

ANALISIS SENTIMEN MASYARAKAT TERHADAP DATA *TWEET* TRAVELOKA SELAMA *RAPID TEST* ANTIGEN MENGGUNAKAN ALGORITMA *NAÏVE BAYES*

Fenny Novianti*¹⁾, Kiky Rizky Nova Wardani²⁾

1. Universitas Bina Darma, Indonesia
2. Universitas Bina Darma, Indonesia

Article Info

Kata Kunci: Traveloka; Twitter; Pandemi; Rapid Test; Naïve Bayes

Keywords: *Traveloka; Twitter; Pandemic; Rapid Test; Naïve Bayes*

Article history:

Received 15 May 2023

Revised 29 May 2023

Accepted 12 June 2023

Available online 1 September 2023

DOI :

<https://doi.org/10.29100/jupi.v8i3.3973>

* Corresponding author.

Corresponding Author

E-mail address:

18141009P@student.binadarma.ac.id

ABSTRAK

Traveloka melalui akun twitter selama pandemi, banyak mengeluarkan pendapat berupa komentar-komentar tentang syarat untuk melakukan setiap perjalanan baik di luar maupun di dalam negeri yaitu surat dari hasil *rapid test* antigen maupun PCR dengan dinyatakan negatif Covid-19. Rumusan masalah dari penelitian ini adalah bagaimana cara mendapatkan data analisis sentimen dari media sosial Twitter pada data *tweet* Traveloka selama *rapid test* antigen dengan algoritma *naïve bayes*. Penelitian ini merupakan penelitian kuantitatif dengan metode yang digunakan dalam penelitian ini adalah analisis sentimen masyarakat terhadap data *tweet* Traveloka selama *rapid test* antigen berdasarkan dari opini masyarakat atau netizen pada Twitter. Sementara itu, analisis pada penelitian ini menggunakan *Naïve Bayes*. Dalam melakukan pengukuran pada analisis sentimen didapatkan hasil dari pembobotan *TF-IDF* serta hasil akurasi algoritma *Naïve Bayes* sebesar 75% dengan membagi data *training* dan data *testing* dengan rasio 80:20 serta mendapatkan nilai *recall* sebesar 67%, nilai *precision* sebesar 86% dan nilai *f1-score* sebesar 76%.

ABSTRACT

Traveloka, through its Twitter account, during the pandemic, issued many opinions in the form of comments regarding the requirements for each trip both outside and within the country, namely letters from the results of rapid antigen and PCR tests which stated negative for Covid-19. The problem formulation of this research is how to obtain sentiment analysis data from social media Twitter on Traveloka tweet data during the antigen rapid test with the Naïve Bayes algorithm. This research is a quantitative research with the method used in this study is the analysis of public sentiment towards Traveloka tweet data during the antigen rapid test based on public opinion or netizens on Twitter. Meanwhile, the analysis in this study uses Naïve Bayes. In measuring the sentiment analysis, the results obtained from the TF-IDF weighting and the accuracy of Naïve Bayes algorithm were 75% by dividing the training data and data testing with a ration 80:20 and obtaining a recall value of 67%, a precision value of 86% and value f1-score of 76%.

I. PENDAHULUAN

Di era industri 4.0 yang terjadi dalam berkembangnya teknologi informasi dan komunikasi mengakibatkan masyarakat mudah dalam menerima informasi dengan kemajuan teknologi yang dapat memberikan informasi dan komunikasi yang sangat luas. Berdasarkan data APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) pengguna internet di Indonesia pada tahun 2019 sebanyak 171,2 juta jiwa mengalami peningkatan 25,5% sedangkan di tahun 2020 sebesar 196,7 juta jiwa[1]. Tahun 2020, Indonesia menjadi tahun yang sangat genting karena seluruh dunia sedang dilanda pandemi Covid-19[2]. *World Health Organization* (WHO) mengumumkan bahwa covid-19 merupakan pandemi global dengan penyebaran virus ini secara tidak langsung melalui kontak fisik[3]. Pandemi Covid-19 yang terjadi di berbagai negara dan termasuk Indonesia merupakan pandemi yang berlangsung pada tahun 2019, dan disebabkan oleh sindrom akut coronavirus pernapasan. Berbagai banyak hal yang telah didiskusikan masyarakat terkait pandemi Covid-19 melalui media sosial[4]. Media sosial

merupakan *platform* yang dapat digunakan untuk komunikasi antara satu dan lainnya dengan dilakukan secara online sehingga pengguna dapat melakukan interaksi tanpa batasan waktu. Ada tiga makna bersosial terhadap media sosial yaitu pengenalan, komunikasi dan kerjasama[5]. Dengan begitu banyak informasi yang didapatkan mulai dari media *digital* maupun media sosial melalui internet yang diterima oleh masyarakat dengan bebas memberikan pendapat dan disebarakan melalui media sosial pribadi. Di Indonesia sendiri, media sosial yang paling banyak digunakan sampai sekarang adalah *Twitter*[6].

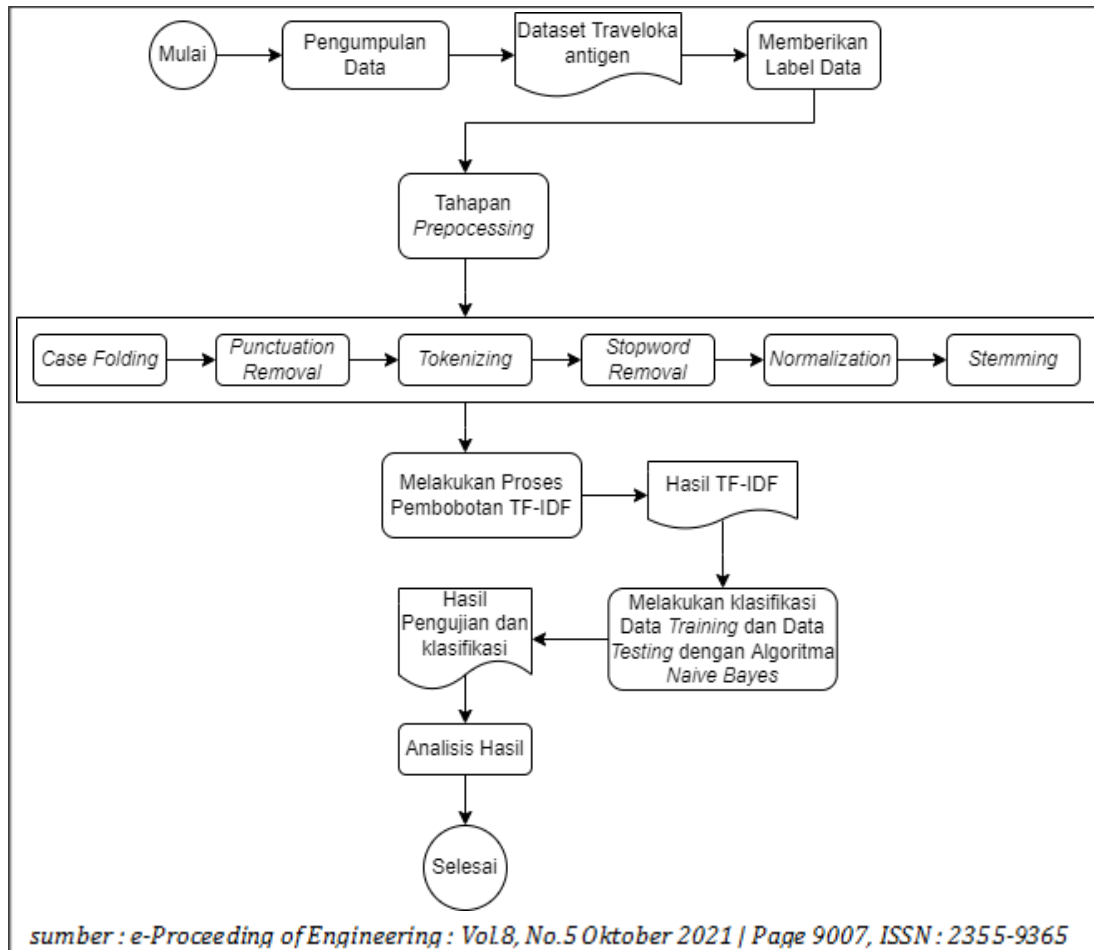
Twitter sendiri merupakan sebuah layanan yang sangat mempermudah pengguna untuk dapat saling berkomunikasi dengan pengguna yang lain dengan cara mengirim sebuah pesan berkapasitas 280 karakter yang dimiliki dan disebut "*tweet*"[7]. Pengguna *Twitter* banyak memanfaatkan media sosial agar dapat menemukan pelaku bisnis, dengan menjadi pengikut (*followers*) dan berinteraksi[8]. Dengan adanya penggunaan *Twitter* pada media sosial, pengguna dengan mudah mengeluarkan pendapat terhadap keluh kesah. Terlebih lagi disaat ini pemerintah di Indonesia banyak mengeluarkan aturan-aturan pembatasan sosial, terutama pada sektor di bidang pariwisata dan akomodasi seperti *Traveloka*[2]. *Traveloka* merupakan perusahaan teknologi terbaik di Asia Tenggara yang banyak menyediakan akses bagi masyarakat, guna menemukan dan memesan berbagai layanan. Pendiri dari *Traveloka* sendiri menyebutkan pada awal terbentuknya *Traveloka* adalah sebuah reservasi untuk melihat perbandingan harga dalam aplikasi tersebut. Selain itu, *Traveloka* merubah situs reservasi pada pemesanan tiket pesawat di tahun 2013. *Traveloka* menambahkan reservasi hotel pada bulan maret 2014, kemudian website resmi pemesanan hotel melalui aplikasi *Traveloka* di bulan Juli 2014. Di tahun 2017, *Traveloka* banyak menyediakan berbagai layanan seperti pemesanan pada tiket pesawat dan kereta api, tiket perjalanan wisata dan pertunjukan seni, internet dari vendor-vendor dengan bekerja sama untuk terus memberikan pelayanan terbaik[9]. *Traveloka* melalui akun *twitter* selama pandemi, banyak mengeluarkan pendapat berupa komentar-komentar tentang syarat untuk melakukan setiap perjalanan baik di luar maupun di dalam negeri yaitu surat dari hasil *rapid test* antigen maupun PCR dengan dinyatakan negatif Covid-19. Dokumen ini menjadi sangat penting mengingat di Indonesia semakin bertambah kasus Covid-19, banyak dari masyarakat yang mengeluh akan aturan tersebut. Dikarenakan mulai dari harga antigen maupun PCR yang berbeda dan sangat mahal.

Oleh karena itu dibutuhkan suatu metode agar dapat mengolah data dalam mengkategorikan dari komentar tersebut di media ke dalam kategori positif dan negatif. Metode ini dapat digunakan agar mengetahui hasil dari data yang telah didapatkan dengan melakukan analisis sentimen. Analisis sentimen adalah bagian dari *text mining*, sehingga dalam melakukannya dibutuhkan pengklasifikasian dengan metode penunjang terhadap banyak komentar atau tanggapan yang berupa sebuah teks. Dalam metode *text mining* berguna untuk menganalisis, evaluasi, penilaian, emosi seseorang serta pendapat sehingga diketahui apakah berkenaan dalam suatu topik, organisasi atau kegiatan tertentu[7]. Kecenderungan penelitian dalam analisis sentimen tersebut berfokus pada pendapat yang menyatakan sentimen positif dan negatif agar dapat digunakan untuk mengetahui pendapat atau opini yang terdapat pada komentar tersebut[10]. Untuk dapat melakukan analisis sentimen masyarakat terhadap data *tweet* *Traveloka* selama *rapid test* antigen menggunakan algoritma *Naïve Bayes*. Berdasarkan dari penelitian sebelumnya, menunjukkan bahwa algoritma *Naïve Bayes* memiliki tingkat akurasi yang lebih tinggi daripada algoritma lain. Pada penelitian Delima Ayu Wulandari, Rd. Rohmat Saedudin, dan Rachmadita Andreswari pada tahun 2020 tentang analisis sentimen media sosial *twitter* terhadap reaksi masyarakat pada ruu cipta kerja menggunakan metode klasifikasi algoritma *Naïve Bayes* memberi kesimpulan bahwa nilai akurasi yang terbaik adalah 92,33% dengan jumlah data 1667 data komentar[7], Pada penelitian dilakukan oleh Fajar Sodik Pamungkas dan Iqbal Kharisudin pada tahun 2021 tentang analisis sentimen dengan SVM, *Naïve Bayes*, dan KNN untuk studi tanggapan masyarakat indonesia terhadap pandemi covid-19 pada media sosial *twitter* menunjukkan bahwa algoritma SVM memiliki kurasi yang lebih tinggi yaitu 90,01% sedangkan pada *naïve bayes* sebesar 79,20% dan *knn* sebesar 62,10% dengan jumlah keseluruhan data adalah 10.000 data[14]. Selanjutnya, penelitian dilakukan oleh Franly Salmon Pattiiha dan Hendy tentang perbandingan metode KNN, *Naïve Bayes*, *Decision Tree* untuk analisis sentimen *tweet* *twitter* terkait opini terhadap PT PAL Indonesia memberikan kesimpulan bahwa metode *Naïve Bayes* merupakan metode dengan tingkat akurasi tertinggi sebesar 84,08% dibanding tingkat akurasi KNN sebesar 83,38% dan *Decision Tree* sebesar 81,09%[20]. Dari penelitian-penelitian ini yang perlu dikembangkan mulai dari jumlah data, tahapan pada proses *preprocessing* data yang ada sedikit berbeda. Jika dari data serta tahapan yang sesuai maka hasil dari menggunakan algoritma *naïve bayes* akan semakin banyak digunakan untuk penelitian selanjutnya.

II. METODE PENELITIAN

A. Tahapan Penelitian

Pada tahapan penelitian ini memiliki tujuan agar hasil yang didapatkan sesuai yang diharapkan. Berikut ini tahapan pada penelitian yang akan dilakukan dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

Penjelasan pada tahapan penelitian dapat dilihat pada gambar 1, yaitu :

1. Langkah pertama dari tahapan penelitian ini adalah mendapatkan data berupa kumpulan *tweets* dari Twitter yang berkaitan dengan penelitian, mengenai tanggapan masyarakat terhadap *rapid test* antigen pada pengguna Traveloka selama melakukan perjalanan dan beberapa *hashtags*. Untuk mendapatkan data dari Twitter harus melakukan pengisian *form* maka akan mendapatkan balasan pesan disetujui atau tidak disetujui melalui gmail dari pihak Twitter. Jika disetujui maka akan mendapatkan API *key* Twitter yaitu API *key*, API *key secret*, *access token* dan *access token secret*. Jika *form* yang telah diisi tidak disetujui, maka akan diberikan lagi pertanyaan oleh pihak Twitter melalui gmail. Pada saat pengambilan data, hal yang pertama dilakukan yaitu membuka *tools jupyter notebook* untuk melakukan proses pada *crawling* data Twitter. Kemudian, memasukkan API *key*, API *key secret*, *access token* dan *access token secret* yang sudah didapat. Dilanjutkan dengan instalasi *packages* untuk dapat berkomunikasi dengan API Twitter. Setelah itu melakukan verifikasi dan autentikasi antara *jupyter notebook* dan Twitter. Ketika prosesnya berhasil, maka dapat melakukan pencarian *keyword* “Traveloka antigen”, “antigen”. Kemudian, setelah pengambilan data selesai maka data akan tersimpan dalam format *csv*.
2. Data yang telah terkumpul selanjutnya dilakukan pelabelan data secara manual agar dapat mengurangi duplikasi pada *tweets* dengan label opini positif dan negatif. Dataset yang dikumpulkan dan diambil termasuk *unsupervised* (tidak berlabel).
3. Melakukan *preprocessing* agar dapat mempersiapkan data untuk diproses dan diawali dengan melakukan proses *case folding*. Dalam hal ini dilakukan agar bentuk huruf seragam dengan mengubah huruf menjadi

huruf kecil/*lowercase* untuk mempermudah proses selanjutnya, pada tahap *punctuation removal* akan melakukan penghapusan pada tanda baca yang terdapat di kalimat pada *tweets*. Setelah tahap *punctuation removal*, pada penelitian dilanjutkan tahap *tokenizing* yaitu tahap pemrosesan pemotongan kata berdasarkan dari tiap tahap pemrosesan pemotongan kata berdasarkan dari tiap kata dalam menyusun pada kalimat tersebut. Selanjutnya, tahap *stopword removal* dilakukan penghapusan kata yang tidak memiliki kontribusi dalam kalimat dan tahap selanjutnya melakukan *stemming*, agar kalimat pada komentar tidak banyak memiliki kata-kata imbuhan. Dalam proses *preprocessing* data yang dihasilkan akan lebih terstruktur dengan baik.

4. Setelah melakukan *preprocessing* langkah selanjutnya melakukan proses pembobotan kata yaitu proses dalam menentukan hubungan kata terhadap dokumen dengan memberikan angka pada setiap kata. Pertama, dengan menghitung nilai TF (*Term Frequency*) yaitu menghitung munculnya kata pada dokumen. Selanjutnya DF (*Document Frequency*) untuk menghitung banyak kata yang sering muncul dan dilakukan perhitungan IDF (*Inverse Document Frequency*). Setelah didapatkan hasil dari TF dan IDF, selanjutnya dilakukan perhitungan TF-IDF serta menghasilkan pembobotan TF-IDF didalam *dataset*[11].
5. Selanjutnya dilakukan pemisahan antara data *training* dan data *testing* yang akan diklasifikasi menggunakan algoritma *Naïve Bayes*. Terakhir proses akan dilanjutkan menghitung nilai akurasi menggunakan *confusion matrix*, dengan melakukan perhitungan *precision*, *recall*, *f1-score* dan *accuracy*.
6. Metode yang digunakan dalam penelitian analisis sentimen masyarakat terhadap data *tweet* Traveloka selama *rapid test* antigen berdasarkan dari opini masyarakat atau netizen pada Twitter menggunakan jenis penelitian kuantitatif. Penelitian kuantitatif yaitu penelitian berupa data yang berbentuk angka dan dianalisis menggunakan data statistika serta angka yang sangat cepat untuk memperoleh hasil dari data penelitian. Dalam hal ini melibatkan banyak tabel dan perhitungan lainnya[12]. Analisis pada penelitian ini menggunakan *Naïve Bayes* dengan menggambarkan suatu konsep algoritma sebagai penentu terhadap kelas dalam sebuah dokumen dan salah satu algoritma dengan klasifikasi melalui pelabelan yang dapat mengelola informasi dari data dengan jumlah besar serta tingkat akurasi dan performa yang tinggi[13]. Dalam memilih metode *Naïve Bayes* untuk penelitian ini dikarenakan analisis sentimen sering menggunakan metode tersebut serta penggunaan yang sederhana dalam menerapkan pada suatu data.

B. Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan algoritma yang dapat digunakan dalam menganalisa sebuah kalimat dengan sekumpulan dokumen serta metode untuk perhitungan setiap kata pada bobot. Rumus RF-IDF merupakan gabungan perhitungan dari TF dan IDF dengan mengalikan nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Tujuan dari penggunaan TF-IDF ini ialah mengekstrak kata menggunakan *machine learning* ke dalam bentuk numerik agar data diolah ke tahap berikutnya[15]. Pada *Term frequency* nilai dari frekuensi ini sering muncul dalam dokumen. Semakin banyak jumlah kemunculan dalam dokumen, maka semakin besar bobot atau nilai yang diberikan sesuai yang lebih besar. Untuk rumus TF didefinisikan sebagai berikut [16] :

$$TF(t, d) = f(t, d) \quad (1)$$

Dalam sebuah dokumen *term* dikoleksikan dapat mendistribusikan secara acak dengan IDF. Karena semakin *term* sering muncul pada dokumen maka nilai IDF akan kecil. Jumlah dokumen dalam *dataset* (N) akan dibagi dengan jumlah dokumen *term* df (t). Rumus IDF sebagai berikut[16] :

$$IDF(t) = \log \left(\frac{N}{df(t)} \right) \quad (2)$$

Dimana :

N : total dokumen

df(t) : banyak jumlah dokumen mengandung kata t

Rumus TF-IDF yaitu [16] :

$$TFIDF(t, d) = TF(t, d) * IDF(t) \quad (3)$$

Keterangan[17] :

- TFIDF (t,d) : banyak kata dari t pada dokumen d
- TF(t,d) : frekuensi muncul kata dari t pada dokumen d
- IDF(t) : nilai idf dalam kata dari t

C. Algoritma Naïve Bayes

Naïve Bayes Classifier adalah metode probabilitas sederhana yang mengklasifikasi dengan pengaplikasian *Teorema Bayes* di dalamnya berasumsikan tidak bergantung (*independent*) tinggi. Dengan mengacu konsep paling dasar dari *Naïve Bayes Classifier* yaitu *Teorema Bayes* yang dikemukakan pertama kali oleh Thomas Bayes[10].

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (4)$$

Keterangan :

- H : Kelas klasifikasi (Positif, Negatif dan Netral)
- X : Data dengan kelas belum diketahui
- P(H|X) : Peluang kelas klasifikasi pada data
- P(X) : Peluang dari X
- P(X|H) : Peluang kejadian data X menjadi sesuai kondisi
- P(H) : Peluang dari H

Keuntungan dalam algoritma ini ialah metode ini menggunakan jumlah data pelatih (*traingin data*) dari data yang kecil untuk ditentukan parameter yang diperlukan proses pengklasifikasi. Karena *variabel independet* yang diasumsikan hanya *variable* dari sebuah kelas yang dibutuhkan bukan dari keseluruhan matriks kovarians[18].

D. Evaluasi Performansi

Evaluasi performansi merupakan tahapan dalam mengukur performansi dari *classifier* dalam melakukan klasifikasi yang digunakan untuk penelitian. Evaluasi ini diperlukan untuk memastikan dari hasil yang telah didapatkan dan diukur menggunakan *Confussion Matrix* dapat dilihat pada tabel 1.

Tabel 1. *Confussion Matrix*

	<i>Actual Positive</i>	<i>Actual Negative</i>
<i>Predicted Positive</i>	TP	FP
<i>Predicted Negative</i>	FN	TN

Keterangan :

- TP = *True Positive*
- FN = *False Negative*
- FP = *False Positive*
- TN = *True Negative*

Dalam melakukan prediksi terhadap menebak label positif dengan benar dan komentar, maka tabel 1 dapat dihitung sebagai *True Positif* (TP). Sebaliknya, jika label negatif maka dapat dihitung sebagai *True Negative* (TN). Jika melakukan prediksi dengan komentar lalu label tersebut positif dan salah maka dihitung sebagai *False Positive* (FP). Sebaliknya, *False Negative* (FN) untuk hasil dengan prediksi salah dengan label negatif[7].

1. Accuracy

Accuracy merupakan jumlah dari proporsi prediksi positif dan negatif yang benar[19]. Adapun rumus perhitungan *accuracy* ialah :

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN} \quad (4)$$

2. Precision

Precision merupakan tingkatan ketepatan dalam rasio benar terhadap jumlah data dengan target positif[19]. Berikut rumus perhitungan *precision* :

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

3. Recall

Recall merupakan jumlah proporsi dokumen berupa teks yang relevan terkendali terhadap semua data yang positif[19]. Adapun rumus perhitungan pada *recall*, yaitu :

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

4. F1-Score

F1-Score merupakan perhitungan dari hasil perkalian nilai antara *precision* dan *recall* dari dua kali pada konstanta, lalu dibagi dari jumlah *precision* dan *recall*[19]. Berikut rumus dari *F1-Score* :

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (7)$$

III. HASIL DAN PEMBAHASAN

A. Analisa Studi Kasus

Penelitian ini melakukan analisis sentimen persoalan terhadap kasus yang ramai dibicarakan dua tahun terakhir ini pada media sosial Twitter khususnya di Indonesia tentang *rapid test* antigen di *tweet* Traveloka. Traveloka merupakan salah satu aplikasi untuk masyarakat melakukan pencarian tiket penerbangan baik darat maupun udara serta mengetahui persyaratan apa saja yang harus dipersiapkan dalam melakukan perjalanan termasuk *rapid test* antigen. Tidak sedikit dari masyarakat banyak mengeluhkan persyaratan tersebut, karena kurangnya informasi dan harga yang berbeda di setiap tempat untuk melakukan *rapid test* antigen. Agar dapat mengetahui komentar-komentar dari masyarakat melalui Twitter terhadap data *tweet* Traveloka tentang *rapid test* antigen lebih banyak komentar setuju atau komentar tidak setuju dengan cara melakukan analisis sentimen.

B. Pengumpulan Data

Dalam pengambilan data yang diperoleh dari komentar-komentar yang ada di media sosial Twitter menggunakan API *key key* yang telah didapat dari Twitter dengan melakukan *crawling* data tersebut menggunakan *python*. Data dari media sosial Twitter yang telah dikumpulkan menggunakan beberapa kata kunci yang berkaitan pada penelitian ini yaitu “Traveloka antigen”. Setelah data terkumpul yang diambil setiap minggu yaitu dari tanggal 01 Januari 2022 sampai 08 September 2022 sebanyak 239 data *tweets*. Kemudian, data yang tersimpan ke dalam format CSV dan akan dilanjutkan ke tahap pelabelan data.

C. Pelabelan Data

Data yang telah didapatkan dari *crawling* data akan diberi label secara manual dan label yang digunakan adalah positif dan negatif. Salah satu contoh data yang telah diberi label dapat dilihat pada tabel 2.

Tabel 2. Data *Tweets* yang Sudah dilabel

No	<i>Tweets</i>	Label
1.	halo ka mau tanya penerbangan dari jakarta menuju kupang dan kupang menuju maumere tidak perlu antigen atau pcr lagi kan yah	Positif
2.	halo ini kok saya ga terverifikasi ya tramsferannya biasanya langsung ini mau antigen nih udah jelas sudah di transfer	Negatif
3.	sekarang tempat yang buat tes antigen amp pcr di bangunan yang baru terakhir tes di sini akhir tahun lalu masih di	Positif
4.	rt bener tdk profesional sudah jam sy kirim email hsl antigen sy berikut no tlp lab nya belum ada jawaban apa	Negatif

5. mohon info ada promo tiket pesawatkah persyaratannya ocr apa antigen ya rencana sby bengkulu Positif

Dari data *tweets* pada tabel 2, salah satu contoh kalimat nomor 1 dengan label positif menunjukkan pada kalimat tersebut terdapat kata-kata yang bermakna positif seperti “halo”, “tanya”, “tidak perlu”. Contoh kalimat nomor 4 menunjukkan kalimat negatif, karena terdapat kata-kata seperti “tdk profesional” dan “belum ada jawaban”. Berdasarkan dari data *tweets* yang telah diberikan label dengan jumlah label positif sebanyak 144 dan label negatif sebanyak 95.

D. Preprocessing Data

Tahapan *preprocessing data* adalah tahapan untuk mempersiapkan data sebelum melakukan pembobotan kata dan klasifikasi dengan metode *Naïve Bayes* serta data yang belum terstruktur menjadi lebih terstruktur. Ada beberapa tahapan pada tahapan ini yaitu *case folding*, *punctuation removal*, *tokenization*, *stopword*, *normalization* dan *stemming*.

1. Case Folding

Tahapan *case folding* adalah mengubah kalimat pada komentar menjadi huruf kecil (*lower case*) agar dalam sebuah kalimat memiliki format kata yang satu dan lainnya dan dapat melanjutkan ke tahap *preprocessing* selanjutnya dengan mudah. Contoh penerapan *case folding* dapat dilihat pada tabel 3.

Tabel 3. Case Folding

No	Sebelum	Sesudah
1.	BESOK TERAKHIR Promo traveloka xperience berlaku hingga 30juni. Klik link dibawah atau masukan MAMSWIDYAREF71 bisaâ€¦ https://t.co/YfXdUqdcYN	besok terakhir promo traveloka xperience berlaku hingga 30juni klik link dibawah atau masukan mamswidyaref71 bisaâ€¦ https://t.co/YfXdUqdcYN
2.	Dahulu ku traveloka cari tiket pesawat dan hotel. Masa kini ku traveloka beli vocer pcr antigen. Sad	dahulu ku traveloka cari tiket pesawat dan hotel masa kini ku traveloka beli vocer per antigen. sad
3.	RT @tepapuru: Gais buat yang di daerah Serpong kalo mau antigen/PCR aku saranin di RS Islam Asshobirin. Utk sekarang belinya via traveloka/â€¦	rt @tepapuru: gais buat yang di daerah serpong kalo mau antigen/pcr aku saranin di rs islam asshobirin. utk sekarang belinya via traveloka/â€¦

2. Punctuation Removal

Tahapan *punctuation removal* merupakan tahapan penghapusan berupa tanda baca, simbol dan angka yang terdapat dalam kalimat. Tujuan dari tahapan ini agar dalam komentar tersebut hanya terdapat sebuah kalimat dengan kata dan tidak terdapat simbol-simbol lainnya. Contoh penerapan *punctuation removal* dapat dilihat pada tabel 4.

Tabel 4. Punctuation Removal

No	Sebelum	Sesudah
1.	besok terakhir promo traveloka xperience berlaku hingga 30juni klik link dibawah atau masukan mamswidyaref71 bisaâ€¦ https://t.co/YfXdUqdcYN	besok terakhir promo traveloka xperience berlaku hingga juni klik link dibawah atau masukan mamswidyaref bisa
2.	dahulu ku traveloka cari tiket pesawat dan hotel masa kini ku traveloka beli vocer pcr antigen. sad	dahulu ku traveloka cari tiket pesawat dan hotel masa kini ku traveloka beli vocer pcr antigen sad
3.	rt @tepapuru: gais buat yang di daerah serpong kalo mau antigen/pcr aku saranin di rs islam asshobirin. utk sekarang belinya via traveloka/â€¦	rt gais buat yang di daerah serpong kalo mau antigenpcr aku saranin di rs islam asshobirin utk sekarang belinya via traveloka

3. Tokenizing

Pada tahapan *tokenizing* yang dilakukan adalah memisahkan *string inout* berdasarkan dari setiap kata yang telah disusun. Contoh penerapan *tokenizing* dapat dilihat pada tabel 5.

Tabel 5. Tokenizing

No	Sebelum	Sesudah
1.	besok terakhir promo traveloka xperience berlaku hingga juni klik link dibawah atau masukan mamswidyaref bisa	['besok', 'terakhir', 'promo', 'traveloka', 'xperience', 'berlaku', 'hingga', 'juni', 'klik', 'link', 'dibawah', 'atau', 'masukan', 'mamswidyaref', 'bisa']
2.	dahulu ku traveloka cari tiket pesawat dan hotel masa kini ku traveloka beli vocer pcr antigen sad	['dahulu', 'ku', 'traveloka', 'cari', 'tiket', 'pesawat', 'dan', 'hotel', 'masa', 'kini', 'ku', 'traveloka', 'beli', 'vocer', 'pcr', 'antigen', 'sad']
3.	rt gais buat yang di daerah serpong kalo mau antigenpcr aku saranin di rs islam asshobirin utk sekarang belinya via traveloka	['rt', 'gais', 'buat', 'yang', 'di', 'daerah', 'serpong', 'kalo', 'mau', 'antigenpcr', 'aku', 'saranin', 'di', 'rs', 'islam', 'asshobirin', 'utk', 'sekarang', 'belinya', 'via', 'traveloka']

4. Stopword Removal

Tahapan *stopword removal* merupakan tahapan untuk menghilangkan kata-kata yang tidak terlalu penting dalam sebuah kalimat. Kata-kata yang tidak terlalu penting antara lain kata ganti, kata-kata yang tidak memiliki arti dan kata ganti. Contoh penerapan *stopword removal* dapat dilihat pada tabel 6.

Tabel 6. *Stopword Removal*

No	Sebelum	Sesudah
1.	['besok', 'terakhir', 'promo', 'traveloka', 'xperience', 'berlaku', 'hingga', 'juni', 'klik', 'link', 'dibawah', 'atau', 'masukan', 'mams-widyaref', 'bisa']	['besok', 'terakhir', 'promo', 'traveloka', 'berlaku', 'hingga', 'juni', 'klik', 'link', 'dibawah', 'atau', 'masukan', 'bisa']
2.	['dahulu', 'ku', 'traveloka', 'cari', 'tiket', 'pesawat', 'dan', 'hotel', 'masa', 'kini', 'ku', 'traveloka', 'beli', 'vocer', 'pcr', 'antigen', 'sad']	['dahulu', 'ku', 'traveloka', 'cari', 'tiket', 'pesawat', 'dan', 'hotel', 'masa', 'kini', 'ku', 'traveloka', 'beli', 'vocer', 'pcr', 'antigen', 'sad']
3.	['rt', 'gais', 'buat', 'yang', 'di', 'daerah', 'serpong', 'kalo', 'mau', 'antigenpcr', 'aku', 'saranin', 'di', 'rs', 'islam', 'asshobirin', 'utk', 'sekarang', 'belinya', 'via', 'traveloka']	['gais', 'buat', 'yang', 'di', 'daerah', 'serpong', 'mau', 'antigenpcr', 'aku', 'saranin', 'di', 'rs', 'islam', 'asshobirin', 'sekarang', 'belinya', 'via', 'traveloka']

5. Normalization

Tahapan normalisasi merupakan tahapan untuk menormalisasikan kata-kata yang mengandung bahasa gaul yang tidak sesuai Kamus Besar Bahasa Indonesia (KBBI) dan kata tidak baku serta singkatan pada media sosial. Contoh penerapan *normalization* dapat dilihat pada tabel 7.

Tabel 7. *Normalization*

No	Sebelum	Sesudah
1.	['besok', 'terakhir', 'promo', 'traveloka', 'berlaku', 'hingga', 'juni', 'klik', 'link', 'dibawah', 'atau', 'masukan', 'bisa']	['besok', 'terakhir', 'promo', 'traveloka', 'berlaku', 'hingga', 'juni', 'klik', 'link', 'dibawah', 'atau', 'masukan', 'bisa']
2.	['dahulu', 'ku', 'traveloka', 'cari', 'tiket', 'pesawat', 'dan', 'hotel', 'masa', 'kini', 'ku', 'traveloka', 'beli', 'vocer', 'pcr', 'antigen', 'sad']	['dahulu', 'aku', 'traveloka', 'cari', 'tiket', 'pesawat', 'dan', 'hotel', 'masa', 'kini', 'aku', 'traveloka', 'beli', 'vocer', 'pcr', 'antigen', 'sedih']
3.	['gais', 'buat', 'yang', 'di', 'daerah', 'serpong', 'mau', 'antigenpcr', 'aku', 'saranin', 'di', 'rs', 'islam', 'asshobirin', 'sekarang', 'belinya', 'via', 'traveloka']	['gais', 'buat', 'yang', 'di', 'daerah', 'serpong', 'mau', 'antigenpcr', 'aku', 'saranin', 'di', 'rumah sakit', 'islam', 'asshobirin', 'sekarang', 'belinya', 'via', 'traveloka']

6. Stemming

Tahapan *stemming* adalah tahapan proses pada pemetaan atau mengembalikan kata ke berbagai bentuk kata-kata ke dalam bentuk kata dasar dengan cara menghilangkan kata akhiran, awalan atau keduanya agar dalam suatu kata dapat kembali ke dalam bentuk kata asal. Contoh penerapan *stemming* dapat dilihat pada tabel 8.

Tabel 8. *Stemming*

No	Sebelum	Sesudah
1.	['besok', 'terakhir', 'promo', 'traveloka', 'berlaku', 'hingga', 'juni', 'klik', 'link', 'dibawah', 'atau', 'masukan', 'bisa']	['besok', 'akhir', 'promo', 'traveloka', 'laku', 'hingga', 'juni', 'klik', 'link', 'bawah', 'atau', 'masuk', 'bisa']
2.	['dahulu', 'aku', 'traveloka', 'cari', 'tiket', 'pesawat', 'dan', 'hotel', 'masa', 'kini', 'aku', 'traveloka', 'beli', 'vocer', 'pcr', 'antigen', 'sedih']	['dahulu', 'aku', 'traveloka', 'cari', 'tiket', 'pesawat', 'dan', 'hotel', 'masa', 'kini', 'aku', 'traveloka', 'beli', 'vocer', 'pcr', 'antigen', 'sedih']
3.	['gais', 'buat', 'yang', 'di', 'daerah', 'serpong', 'mau', 'antigenpcr', 'aku', 'saranin', 'di', 'rumah sakit', 'islam', 'asshobirin', 'sekarang', 'belinya', 'via', 'traveloka']	['gais', 'buat', 'yang', 'di', 'daerah', 'serpong', 'mau', 'antigenpcr', 'aku', 'saranin', 'di', 'rumah sakit', 'islam', 'asshobirin', 'sekarang', 'beli', 'via', 'traveloka']

E. Pembobotan TF-IDF

Pada tahapan pembobotan TF-IDF ini merupakan penentuan terhadap kata-kata (*term*) seberapa jauh keterhubungannya pada kalimat dengan diberikan bobot di setiap kata-kata yang terdapat didalam dokumen pada data *tweet*. Tahapan ini diawali menghitung nilai TF (*Term Frequency*) yang tujuannya menghitung frekuensi seberapa banyak kemunculan kata pada data *tweet*. Selanjtnya, DF (*Document Frequency*) tujuannya menghitung kata yang sering muncul dalam data *tweet* serta dilakukan perhitungan IDF (*Inverse Document Frequent*). Setelah nilai TF dan IDF didapatkan, langkah selanjutnya akan dilakukan perhitungan TF-IDF dan hasilnya berupa pembobotan TF-IDF. Dalam pembobotan TF-IDF dokumen dari data *tweet* menggunakan metode klasifikasi beruka teks.

Kemunculan kata pada data *tweet* dengan frekuensi kata dan jumlah kata secara keseluruhan yang merupakan nilai dari TF-IDF. Contoh kalimat yang telah diproses saat *preprocessing* yang akan dilakukan perhitungan dan pembobotan kata dengan kalimat "halo traveloka antigen ini wajib". Cotoh kalimat pada data *tweet* ke 215 dan selanjutnya akan dihitung TF-IDF. Perhitungan TF-IDF secara manual dapat dilihat pada tabel 9.

Tabel 9. Perhitungan TF (*Term Frequency*)

Kata/Term	Urutan <i>Tweet</i>				
	1	...	215	...	239
Halo	0	...	$\frac{1}{5} = 0,2$...	0
Traveloka	0	...	$\frac{1}{5} = 0,2$...	0
Antigen	0	...	$\frac{1}{5} = 0,2$...	0
Ini	0	...	$\frac{1}{5} = 0,2$...	0
Wajib	0	...	$\frac{1}{5} = 0,2$...	0
Jumlah	0	...	2	...	0

Berdasarkan pada tabel 9 dengan data *tweet* urutan ke 216 berjumlah 2 dan jumlah kata sebanyak 5 kata. Selanjutnya akan dilakukan perhitungan yang sama untuk menentukan nilai DF (*Document Frequency*). Nilai DF ini dengan banyaknya kalimat pada data *tweet* yang muncul.

Tabel 10. Perhitungan DF (*Document Frequency*)

Kata/Term	Urutan <i>Tweet</i>					DF (<i>Document Frequency</i>)
	1	...	215	...	239	
Halo	0	...	2	...	0	29
Traveloka	0	...	2	...	0	96
Antigen	0	...	2	...	0	163
Ini	0	...	2	...	0	20
Wajib	0	...	2	...	0	31

Dapat dilihat pada tabel 10 setelah melakukan perhitungan nilai TF dan DF, kemudian akan dilanjutkan menghitung nilai IDF (*Inverse Document Frequency*) nilai dari log pada hasil D dengan jumlah 239 data *tweet* dan akan dibagi dengan nilai DF (*Document Frequency*) dapat dilihat pada tabel 11 yang menunjukkan hasil setelah dihitung.

Tabel 11. Perhitungan IDF (*Inverse Document Frequency*)

Kata/Term	DF (<i>Document Frequency</i>)	IDF = $\log \left(\frac{D}{DF+1} \right)$
Halo	29	$\log \left(\frac{239}{29+1} \right) = 1,477$
Traveloka	96	$\log \left(\frac{239}{96+1} \right) = 0,391$
Antigen	163	$\log \left(\frac{239}{163+1} \right) = 0,163$
Ini	20	$\log \left(\frac{239}{20+1} \right) = 1,056$
Wajib	31	$\log \left(\frac{239}{31+1} \right) = 0,873$

Setelah hasil diperoleh nilai TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*), maka akan dilakukan perhitungan TF-IDF yang dapat dilihat pada tabel 12. Dari perhitungan ini diperoleh dengan cara mengalikan hasil dari nilai TF dan nilai IDF.

Tabel 12. Perhitungan TF-IDF

Kata/Term	Urutan <i>Tweet</i>					IDF	TF-IDF				
	1	...	215	...	239		1	...	215	...	239
Halo	0	...	0,2	...	0	1,477	0	...	0,2954	...	0
Traveloka	0	...	0,2	...	0	0,391	0	...	0,0782	...	0
Antigen	0	...	0,2	...	0	0,163	0	...	0,0326	...	0
Ini	0	...	0,2	...	0	1,056	0	...	0,2112	...	0
Wajib	0	...	0,2	...	0	0,873	0	...	0,1746	...	0

F. Hasil Klasifikasi Menggunakan Algoritma Naïve Bayes

Setelah dilakukan tahapan *preprocessing* data, maka selanjutnya akan dilakukan klasifikasi dengan metode *Naïve Bayes*. Pada tahapan ini data *tweet* dibagi menjadi dua bagian yaitu data *training* dan data *testing* untuk mengukur seberapa jauh performa menggunakan algoritma *Naïve Bayes* dengan rasio yang telah ditentukan. Pada data *training* digunakan dengan melatih dari hasil pembobotan dan label yang telah ditentukan oleh algoritma *Naïve Bayes*, pada data *testing* untuk mengukur performa *classifier* dari keberhasilan pada saat melakukan klasifikasi dengan prediksi. Lalu, dilakukan klasifikasi algoritma *Naïve Bayes* sampai menghasilkan *dataset* dari hasil klasifikasi. Data *training* akan mempengaruhi tingkat akurasi dihasilkan sedangkan data *testing* akan menguji tingkat akurasi dari model yang telah dibuat data *testing*. Berikut dapat dilihat pada tabel 13 tentang *ratio* pada data *training* dan data *testing*.

Tabel 13. *Ratio Testing*

Rasio	Data Training	Data Testing	Jumlah
60:40	143	96	239
70:30	167	72	
80:20	191	48	

Dari tabel tersebut terdapat rasio perbandingan 80:20 dengan jumlah data *training* 191 data dan data *testing* 48 data.

G. Analisis Sentimen

Dalam melakukan analisis sentimen, data yang telah diproses dengan tahapan *preprocessing* dibagi menjadi data *training* dan data *testing* dengan rasio 80:20. Dari perbandingan 80:20 ini adalah 80% akan digunakan sebagai data *training* dan 20% sebagai data *testing*. Dengan melakukan proses klasifikasi *Naïve Bayes* setelah proses F-IDF dan menggunakan *python* mendapatkan hasil pada label data *testing* sebesar 48 data dengan jumlah data positif 32 data dan jumlah data negatif 16 data.

H. Hasil Evaluasi Performansi

Hasil evaluasi performansi ini dilakukan dengan menggunakan data dari hasil tahapan klasifikasi, dengan metode *confusion matrix* akan didapatkan hasil *accuracy*, *recall*, *precision*, *f1-score* yang dapat dilihat pada tabel 14.

Tabel 14. Hasil Evaluasi Performansi

<i>Confusion Matrix</i>	<i>Predicted</i>		<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
	Negatif	Positif				
<i>Actual</i> Negatif	12	4	75%	60%	75%	67%
Positif	8	24		86%	75%	80%
Avg				77%	75%	76%

Dalam melakukan evaluasi performansi menggunakan *python* untuk mendapatkan hasil yang terdapat pada tabel diatas secara otomatis. *Confusion matrix* diperoleh dengan nilai *True Positive*(TP) sebanyak 24 *tweet*, *True Negative* (TN) sebanyak 12 *tweet*, *False Positive* (FP) sebanyak 4 *tweet*, dan *False Negative* (FN) sebanyak 8 *tweet* sehingga kita dapat lihat juga nilai presisi sebesar 77%, *recall* sebesar 75% dan *f1-score* sebesar 76% sehingga nilai akurasi yang didapatkan sebesar 75%. Adapun perhitungan untuk menentukan *accuracy* dari rumus persamaan 4, *recall* dari rumus persamaan 5, *precision* dari rumus persamaan 6 dan *f1-score* dari rumus persamaan 7 secara manual.

1. *Accuracy*

$$AC = \frac{TP+TN}{TP+TN+FN+FP} = \frac{24+12}{24+12+4+8} = \frac{36}{48} = 0,75$$

2. *Recall*

$$R = \frac{TP}{TP+FN} = \frac{24}{24+12} = \frac{24}{36} = 0,67$$

3. Precision

$$P = \frac{TP}{TP+FP} = \frac{24}{24+4} = \frac{24}{28} = 0,86$$

4. F1-Score

$$F1 - Score = \frac{2xPxR}{P+R} = \frac{2x0,86x0,67}{0,86+0,67} = \frac{1,1524}{1,53} = 0,76$$

I. Wordcloud

Wordcloud dibuat sebagai visualisasi untuk mewakili masing-masing label yang ada. Pada gambar 2 kata pada label positif yang paling sering muncul adalah antigen, traveloka, di, dosis, admin, untuk dan beberapa kata lain yang menunjukkan kata positif pada layanan traveloka antigen dalam sebuah kalimat.



Gambar 2. Wordcloud Positif

Pada gambar 3 label negatif menunjukkan kata dengan kalimat negatif atau kecewa, ada beberapa kata yang sama seperti antigen, traveloka, di pcr dan yang lainnya pada sentimen positif. Pada kata sentimen negatif ini berisi kata negatif selama pelayanan traveloka antigen. Berikut ini merupakan kata yang menampilkan katakata yang sering muncul pada sentimen negatif.



Gambar 3. Wordcloud Negatif

IV. KESIMPULAN

Berdasarkan dari hasil penelitian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut :

1. Berdasarkan dari analisis sentimen yang telah dilakukan menggunakan data tweet dari pengguna Twitter terhadap masyarakat menggunakan rapid test antigen pada Traveloka dengan jumlah 239 data tweet serta

- dilakukan proses *preprocessing* data yang terdiri dari tahapan *case folding*, tahapan *punctuation removal*, tahapan *tokenizing*, tahapan *stopword removal*, tahapan *normalization* dan tahapan *stemming*. Dilanjutkan dengan klasifikasi *Naïve Bayes* dengan membagi data *training* dan data *testing* secara otomatis pada *Python* untuk mendapatkan prediksi dari proses algoritma *Naïve Bayes*. Kemudian melakukan proses perhitungan evaluasi dengan metode *confusion matrix* untuk mengetahui nilai *recall*, *preciion*, *f1-score* dan *accuracy*.
2. Hasil performansi dari analisis sentimen terhadap data *tweet* Traveloka selama *rapid test* antigen menggunakan algoritma *Naïve Bayes* dengan membagi data *training* dan data *testing* dengan rasio 80:20 serta mendapatkan nilai *recall* sebesar 67%, nilai *precision* sebesar 86%, dan nilai *f1-score* sebesar 76%. Dari nilai *recall*, *precision* dan *f1-score* tersebut didapatkan juga nilai akurasi sebesar 75%.
 3. Dampak dari analisis sentimen masyarakat terhadap data *tweet* Traveloka selama *rapid test* antigen menggunakan algoritma *Naïve Bayes* dari komentar pada Twitter adalah positif. Dimana komentar positif terhadap pelayanan Traveloka dalam melayani *rapid test* antigen dengan baik dan sesuai prosedur yang sudah ada, harga sesuai, pelayanan yang baik dan penyediaan pada tempat berbeda dengan komentar negatif dengan berbagai argumen kurangnya pelayanan.

DAFTAR PUSTAKA

- [1] T. APJII, "Survei Pengguna Internet APJII 2019-Q2 2020: Ada Kenaikan 25,5 Juta Pengguna Internet Baru di RI." *Bul. APJII*, vol. 74, 10, 2020.
- [2] K. Ikhsani, C. Catur Widayati, and N. Endah Retno Wuryandari, "ANALISIS PENGARUH PERSEPSI RESIKO, PROMOSI, DAN KEPERCAYAAN MEREK TERHADAP NIAT BELI PASCA COVID-19," *J. Bisnis, Ekon. Manajemen, dan Kewirausahaan*, pp. 81–90, May 2021, doi: 10.52909/jbemk.v1i1.31.
- [3] S. Samsir, A. Ambiyar, U. Verawardina, F. Edi, and R. Watrionthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 1, p. 157, Jan. 2021, doi: 10.30865/mib.v5i1.2580.
- [4] Y. Yunitasari and A. R. Putera, "Analisis Sentimen Masyarakat di Twitter Terkait Pandemi Covid-19," *SMATIKA J.*, vol. 11, no. 01, pp. 22–26, Jun. 2021, doi: 10.32664/smatika.v11i01.520.
- [5] A. Rafiq, "Dampak Media Sosial Terhadap Perubahan Sosial Suatu Masyarakat," *Glob. Komunika*, vol. 1, 1, pp. 18–29, 2020.
- [6] E.S. Negara, "Kajian terhadap tools dan framework social media analytics untuk pemanfaatan data social media dalam penelitian ilmu sosial," *J. Teknol. Technoscientia*, vol. 9, 2, pp. 132–138, 2017.
- [7] D.A. Wulandari, R.R. Saedudin and R. Andreswari, "Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada Ruu Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naive Bayes," *eProceedings Eng.*, vol. 8, 5, 2021.
- [8] D. Setian, I. Iin Seprina, "Analisis Sentimen Masyarakat Terhadap Data Tweet Lazada Indonesia Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier," *Bina Darma Conf. Comput. Sci.*, vol. 1, pp. 998–1004, 2019.
- [9] R.A. Ghifary and others, "Analisis kualitas layanan pada perusahaan e-commerce Traveloka," *J. Chem. Inf. Model.*, vol. 53, 15, pp. 1689–1699, 2018.
- [10] R. Meifitrah, I. Darmawan, and O. Nurul Pratiwi, "Sentiment analysis of tokopedia application review to service product recommender system using neural collaborative filtering for marketplace in Indonesia," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 909, no. 1, p. 012071, Dec. 2020, doi: 10.1088/1757-899X/909/1/012071.
- [11] M. W. A. Putra, Susanti, Erlin, and Herwin, "Analisis Sentimen Dompok Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, Jul. 2020, doi: 10.25299/itjrd.2020.vol5(1).5159.
- [12] A.F.O. Pasaribu and others, "Analisis Pola Menggunakan Metode C4. 5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus: Sman 1 Natar)," *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 1, pp. 80–85, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTISI>
- [13] D. Yoga Pratama and others, "Implementasi Analisis Sentimen Menggunakan Naive Bayes Terhadap Keluhan Sarana dan Prasarana Di Biro Administrasi Umum Universitas Muhammadiyah Surakarta," 2020.
- [14] F.S. Pamungkas, I. Kharisudin "Analisis Sentimen dengan SVM, Naïve Bayes, dan KKN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *In PRISMA, Prosiding Seminar Nasional Matematika.*, vol. 4, pp. 628-634. 2021, doi: 10.32664/smatika.v10(02).7176.
- [15] H.Humairah, I. Darmawan, And O.N. Pratiwi, "Analisis Sentimen Ulasan Produk Toko Online Rubylicious Untuk Peningkatan Layanan Menggunakan Algoritma Naive Bayes," *eProceedings Eng.*, vol. 7, no. 2, pp. 7026–7034, 2020.
- [16] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, p. 422, Apr. 2021, doi: 10.30865/mib.v5i2.2845.
- [17] S.W.U. Vitandy, A.A. Supianto, And F.A. Bachtiar, "Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency-Inverse Document Frequency dan Naive Bayes Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, pp. 6082–6083, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5645>
- [18] Alvina Felicia Watratan, Arwini Puspita. B, and Dikwan Moelis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, Jul. 2020, doi: 10.52158/jacost.v1i1.9.82–6083, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5645>
- [19] T. N. Wijaya, R. Indriati, and M. N. Muzaki, "ANALISIS SENTIMEN OPINI PUBLIK TENTANG UNDANG-UNDANG CIPTA KERJA PADA TWITTER," *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 78–83, Jul. 2021, doi: 10.37905/jjee.v3i2.10885.
- [20] F.S. Pattiha, Hendry, "Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia," *JURIKOM (Jurnal Riset Komputer)*, 9(2),506-514,2022,doi: 10.30865/jurikom.v9i2.4016.