekonomi di seluruh dunia dan di berbagai sector, salah satunya sektor pariwisata.

Pandemi Covid-19 yang terjadi di Indonesia sendiri memberikan dampak sangat luas, termasuk sektor pariwisata dan ekonomi kreatif. Menurunnya kunjungan wisatawan mancanegara, mengakibatkan kerugian yang sangat besar. Keterpurukan ini dikarenakan sektor pariwisata sangat mengandalkan pergerakan manusia. Dalam pandemi ini pemerintah masih minim memberikan kebijakan dalam sektor pariwisata, terutama dalam proses penyampaian sosialisasinya kepada publik. Kejelasan dan detail akan informasi mengenai langkah yang diambil dalam merespon penanganan sektor pariwisata di masa pandemi ini oleh Kementrian Pariwisata dan Ekonomi Kreatif (Kemenparekraf) masih belum jelas. Hal ini menimbulkan respon masyarakat atas kebijakan pemerintah.

Untuk melihat respon masyarakat terhadap pariwisata dapat dilihat dari berbagai sarana dan media, salah satunya yaitu melalui media sosial Twitter. Twitter ialah *microblog* terpopuler di dunia saat ini dengan total pengguna mencapai 324 juta. Twitter saat ini dapat mengirim teks hingga 280 karakter. Kemudahan dalam penggunaan Twitter merupakan alasan Twitter lebih digemari dibandingkan media sosial lainnya. Isi *tweet* dijadikan sebagai sumber data opini dan sentimen masyarakat terhadap kegiatan pariwisata yang dapat dianalisis dan dijadikan bahan evaluasi bagi pihak terkait.

Analisis sentimen atau *opinion mining* adalah bidang studi yang menganalisis pendapat orang, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atributnya [2]. Data yang diambil menggunakan teknik *crawling* dengan memanfaatkan *Application Programming Interface* (API) yang disediakan oleh Twitter. Berdasarkan data dari Datereportal saat ini pengguna internet aktif di Indonesia sekitar 175,4 juta dan total pengguna Twitter mencapai sekitar 63,6%. Analisis sentimen akan dilakukan dengan mengelompokkan polaritas dari teks ke dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam bentuk kalimat atau dokumen yang bersifat positif, negatif, atau netral. Data yang telah diolah akan diklasifikasi dengan metode Naïve Bayes, lalu akan diklasifikasikan dengan model *confusion matrix*.

Algoritme Naive Bayes merupakan metode klasifikasi dengan statistik berdasarkan pada teorema Bayes [3]. Naive Bayes merupakan sebuah pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan [4]. Untuk beberapa jenis model probabilitas, *Naive Bayes clasifie*r dapat dilatih dengan sangat efisien dalam pengaturan *supervised learning*. Kelebihan dari *Naive Bayes classifier* adalah metode ini hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter yang diperlukan dalam klasifikasi.

Pada penelitian yang sudah dilakukan mengenai analisis sentimen terdapat penelitian tentang analisis sentimen Pemindahan Ibu Kota Negara dengan menggunakan *feature selection*. Penelitian tersebut ingin mengetahui kecenderungan tanggapan masyarakat terkait pemindahan ibu kota negara serta mengetahui rasio terbaik antara *data training* dan *testing* dengan memakai *feature selection* pada algoritme *Support Vector Machine* dan *Naive Bayes*. Tahapan penelitian tersebut terdiri dari pengumpulan data melalui media sosial Twitter, *polarity*, serta *preprocessing* yang terdiri dari proses *transform case*, *cleansing*, *tokenizing*, *filtering*, dan *stemming*. *Feature selection* digunakan untuk meningkatkan nilai akurasi dan selanjutnya memasukan rasio yang telah ditentukan untuk digunakan oleh *data training* dan *testing*. Jumlah data yang diambil pada periode 23 April 2020 hingga 2 Mei 2020 mendapatkan hasil sebanyak 849 data *tweet*. Dalam periode tersebut didapatkan 24,26% sentimen positif dan 75,74% sentimen negatif. Akurasi memakai perangkat lunak Rapid Miner menghasilkan nilai akurasi terbaik untuk algoritme Naive Bayes dengan *feature selection*, ada pada rasio 9:1 dan dengan akurasi 88,24%. Hasil akurasi terbaik untuk algoritme *Support Vector Machine* dengan *feature selection* ada pada rasio 5:5 dengan akurasi 78,77% [5].

Penelitian yang dilakukan ingin mengetahui hasil analisis sentimen masyarakat mengenai pariwisata di masa pandemi dan tingkat akurasi yang didapatkan adalah dari hasil klasifikasi sentimen menggunakan metode Naïve Bayes.

II. METODOLOGI

Metode penelitian yang dilakukan dalam penelitian ini terbagi menjadi beberapa tahapan. Berikut ini adalah penjelasannya.

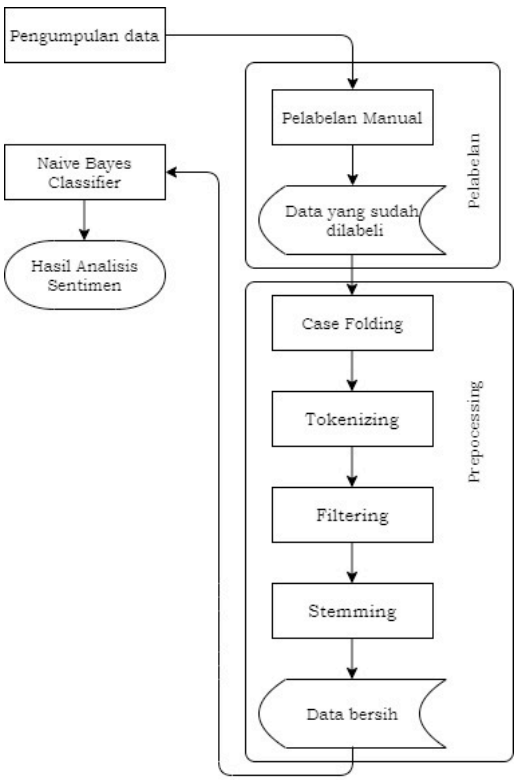
A. Arsitektur Sistem

Sistem yang dibangun ditunjukkan pada Gambar 1. Secara umum sistem ini terdiri dari empat bagian di antaranya adalah pengambilan data, pelabelan, *preprocesing*, dan klasifikasi data.

Proses *text mining* dalam penelitian adalah pengumpulan data, pelabelan, *preprocessing* data, dan pembuatan model klasifikasi [6]. Untuk mengumpulkan data dilakukan proses *crawling* dengan memanfaatkan Twitter API. Data yang telah terkumpul selanjutnya diberikan label secara manual sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Dibantu juga oleh pihak ketiga, yaitu Badan Pengembangan dan Pembinaan Bahasa Indonesia Jambi berdasarkan kelasnya: positif, netral, dan negatif. Selanjutnya dilakukan proses *preprocessing* data, yaitu *case folding, tokenizing, filtering,* dan *stemming.* Setelah melalui *preprocessing* data, maka didapatkan data bersih yang digunakan untuk membuat dan menguji model. Didapatkan hasil akhir berupa analisis sentimen opini pengguna Twitter mengenai penerapan pariwisata selama pandemi Covid-19.

B. Pengumpulan Data

Tahap pengumpulan data dilakukan dengan cara *crawling* melalui program Python dari API Twitter yang didaftarkan.



Gambar 1 Arsitektur sistem

Proses *crawling* dimulai dengan memasukkan *consumerKey, consumerSecret, accessToken,* dan *accessTokenSecret* yang didapat dari API Twitter. Yang dapat dilakukan adalah berupa *keyword*, *mention*, *hashtag*, *reply*, dan rentang tanggal *tweet* dibuat.

Setelah terkoneksi dengan Twitter, program akan meminta data tentang kata kunci maupun *hashtag* yang dicari. Berikutnya, memasukkan ke dalam variabel *keyword* dan akan meminta jumlah data *tweet* yang akan dianalisis, kemudian memasukkan ke dalam variabel *noOfTweet* berupa *integer*. Variabel *tweet* akan melakukan operasi pengambilan data yang telah disimpan pada variabel *searchTerm* dan *noOfSearchTerm*. Pada penelitian ini dimasukkan kata kunci pariwisata dengan jumlah *tweet* yang diambil sebanyak 3026 *tweet*. Data yang terkumpul sebanyak 3026 *tweet* dikumpulkan dan disimpan ke dalam *file* Excel dengan format .xlsx. Pengambilan data diambil per hari sejak dimulainya penelitian ini, yaitu pada bulan Maret 2021 hingga Juli 2021.

C. Preprocessing

Preprocessing merupakan tahap di mana teks akan diseragamkan bentuk dan formatnya agar dapat dipersiapkan menjadi data yang dapat diolah pada tahap selanjutnya. *Text preprocessing* meliputi *case folding*, *tokenizing*, *filtering*, dan *stemming* [7].

1) Case Folding

Pada tahap ini, semua huruf akan diubah menjadi *lowercase*, atau huruf kecil [6].

2) Tokenizing

Tokenizing dalam penelitian ini merupakan tahap pemotongan kalimat menjadi kata, lalu menyusunnya serta menghilangkan URL, *@mention*, dan *hashtag* yang ada pada dokumen (*tweets*). Pada tahap ini juga menghilangkan delimeter seperti tanda titik (.), koma (,), spasi, dan karakter, serta angka dan tanda baca yang ada pada dokumen (*tweets*) [8].

3) Filtering

Membuang kata-kata yang sering muncul secara umum dan kurang relevan dilakukan untuk mengubah kata berimbuhan dari setiap kata yang sudah disaring menjadi kata dasar dengan teks akan dihapus [7]. Proses ini disebut juga dengan *stopword removing*, yaitu menghapus kata-kata yang tidak bermakna dan tidak memiliki pengaruh terhadap analisis sentimen.

4) Stemming

Stemming adalah tahap mencari kata dasar dari tiap kata hasil *filtering* [7].

D. Pemodelan Data

Pemodelan data dibuat dengan tujuan untuk menguji keakuratan prediksi sistem berdasarkan data model yang sudah dibuat. Proses *training* terlebih dahulu dibangun untuk membuat model *probabilitas* dengan menggunakan data *training*. Pada data *training*, data yang digunakan harus sudah terlebih dahulu dilabeli dan juga harus melalui proses *preprocesing* untuk dijadikan *input* metode *Naive Bayes classifier* dan disimpan sebagai model data dengan ekstensi *.pickle* (*dot pickle*). Pada proses *training*, sistem akan membaca data *tweet* yang dimasukkan. Selanjutnya sistem akan melakukan proses ekstraksi fitur kata dan diselesaikan dengan menyimpan model data. Selanjutnya dilakukan pengujian model klasifikasi sentimen yang dihasilkan pada proses pengujian dengan menggunakan data *tweet* baru (*data testing*).

E. Rancangan Pengujian Model

Dalam rancangan pengujian model, model yang digunakan untuk evaluasi adalah *confusion matrix*. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Dilihat pada Gambar 2, *confusion matrix* biasanya dijelaskan dalam kondisi 2x2 *class,* yaitu *true* dan *false*, atau positif dan negatif. Pada penelitian ini menggunakan *confusion matrix* 3x3, akan ada perubahan dalam penilaiannya.

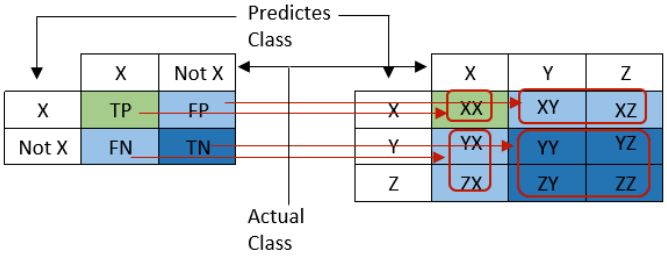
Untuk pengukuran performa klasifikasi, cara yang digunakan adalah menghitung akurasi*, precission*, *recall*, dan *F1-score* [10]. Akurasi merupakan persentase dari total sentimen yang benar dikenali. Untuk menghitung nilai akurasi digunakan persamaan di bawah ini:

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

Precision merupakan perbandingan jumlah data relevan yang ditemukan terhadap jumlah data yang ditemukan. Untuk menghitung nilai *precision* digunakan persamaan di bawah ini:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall merupakan perbandingan jumlah materi relevan yang ditemukan terhadap jumlah materi yang relevan. Untuk menghitung nilai *recall* digunakan persamaan di bawah ini:



Gambar 2 *Extended confusion matrix* [9]

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score merupakan parameter tunggal ukuran keberhasilan *retrieval* yang menggabungkan *recall* dan *precision*. Untuk menghitung nilai *F-measure* digunakan persamaan di bawah ini:

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Total data yang terkumpul serta data yang telah dilabeli secara manual sebanyak 3.026. Data yang dipakai dalam penelitian ini sebanyak 2.086 dikarenakan dalam pelabelan manual 3.026 data terdapat kelas mayoritas yang berpotensi mempengaruhi hasil yang ada. Dari 2.086 data dibagi menjadi 80% data *training* dan 20% data *testing*. Sentimen pada data *training* sejumlah 1.668 dibagi secara manual sesuai kelasnya, 418 data lainnya akan digunakan sebagai data uji. Pelabelan data hasil dilakukan secara manual dalam penelitian ini dan diverifikasi oleh pihak ketiga, yaitu Badan Pengembangan dan Pembinaan Bahasa Indonesia Jambi. Didapatkan hasil akhir dari pelabelan, yaitu 544 *tweet* positif, 747 *tweet* netral, dan 795 *tweet* negatif.

B. Preprocessing

Setelah diberi label manual, selanjutnya masuk ke proses *preprocessing*. Tahap *preprocessing* data perlu dilakukan karena beberapa kalimat *tweet* yang didapatkan tidak sepenuhnya menggunakan kata baku dan menggunakan bahasa Indonesia yang baik. *Preprocessing* dilakukan menggunakan bantuan *library* pada bahasa pemrograman Python. *Preprocessing* data dilakukan dengan tahap *case folding*, *tokenizing*, *filtering*, dan *stemming* sehingga menghasilkan data bersih.

1) Case Folding

Case folding adalah proses mengubah data *tweet* menjadi *lowercase*. Tabel I memuat contoh data penelitian yang dilakukan dalam proses *case folding*.

TABEL I

PREPROCESSING CASE FOLDING

No.	*Tweet*	*Case Folding*
1	Kalian mau pariwisata? Pastkan peduli lindungi udah ada ya, supaya aman dgn cara ini, jngan lupa protocol keshtannya https://t.co/1qlsrxAVqi	kalian mau pariwisata? pastkan peduli lindungi udah ada ya, supaya aman dgn cara ini, jngan lupa protocol keshtannya https://t.co/1qlsrxAVqi

2) Tokenizing

Tokenizing dalam penelitian ini merupakan tahapan dalam memecah *string* atau *input* terhadap suatu teks yang telah melewati tahap *case folding*. Hal tersebut dilakukan berdasarkan tiap kata yang menyusunnya dengan menghilangkan URL, *@mention*, dan *hashtag*. Tabel II berisi contoh proses *tokenizing* yang dilakukan pada penelitian ini.

Pada tahap ini juga dilakukan proses normalisasi, seperti yang ada pada Tabel III, yaitu mengubah kata yang tidak lengkap dan kesalahan dalam pengetikan (*typo*) ke dalam kata yang normal sehingga dapat dipahami dengan baik.

3) Filtering

Filtering disebut juga dengan *stopword removal*. Proses ini membuang kata yang tidak memiliki arti. Pada tahap ini digunakan NLTK (*Natural Language ToolKit*) yang dapat dilihat pada Tabel IV. NLTK adalah *library* yang disediakan oleh Python untuk membangun program analisis teks.

4) Stemming

Stemming adalah tahap mencari *root* (dasar) kata dari tiap kata hasil *filtering* dengan menghapus kata imbuhan di depan maupun imbuhan di belakang kata. Pada Tabel V terdapat beberapa contoh proses *stemming* yang dilakukan, seperti kata "penebangan" menjadi "tebang".

TABEL II

PREPROCESSING TOKENIZING

No.	*Tweet*	*Tokenizing*
1	Kalian mau pariwisata? Pastkan peduli lindungi udah ada ya, supaya aman dgn cara ini, jngan lupa protocol keshtannya https://t.co/1qlsrxAVqi	'kalian', 'mau', 'pariwisata', 'pastkan', 'peduli', 'lindungi', 'udah', 'ada', 'ya', 'supaya', 'aman', 'dgn', 'cara', 'ini', 'jngan', 'lupa', 'protocol', keshtannya

TABEL III

PREPROCESSING NORMALIZATION

No.	*Tweet*	*Normalization*
1	Kalian mau pariwisata? Pastkan peduli lindungi udah ada ya, supaya aman dgn cara ini, jngan lupa protocol keshtannya https://t.co/1qlsrxAVqi	'kalian', 'mau', 'pariwisata', 'pastikan', 'peduli', 'lindungi', 'sudah', 'ada', 'ya', 'supaya', 'aman', 'dengan', 'cara', 'ini', 'jangan', 'lupa', 'protocol', 'kesehtannya'

TABEL IV

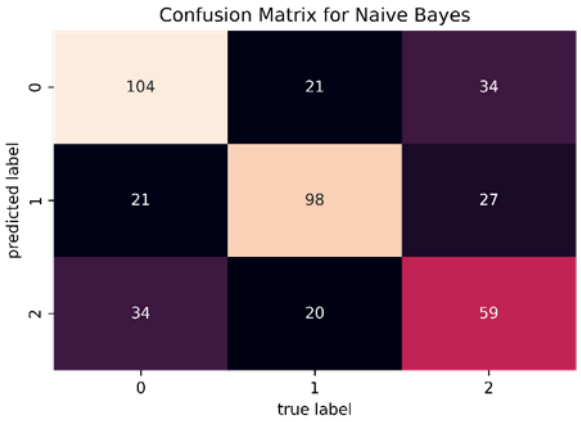
PREPROCESSING FILTERING

No.	*Tweet*	*Filtering*
1	'kalian', 'mau', 'pariwisata', 'pastikan', 'peduli', 'lindungi', 'sudah', 'ada', 'ya', 'supaya', 'aman', 'dengan', 'cara', 'ini', 'jangan', 'lupa', 'protocol', 'kesehtannya'	'pariwisata', 'pastikan', 'peduli', 'lindungi', 'aman', 'lupa', 'protocol', 'kesehatan'

TABEL V

PREPROCESSING STEMMING

No	Kata Imbuhan	Kata Dasar
1	Penebangan	Tebang
	Temannya	Teman
	Keistimewaan	Istimewa
	Laporan	Lapor
	Bergerak	Gerak
	Diharapkan	Harap
	Pastikan	Pasti
	Menciptakan	Cipta
	Pemanfaatan	Manfaat
	Perdagangan	Dagang
	Pembatasan	Batas
	Kesehatan	Sehat



Gambar 3 Hasil *confusion matrix*

C. *Ektrasi fitur*

Setelah data *tweet* melalui proses *preprocessing*, selanjutnya akan dibentuk menjadi sebuah model klasifikasi. Dalam tahapan ekstraksi fitur, tahap awal yang dilakukan adalah mengubah *dataset* penelitian ke dalam representasi vektor.

D. *Implementasi Klasifikasi Naïve Bayes*

Setelah dilakukan *preprocessing* data dan *vectorizer*, selanjutnya dibuat model yang akan digunakan untuk melakukan klasifikasi data uji. Pada proses klasifikasi digunakan data uji sebanyak 20% dari data keseluruhan. Proses klasifikasi dilakukan dengan perhitungan probabilitas antarkalimat terhadap setiap kelas agar dapat menghasilkan dengan jelas prediksi data yang dimasukkan.

E. *Pengujian dan Evaluasi Model Klasifikasi Naïve Bayes*

Untuk mengetahui performa dari Algoritma Naive Bayes, maka dilakukan pengujian terhadap model yang telah dibuat. Hasil klasifikasi akan divisualisasi dalam bentuk *confusion matrix* yang ditunjukkan pada Gambar 3.

Setelah pengujian model selesai dilakukan, langkah selanjutnya yaitu evaluasi model. Evaluasi model bertujuan untuk menghasilkan *confusion matrix* dengan ukuran 3 x 3. Gambar 3 memperlihatkan *confusion matrix* berupa matrik dengan ukuran 3 x 3 yang mewakili setiap kelas klasifikasi positif, netral, dan negatif. Dari *confusion matrix* dapat dijelaskan bahwa model mengklasifikasikan secara benar sebanyak 104 data sebagai negatif, 59 data sebagai positif, dan 98 sebagai data netral.

Setelah diketahui *confusion matrix* dari model yang dibuat, selanjutnya dihitung akurasi dari model klasifikasi *Naive Bayes* yang telah dibuat. Akurasi menggambarkan seberapa besar tingkat akurat model yang telah dibuat dapat mengklasifikasikan dengan benar. Akurasi didapatkan dari perhitungan rasio prediksi benar terhadap keseluruhan data. Dengan mengetahui besarnya nilai akurasi pada kinerja keseluruhan sistem, dapat dinyatakan tingkat kemampuan sistem dalam mencari ketepatan antara informasi yang diinginkan penguna dengan jawaban yang diberikan oleh sistem [11]. Tingkat keberhasilan sistem dalam menemukan sebuah informasi sebesar 62,4%.

Selanjutnya, untuk melihat performa klasifikasi dari setiap kelas dapat diketahui melalui nilai *precision*, *recall*, dan *F1-score* pada setiap kelas klasifikasi. Hasil nilai *precision*, *recall*, dan *F1-score* memiliki nilai sebesar 0-1. Semakin tinggi nilainya, maka semakin baik hasil model yang dibuat. Nilai *precision* untuk kelas positif sebesar 52%, untuk kelas netral sebesar 67%, dan untuk kelas negatif sebesar 65%. Tingkat keberhasilan sistem dalam *recall*, atau menemukan kembali sebuah informasi, untuk kelas positif sebesar 49%, untuk kelas netral sebesar 71%, dan kelas negatif sebesar 65%. Terakhir, didapatkan rata rata nilai *precision* sebesar 62%, nilai *recall* sebesar 62%, dan nilai *F1-score* sebesar 62%.

IV. SIMPULAN

Pada penelitian ini terbukti dapat dilakukannya klasifikasi menggunakan algoritme *Naive Bayes classifier* dan menghasilkan nilai akurasi sebesar 62,4%. Hasil analisis sentiment dari Twitter dengan kata kunci pariwisata pada penelitian ini memiliki nilai rata-rata pada *precision* sebesar 62%, nilai *recall* sebesar 62%, dan nilai *F1-score* sebesar 62%. Hasil dari penelitian ini masih belum maksimal. Hal ini disebabkan beberapa kemungkinan, di antaranya pada tahap *preprocessing* data tidak dilakukan secara maksimal sehingga ragam kata yang tidak termuat dalam kamus sentimen kata tersebut tidak dikalkulasi dengan baik. Hal tersebut dalam proses implementasi mendapatkan hasil yang tergolong buruk.

DAFTAR REFERENSI

[1] Direktorat Jenderal Pencegahan dan Pengendalian Penyakit (P2P), "Pedoman Pencegahan dan Pengendalian Coronavirus Disease (Covid-19)," Jakarta: Kementerian Kesehatan RI. 2020. hlm.0–115.

[2] A. Rausanfita, P. P. Adikara, dan S. Adinugroho. "Analisis sentimen Twitter menggunakan *Ensemble Feature* dan metode *Extreme Learning Machine* (ELM) (studi kasus: Samsung Indonesia)", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2, no. 12, hlm. 6409-6417.

[3] A. F. Watratan, Arwini Puspita. B, dan D. Moeis, "Implementasia algoritme Naive Bayes untuk memprediksi tingkat penyebaran Covid-19 di Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, hlm. 7–14, 2020. DOI: 10.52158/jacost.v1i1.9.

[4] T. R. Patil dan S. S. Sherekar, "Performance analysis of ANN and Naive Bayes classification algorithm for data classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 2, hlm. 88–91, 2013. DOI: 10.18201/ijisae.2019252786.

[5] F. Zamachsari, G. V. Saragih, Susafa'ati, dan W. Gata, "Analisis sentimen pemindahan ibu kota negara dengan *Feature Selection*," *J. Resti (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 3, hlm. 504–512, 2020.

[6] A. M. Pravina, I. Cholissodin, dan P. P. Adikara, "Analisis sentimen tentang opini maskapai penerbangan pada dokumen Twitter menggunakan algoritme *Support Vector Machine* (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, hlm. 2789–2797, 2019.

[7] D. Normawati dan S. A. Prayogi, "Implementasi *Naïve Bayes classifier* dan *confusion matrix* pada analisis sentimen berbasis teks pada Twitter," *J. Sains Komput. Inform.*, vol. 5, no. 2, hlm. 697–711, 2021.

[8] F. A. Muttaqin dan A. M. Bachtiar, "Implementasi *text mining* pada aplikasi pengawasan penggunaan internet anak 'Dodo Kids Browser'," *J. Ilm. Komput. dan Inform.*, hlm. 1–8, 2016.

[9] A. Alfiani Mahardhika, R. Saptono, dan R. Anggrainingsih, "Sistem klasifikasi *feedback* pelanggan dan rekomendasi solusi atas keluhan di UPT Puskom UNS dengan algoritme *Naive Bayes classifier* dan *cosine similiarity*," *J. Teknol. Inf. ITSmart*, vol. 4, no. 1, hlm. 36, 2016, DOI: 10.20961/its.v4i1.1806.

[10] H. Yun, "Prediction model of algal blooms using logistic regression and confusion matrix," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 3, hlm. 2407–2413, 2021. DOI: 10.11591/ijece.v11i3.pp2407-2413.

[11] M. Z. Al-Taie, S. Kadry, dan J. P. Lucas, "Online data preprocessing: A case study approach," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 4, hlm. 2620–2626, 2019. DOI: 10.11591/ijece.v9i4.pp2620-2626.

Indra Weni, menerima gelar Sarjana dari Universitas Putra Indonesia YPTK Padang Jurusan Ilmu Komputer bidang Sistem Informasi pada tahun 1989 dan gelar Magister Ilmu Komputer dari Universitas Putra Indonesia YPTK Padang Jurusan Ilmu Komputer bidang Ilmu Komputer pada tahun 1993. Saat ini aktif sebagai dosen tetap di Jurusan Teknik Elektro dan Informatika.

Daniel Arsa, menerima gelar Sarjana dari Universitas Sriwijaya Jurusan Ilmu Komputer bidang Teknik Informatika pada tahun 2010 dan gelar Magister Komputer dari Stikom Dinamika Bangsa Jurusan Sistem Informasi bidang Sistem Informasi pada tahun 2013. Saat ini aktif sebagai dosen tetap di Jurusan Teknik Elektro dan Informatika.

Agil Fahreza, lahir di Jambi, tahun 1999. Saat ini sedang menempuh Pendidikan Sarjana di Universitas Jambi. Minat penelitian pada bidang *Big Data* dan *Machine Learning*.