PENERAPAN METODE *MACHINE LEARNING NAÏVE BAYES* PADA TWITTER TERHADAP SENTIMEN PADA PEMBELAJARAN DARING DI INDONESIA

Sri Devi Wahyu Pratama, Febriyanti Panjaitan, M. Izman Herdiansyah, Susan Dian Purnamasari, Nurul Adha Oktarini Saputri
Program Studi Teknik Informatika, Sains Teknologi, Universitas Bina Darma,
Jalan Jenderal Ahmad Yani, Kota Palembang, Sumatera Selatan, Indonesia
srideviwahyuprtama@gmail.com

ABSTRAK

Pandemi COVID-19 merupakan penyakit yang skala penyebarannya terjadi secara global di seluruh dunia,termasuk Indonesia. Banyak bidang yang terkena dampak pandemi ini termasuk pendidikan. Indonesia saat ini sedang menjalankan strategi pembelajaran daring yang menimbulkan banyak opini masyarakat. Tanggapan dari masyarakat terhadap tindakan vaksinasi cukup beragam di media sosial salah satunya media sosial *Twitter*. Tanggapan masyarakat ada yang mendukung dan ada juga yang tidak setuju. Maka dari itu, dalam penelitian ini dilakukan analisis sentimen terhadap *tweets* yang berhubungan dengan pembelajaran daring. Pembelajaran daring sederhananya dapat diartikan sebagai sebuah sistem kegiatan pembelajaran yang dilakukan tanpa melalui tatap muka secara langsung melainkan melalui jaringan internet. Metode penelitian melibatkan pengumpulan data dari twitter dengan metode machine learning. Berdasarkan hasil yang di dapat menggunakan data dari tahun 2020-2022. Dari hasil pengujian didapatkan bahwa Naive Bayes dapat mengklasifikasikan data berupa teks, terutama teks yang berasal dari komentar (tweet), tingkat AUC atau akurasi sebesar 1.00 atau 100%, precision atau keakuratan data sebesar 0.98 atau 98%, *recall* atau keberhasilan model dalam menemukan sebuah informasi sebesar 0.94 atau 94%, dan f-1 Score sebesar 0.96 atau sebesar 96% berdasarkan dengan jumlah data yang ada pada pelatihan sebanyak 248 data komentar. Disimpulkan bahwa menurut uji model, metode Naive Bayes mempunyai nilai akurasi diatas 90% menunjukkan hal ini, bahwa metode Naive Bayes punya kekuatan yang sangat baik dalam klasifikasi analisis sentimen teks.

***Kata kunci**: Analisis sentimen, Machine Learning, Naive Bayes, Pembelajaran Daring*

1. PENDAHULUAN

Pandemi COVID-19 telah mengubah banyak aspek kehidupan, termasuk sistem pendidikan. Salah satu perubahan yang signifikan adalah peralihan dari pembelajaran tatap muka ke pembelajaran daring. Di Indonesia, perubahan ini membawa tantangan dan respons yang beragam dari siswa, orang tua, dan pendidik.

Media sosial, terutama *Twitter*, menjadi salah satu platform utama di mana berbagai sentimen dan opini mengenai pembelajaran daring diekspresikan. Sentimen terhadap pembelajaran daring dapat mencerminkan banyak aspek seperti efektivitas pembelajaran, kesiapan infrastruktur, keterlibatan siswa, dan tantangan teknis yang dihadapi. Oleh karena itu, penting untuk menganalisis sentimen ini agar dapat memberikan masukan yang bermanfaat bagi pembuat kebijakan dan pihak terkait untuk meningkatkan kualitas pembelajaran daring.

Metode Machine Learning, khususnya Naïve Bayes, telah terbukti efektif yang mana metode Naïve Bayes adalah metode probabilistik yang sederhana namun kuat dalam pengklasifikasian teks berdasarkan pola-pola yang ada pada data pelatihan.metode naïve bayes Dengan menggunakan metode ini, kita dapat membagi tweet yang berkaitan dengan pembelajaran online menjadi kelas negatif, positif dan netral.

Memprediksi dengan menggunakan metode machine learning telah dilakukan oleh banyak peneliti di terdahulu, metode naïve bayes merupakan metode probabilitas tertinggi untuk mengklasifikasikan data uji kedalam kategori yang paling sesuai.

Sentimen terhadap keberanian pembelajaran dapat mencerminkan apakah aspek yang dimaksud, keterlibatan siswa, kesiapan infrastruktur, dan tantangan teknis yang dihadapi. Oleh karena itu, penting untuk menganalisis sentimen tersebut guna memberikan informasi yang berguna bagi pembuat kebijakan dan pihak terkait guna meningkatkan standar pembelajaran bahasa yang berani.

Pada kali ini penelitian ditujukan agar kelak bisa mengimplementasikan metode pembelajaran mesin "Naive Bayes" untuk menganalisis sentimen tentang hubungan antara pembelajaran *online* di Indonesia dan cuitan *Twitter*. Hasil dari kesabaran penelitian ini semoga kelak diberikan Gambaran yang nyata serta *detail* terhadap masyarakat dengan bagaimana agar melihat pembelajaran *online* dan memberikan saran yang bermanfaat untuk meningkatkan kualitas pembelajaran *online* di masa mendatang.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu pertama oleh Winanto, A., & Budihartanti, C [1] yang berjudul Comparison of the Accuracy of Sentiment Analysis on the Twitter of the DKI Jakarta Provincial Government during the COVID-19 Vaccine Time. Penelitian ini membahas

Hasil akhir pengujian sampel. Dataset menggunakan metode naive Bayes dengan akurasi 84,80% dengan class recall sebesar 85,01% sentimen positif dan 84,59% sentimen negatif dan pada presisi kelas menunjukkan positif pada 85,21% dan negatif pada 84,38% dalam aplikasi penambangan cepat.

Penelitian selanjutnya dilakukan oleh Yulita w[2], yang berjudul Analisis sentimen terhadap opini masyarakat tentang vaksin covid-19 menggunakan algoritma naïve bayes classifier, Penelitian ini bertujuan untuk menganalisis pendapat tentang vaksinasi COVID-19 di Indonesia. Analisis dilakukan terhadap data 3780 tweet yang berkaitan vaksinasi dengan menggunakan algoritma *Naïve Bayes Classifier*. Berdasarkan analisis, dapat diamati bahwa sebagian besar tweet memiliki sikap positif (60,3 %), sementara jumlah tweet yang netral (34,4 %) melebihi jumlah tweet yang menentang (5,4 %). Nilai akurasi yang dihasilkan sebesar 0,93 (93 %). *Naïve Bayes Classifier* dalam analisa data tweet yang berjumlah 3780 tweet menghasikan sikap Positif 60.3%, Netral 34.4% dan Negatif atau menentang sebesar 5.4%. Dari metode *Naïve Bayes Classifier* menghasilkan tingkat akurasi yang sangat tinggi yaitu sebesar 93%.

Peneliti menggunalkaln beberalpal jurnall penelitialn terdalhulu sebalgali sallalh saltu alcualn penulis dallalm melalkukaln penelitialn. Hal ini ditunjukkan agar dapat memperkaya teori dalam memperkaya penelitian.

2.2. Text Minning

Metode *text* mining merupakan pengembangan dari metode data mining yang dapat diterapkan untuk mengatasi masalah tersebut. Algoritma-algoritma dalam *text* mining dibuat untuk dapat mengenali data yang sifatnya semi terstruktur misalnya data-data, abstrak maupun isi dari dokumen-dokumen[3].

Text mining, atau analisis teks, adalah proses ekstraksi informasi berkualitas tinggi dari data teks yang tidak terstruktur melalui penggunaan pembelajaran mesin, statistik, dan linguistic. Teknik ini digunakan untuk mengubah teks yang tidak terstruktur menjadi format terstruktur guna mengidentifikasi pola, topik, kata kunci, dan atribut lainnya dalam data. *Text* mining merupakan alat berharga bagi organisasi yang berurusan dengan jumlah data teks yang besar, karena memungkinkan untuk mengekstraksi informasi yang sebelumnya tidak diketahui dan berpotensi bermanfaat dari data tersebut[4].

2.3. *Preprocessing*

Preprocessing adalah tahap awal yang mesti dilakukan sebelum menganalisis data. *Preprocessing* data bertujuan untuk merubah data yang tidak terstruktur menjadi terstruktur guna mempermudah proses analisis. Proses preprocessing melibatkan beberapa tahapan seperti cleaning, trnsform cases, tokenizing, stopword, removal, dan stemming[5].

2.4. *Tf-idf*

Algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan algoritma yang digunakan untuk menghitung bobot kata dengan cara mempertimbangkan ada berapa banyak kata muncul (frekuensi kata) dan berapa banyak kata tersebut ditemukan dalam dokumen. Oleh karena itu IDF dapat digunakan utuk mengontrol bobot dari kata. Sehingga ketika kata tersebut selalu ada dalam setiap file atau kalimat maka bias dikaakan bahwa kata tersebut tidak penting atau kata umum [6].

2.5. *ALlgoritmal Nalïve Balyes*

Algoritma *naïve bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat, algoritma ini merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas[7].

Algoritma ini bekerja dengan cara mengklasifikasikan kelas berdasarkan pada probabilitas sederhana dimana dalam hal ini diasumsikan setiap attribut yang ada sifatnya saling terpisah, adapun persamaan rumus metode tersebut adalah seperti persamaan (1) dan (2) berikut[8].

2.6. Evaluasi Kinerja *Classifier / Confusion Matrix*

Confusion Matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Presisi atau adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data sebenarnya. *Recall* atau *sensitivity* adalah proporsi kasus positif yang sebenarnya diprediksi positif secara benar[9].

2.7. Google colaboratory

Google Colab sendiri merupakan bentuk modifikasi *dari Jupyter Notebook* yang disediakan untuk Google, dimana *platform* ini sering dimanfaatkan untuk *Machine Learning* maupun *Deep Learning*[10].

2.8. Pemrograman *python*

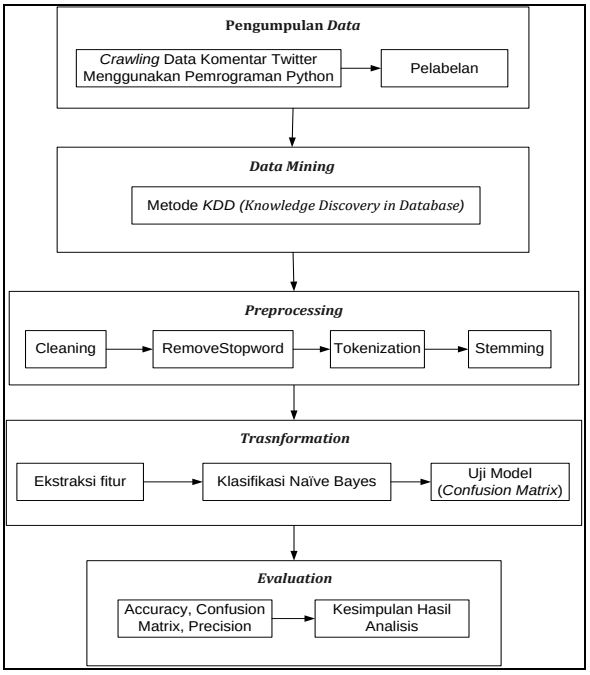
Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Python bisa dibilang bahasa pemrograman dengan tujuan umum yang dikembangkan secara khusus untuk membuat source code mudah dibaca[11].

3. METODE PENELITIAN

Dalam penelitian ini peneliti menggunakan metode KDD (*Knowledge Discovery in Database*). *Knowledge Discovery in Database* (KDD) *is the method of extracting useful information from*

subordinate databases. Terdapat lima tahapan dalam KDD yaitu, data *selection*, *processing*, *tranformation*, data *mining* dan *evaluation*[12].

Dalam penelitian ini penulis membuat suatu rancangan kegiatan yang berisi sebuah rangkaian kegiatan yang akan dilakukan untuk menyelesaikan penelitian, proses yang dilakukan yaitu melalui tahapan:



Gambar 1. Alur Penelitian

Keterangan :

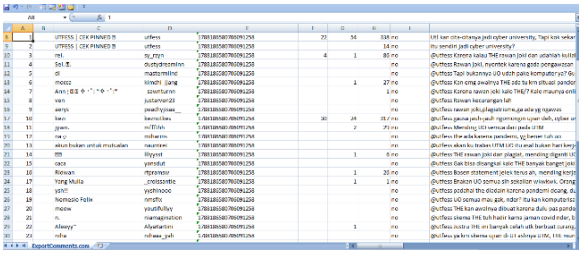
a. Pada bagian ini komentar masyararakat dikumpulkan melalui *twitter*
b. Lalu dikerjakan proses Preprocessing dan Pembobotan, bagian ini telah masuk ke tahapan Naïve Bayes.
c. Pada bagian ini *terjadi* proses mentransformasi data terpilih, sehingga data sesuai dan dikerjakan proses *data mining* terhadap data.
d. *Pada bagian ini dimana Data mining* yang bekerja untuk mencari pola yang menarik dengan metode tertentu.
e. Pada bagian ini *Interpretation / Evaluation* terjadi proses penerjemahan pola yang telah diproses *data mining* , dimana pola informasi yang sebelumnya diperoleh itu hasil dari proses *data mining* maka perlu ditampakkan kedalam bentuk yang dapat dimengerti agar bisa dimengerti secara mudah.

4. HASIL DAN PEMBAHASAN

Beberapa tahapan yang perlu dilakukan Klasifikasi Naïve Bayes dalam menganalisis sentimen di media sosial Twitter, khususnya terkait tanggapan pengguna mengenai pembelajaran daring di Indonesia.

4.1. Pengumpulan Data

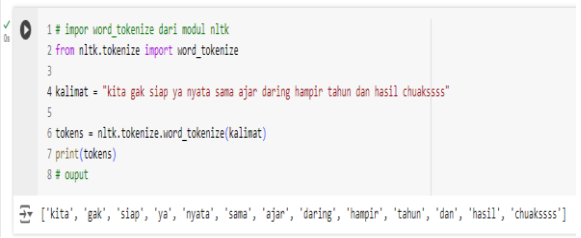
Pada proses ini dilakukan pengumpulan data yang diperlukan menggunakan metode monitoring atau pengamatan, kemudian menggunakan metode studi pustaka, dan selanjutnya metode crawling. Data yang terkumpul terdiri dari 1365 tweet, setelah di processing menjadi 1286 tweet, untuk data yang dipakai berupa *teks post user twitter* diambil dari situs *Twitter.com* dari tahun 2020-2022.



Gambar 2. Excel Hasil Crawling

4.2. Preprocessing

Pada proses ini berguna untuk menghapus kata penghubung dari suatu kalimat seperti yang, dan, tetapi, dan sebagainmya. Dalam proses pengahpusan stopword ini terlebih dahulu dilakukan mendefinisikan kata-kata yang nantinya akan terhapus ketika proses ini dijalankan.



```
1 # impor word_tokenize dari modul nltk
2 from nltk.tokenize import word_tokenize
3
4 kalimat = "kita gak siap ya nyata sama ajar daring hampir tahun dan hasil chuakssss"
5
6 tokens = nltk.tokenize.word_tokenize(kalimat)
7 print(tokens)
8 # ouput

['kita', 'gak', 'siap', 'ya', 'nyata', 'sama', 'ajar', 'daring', 'hampir', 'tahun', 'dan', 'hasil', 'chuakssss']
```

Gambar 3. Proses Pengahapusan Kata Penghubung

4.3. Analisa data computer

a. Membaca data komentar
Proses pertama yaitu membaca data komentar yang sudah dalam bentuk excel, kemudian menghubungkan ke editor *google collaborator* seperti di bawah ini:

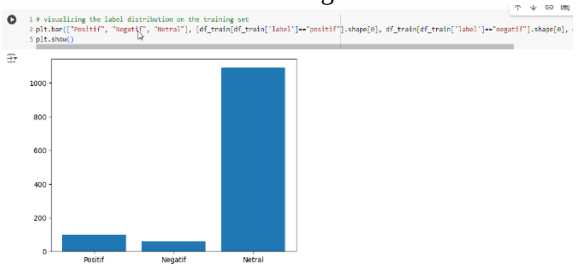
```
[4]  1 from google.colab import drive
     2 drive.mount('/content/drive/')

Drive already mounted at /content/drive/; to attempt to forcibly remount, call drive.

1 # menghubu8ngkan ke dataset training 80% dan tester 20%
2 df_train = pd.read_csv("drive/MyDrive/Training.csv", encoding='ISO-8859-1')
3 df_val = pd.read_csv("drive/MyDrive/Tester.csv", encoding='ISO-8859-1')
4 df_train
```

Gambar 4. Menghubungkan File Data Komentar

b. Menampilkan visualisasi data komentar
Setelah data terbaca selanjutnya menampilkan hasil visualisai data. Lihat gambar 12 di bawah ini.



Gambar 5. Visuallisalsi Klalsifikalsi Sentimen

Berdasarkan data Gambar 5 diatas di dapatkan bahwa sentimen "Netral" lebih tinggi, kemudian dilanjutkan dengan sentimen "Positif", dan terakhir sentiment "Netral".

c. Menampilkan *Confusion Matrix*
Selanjutnya penelitti menapilkan hasil *confusion matrix* berdasarkan metode naïve bayes sehingga hasil akan tampak seperti gambar 13 di bawah ini.

```
12 print("Hasil Confusion Martrix Metoc

Hasil Confusion Martrix Metode Naive Ba
[[ 12   2   0]
[  0 204   1]
[  0   1  28]]
```

Gambar 6. Visualisasi *Confusion Matrix*

4.4. Analisa Model *Naïve Bayes*

Hasil dari klasifikasi nantinya akan ditampilkan dalam bentuk *confusion matrix*. Tabel yang ditampilkan di dalam *confusion matrix* ini terdiri dari kelas *predicted* dan juga kelas *actual*. Model dari *confusion matrix* ini dapat dilihat pada Tabel 1.

Tabel 1. Model Confusion Matrix

		Predict Class	
		Pos	*Neg*
Actual Class	Pos	True Positif	False Negatif
	Neg	False Positif	True Negatif
	Net	True Netral	False Netral

Untuk mengetahui nilai dari akurasi model diperoleh dari banyak jumlah data yang tepat hasil klarifikasi dibagi dengan total dari data, seperti pada Gambar 7 di bawah ini.

$$\boldsymbol{Akurasi} = \frac{\boldsymbol{pos}}{(\boldsymbol{pos+neg+net})} + \frac{\boldsymbol{neg}}{(\boldsymbol{pos+neg+net})} + \frac{\boldsymbol{net}}{(\boldsymbol{pos+neg+net})}$$

Gambar 7. Rumus Akurasi

Didapatkan *confused matrix* matriks yang setiap kolomnya mewakili nilai dari setiap kelas yaitu kelas positif, *netral* dan kelas negatif. Lihat Gambar 8 di bawah ini.

```
13 print("Hasil Classification Report Algoritma Naive Bayes:")

Hasil Classification Report Algoritma Naive Bayes:
              precision    recall  f1-score   support

     negatif       1.00      0.86      0.92        14
      netral       0.99      1.00      0.99       205
     positif       0.97      0.97      0.97        29

    accuracy                           0.98       248
   macro avg       0.98      0.94      0.96       248
weighted avg       0.98      0.98      0.98       248
```

Gambar 8. Halsil uji model NBC

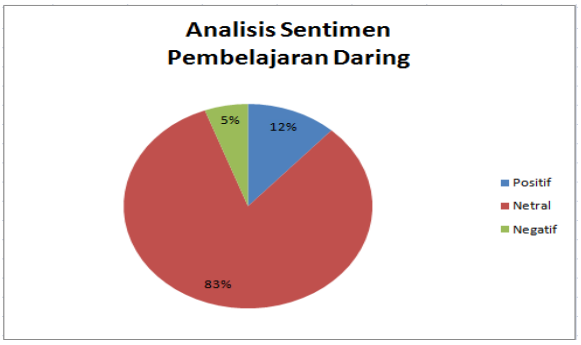
Dengan diketahuinya nilai dari *precision, recall,* dan *f-1 Score* dalam kinerja di keseluruhan sistem, maka dapat mengetahui kemampuan dari sistem untuk mencari ketepatan atau kebenaran dari informasi yang diminta oleh pengguna dengan hasil jawaban yang dikeluarkan oleh sistem dan memberitahu tingkat keberhasilan dari suatu sistem dalam menentukan kembali suatu informasi atau nilai *accuracy* sebesar **98%** pada algoritma *Naïve Bayes***.**

Berdasarkan hasil yang telah didapat dengan tingkat akurasi sebesar 98%, maka hasil sentimen masyarakat mengenai Pembelajaran Daring Di Indonesia, dapat dilihat pada tabel di bawah ini :

Tabel 2. Hasil Analisis Sentimen

Sentimen	**Jumlah**	**Persentase**
Positif	29	12%
Netral	205	83%
Negatif	14	5%
Total	248	100%

Dengan diketahuinya hasil analisis sentiment diatas maka dapat diketahui bahwa sentimen masyarakat yang bersifat Netral lebih tinggi dibandingkan dengan Positif dan Negatif, ini artinya masyarkat tidak terlalu terganggu dengan adanya sistem pembelajaran daring atau *online*.



Gambar 9. Histogram Hasil Analisis Sentimen

5. KESIMPULAN DAN SARAN

Dari hasil pengujian didapatkan bahwa Naive Bayes dapat mengklasifikasikan data berupa teks, terutama teks yang berasal dari komentar (tweet), tingkat AUC atau akurasi sebesar 1.00 atau 100%, precision atau keakuratan data sebesar 0.98 atau 98%, *recall* atau keberhasilan model dalam menemukan sebuah informasi sebesar 0.94 atau 94%, dan f-1 Score sebesar 0.96 atau sebesar 96% berdasarkan dengan jumlah data yang ada pada pelatihan sebanyak 248 data komentar. *D*isarankan data komentar *training* dapatdiperbanyak untuk meningkatkan akurasi klasifikasi sistem. Memperbanyak jumlah kata baku dan tidak baku pada kamus kata baku mengingat gaya penulisan pesan singkat tidak terpaku pada kata baku saja, serta dapat *m*enambahkan analisis sentimen dengan data komentar didalam dataset menggunakan bahasa asing selain bahasa Indonesia atau dapat menambahkan bahasa daerah tertentu karena semua

kalangan dapat mengakses media sosial khususnya dalam penelitian ini media sosialnya adalah *tweet*.

DAFTAR PUSTAKA

[1] A. Winanto and C. Budihartanti, "Comparison of the Accuracy of Sentiment Analysis on the Twitter of the DKI Jakarta Provincial Government during the COVID-19 Vaccine Time," *J. Comput. Sci. Eng.*, vol. 3, no. 1, pp. 14–27, 2022, doi: 10.36596/jcse.v3i1.249.

[2] W. Yulita, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 2, p. 1, 2021, doi: 10.33365/jdmsi.v2i2.1344.

[3] C. Joergensen E Munthe, N. Astuti Hasibuan, and H. Hutabarat, "Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace," *Resolusi Rekayasa Tek. Inform. dan Inf.*, vol. 2, no. 3, pp. 110–115, 2022, doi: 10.30865/resolusi.v2i3.309.

[4] A. Satria, "Implementasi Text Mining untuk Analisis Review pada Aplikasi Crowdfunding LX dan ST Menggunakan Metode Sentiment Analysis," *LANCAH J. Inov. Dan Tren*, vol. 2, no. 1, pp. 124-130file:///C:/Users/User/Pictures/PALEMBANG/JU, 2024.

[5] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.

[6] I. Widaningrum, D. Mustikasari, R. Arifin, S. L. Tsaqila, and D. Fatmawati, "Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen," *Pros. Semin. Nas. Sist. Inf. dan Teknol.*, pp. 145–149, 2022.

[7] N. T. Romadloni, I. Santoso, and S. Budilaksono, "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line," *J. IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 3, no. 2, pp. 1–9, 2019.

[8] O. Somantri and D. Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 191, 2019, doi: 10.26418/jp.v5i2.32661.

[9] P. Mayadewi and E. Rosely, "Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining," *Semin. Nas. Sist. Inf. Indones.*, vol. 2019, no. November, pp. 329–334, 2019.

[10] D. F. Sengkey, F. D. Kambey, S. P. Lengkong, S. R. Joshua, and H. V. F. Kainde, "Pemanfaatan Platform Pemrograman Daring dalam Pembelajaran Probabilitas dan Statistika di Masa Pandemi CoVID-19," *J. Inform.*, vol. 15, no. 4, pp. 217–224, 2020, [Online]. Available: https://ejournal.unsrat.ac.id/index.php/informatika/article/view/31685

[11] D. N. Zuraidah, M. F. Apriyadi, A. R. Fatoni, M. Al Fatih, and Y. Amrozi, "Sistem Informai Penerimaan Siswa Baru Berbasis Web dan Wab," *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 11, no. 2, pp. 1–6, 2021.

[12] N. Purwati, "Analisa Pengaruh Iklan tanpa Label Harga pada media sosial Menggunakan Algoritma Naive Bayes," *Infomatek*, vol. 24, no. 1, pp. 45–50, 2022, doi: 10.23969/infomatek.v24i1.4622.