

NAÏVE BAYES ALGORITHM CLASSIFICATION IN SENTIMENT ANALYSIS COVID-19 WIKIPEDIA

Jessica Margaret Br Sembiring^{*1}, Hendry²

^{1,2}Program Studi Teknik Informatika, Universitas Kristen Satya Wacana, Indonesia
Email: ¹672018238@student.uksw.edu, ²hendry@uksw.edu

(Naskah masuk: 29 April 2022, Revisi : 15 Mei 2022, diterbitkan: 20 Agustus 2022)

Abstract

In recent years during the pandemic Wikipedia created more than 5,200 new pages regarding COVID-19 cases, with an accumulation of more than 400 million pages by mid-June 2020. Wikipedia is one of the most popular websites of our time. In this case Wikipedia always integrates new and fast research. To get an opinion from wikipedia text, sentiment analysis is needed. The analysis was conducted using a classification containing public sentiment regarding the issue of COVID-19 in Indonesia. The classification method used in this study is naive bayes classifier (NBC). Naïve Bayes Classifier is a popular method of solving classification problems. This classification method is often used in sentiment analysis in both precision and data computing. This wikipedia classification is obtained from each label, namely positive, negative and neutral classes. The results of tests conducted in the classification of naive bayes get a high accuracy of 81%.

Keywords: Classification, COVID-19, Naïve Bayes Classifier, Sentimen Analysis, Text Mining, Wikipedia.

KLASIFIKASI ALGORITMA NAÏVE BAYES DALAM ANALISIS SENTIMEN DATA COVID-19 WIKIPEDIA

Abstrak

Beberapa tahun terakhir di masa pandemi Wikipedia membuat lebih dari 5.200 halaman baru mengenai kasus COVID-19, dengan akumulasi lebih dari 400 juta halaman pada pertengahan juni 2020. Keberadaan Wikipedia merupakan salah satu website yang paling populer di masa sekarang. Dalam hal ini Wikipedia selalu mengintegrasikan penelitian baru dan cepat. Untuk mendapatkan suatu opini dari teks wikipedia, dibutuhkan analisis sentimen. Analisis yang dilakukan menggunakan klasifikasi yang berisi sentimen masyarakat mengenai isu COVID-19 di Indonesia. Metode klasifikasi yang digunakan pada penelitian ini adalah Naive Bayes Classifier (NBC). Naïve Bayes Classifier merupakan suatu metode yang populer dalam menyelesaikan suatu masalah klasifikasi. Metode klasifikasi ini sering digunakan pada analisis sentimen baik dalam hal presisi dan komputasi data. Klasifikasi wikipedia ini diperoleh dari tiap label yaitu kelas positif, negatif dan netral. Hasil pengujian yang dilakukan dalam klasifikasi naive bayes ini mendapatkan akurasi yang tinggi sebesar 81%.

Kata kunci: Analisis Sentimen, COVID-19, Data Mining, Klasifikasi, Naïve Bayes Classifier, Wikipedia.

1. PENDAHULUAN

Pada saat ini teknologi di Indonesia semakin berkembang secara pesat, dalam hal mencari suatu informasi tidaklah sulit untuk dilakukan. Kita dapat mencari informasi menggunakan gadget, IOS, laptop dan sebagainya. Sebuah informasi yang kita punya dapat dicari di website yang diakses melalui web browser(Chrome, Firefox dan lainnya) dengan memasukkan alamat url website maupun mencarinya di mesin pencarian(Google, Bing dan lainnya). Wikipedia merupakan salah satu website yang paling populer di masa sekarang, dengan jutaan user yang mempercayakan website tersebut untuk memenuhi berbagai kebutuhan informasi yang

diinginkan setiap hari[3]. Beberapa tahun terakhir di masa pandemi Wikipedia membuat lebih dari 5.200 halaman baru mengenai kasus COVID-19, akumulasi lebih dari 400 juta halaman pada pertengahan juni 2020. Segala sesuatu yang ada di Wikipedia didasarkan pada sumber yang dapat dipercaya, seperti literatur ilmiah. Wikipedia dapat diakses oleh semua orang karena bersifat publik juga menggunakan hasil ilmiah yang representative di masa sekarang ini. Dalam hal ini dapat dilihat bahwa editor Wikipedia selalu mengintegrasikan penelitian baru dengan tanggap dan cepat yang mengutip hampir 2% dari literatur COVID-19 [1].

Python merupakan salah satu bahasa pemrograman yang paling cepat berkembang di dunia. Pada proses ini membutuhkan library NLTK untuk menentukan klasifikasi sentiment melalui Jupyter Notebook.

Analisis Sentimen merupakan suatu proses untuk memahami serta mengelompokkan emosi (positif, negatif, netral) berdasarkan tulisan dengan menggunakan teknik analisis teks [2].

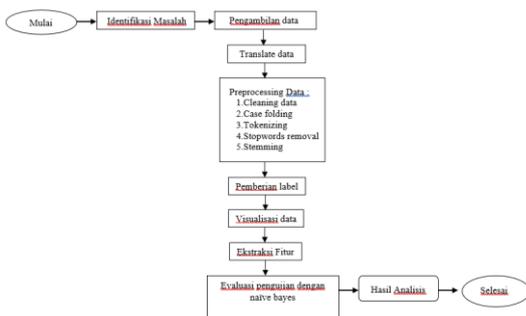
Data yang diambil melalui Wikipedia tentang COVID-19 dan dalam kasus ini berfokus pada web crawling untuk mengetahui isi teks wikipedia. Untuk memodelkan kutipan dari Wikipedia menggunakan metode statistik untuk melihat polaritas kalimat berdasarkan pengelompokan emosi. Algoritma yang digunakan dalam kasus ini Naïve Bayes Classifier akurasi secara keseluruhan dari data yang didapat, presisi serta nilai-nilai yang terdapat didalamnya.

Adapun yang menjadi masalah yaitu menghitung tingkat akurasi dari opini mengenai COVID-19 yang diambil dari sumber wikipedia. Sehingga, dengan menggunakan algoritma naïve bayes analisis permasalahan ini dapat teratasi.

Adapun tujuan yang dilakukan untuk menerapkan metode Naïve Bayes Classifier dalam menganalisis isi teks wikipedia terkait COVID-19 apakah teks tersebut termasuk ke dalam teks dengan label positif, negatif, netral dan mengetahui tingkat akurasi dalam melakukan analisis sentimen pada situs wikipedia terkait COVID-19.

2. METODE PENELITIAN

Penelitian dilakukan berdasarkan data yang diambil melalui wikipedia menggunakan jupyter notebook.



Gambar 1. Tahapan penelitian

Pada Gambar 1 merupakan proses dalam tahapan penelitian. Untuk memudahkan penelitian yang dilakukan.

2.1. Identifikasi masalah

Penelitian yang dilakukan harus berdasarkan permasalahan yang telah dijelaskan sebelumnya yaitu:

1. Menerapkan metode Naïve Bayes Classifier (NBC) dalam menganalisis isi teks wikipedia terkait Covid-19 apakah teks tersebut termasuk

ke dalam teks dengan label positif, negatif dan netral.

2. Untuk mengetahui tingkat akurasi menggunakan metode Naïve Bayes Classifier (NBC) dalam melakukan analisis sentiment situs Wikipedia terkait Covid-19.

2.2. Pengambilan data

Setelah melakukan identifikasi masalah, proses selanjutnya adalah pengambilan data dengan menggunakan jupyter.

```

In [3]: import wikipedia
        from pprint import pprint

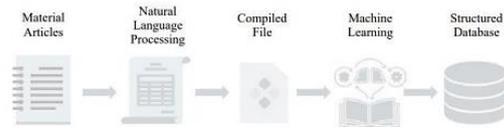
In [4]: wikipedia.search("covid-19 pandemic in indonesia")
    
```

Gambar 2. Penarikan data

Pada gambar 2 merupakan library yang digunakan untuk pengambilan data yaitu import Wikipedia dan melakukan pencarian keyword berdasarkan key "covid-19 pandemic in Indonesia".

2.3. Text mining

Text mining merupakan pemrosesan teks dalam skala yang besar untuk melakukan klasifikasi dokumen, clustering, information extraction, analisis sentimen dan information retrieval yang menghasilkan informasi atau insights [2]. Untuk menghasilkan suatu informasi dari text mining membutuhkan beberapa metode, termasuk NLP. Natural Language Processing merupakan dasar dari text mining. Proses analisis sentimen dapat dilihat pada gambar 3.



Gambar 3. Proses dalam Sentiment Analysis

2.4. Translate data

Setelah melakukan penarikan data, untuk proses selanjutnya yaitu melakukan translate data dengan menggunakan library TextBlob pada gambar 4.

```

from textblob import TextBlob
    
```

Gambar 4. Library translate data

2.5. Naïve bayes classifier

Naïve bayes classifier merupakan suatu algoritma yang populer untuk menyelesaikan suatu masalah dalam pengklasifikasian.

$$P(A|B) = \frac{p(b|a)p(b)}{p(a)} \tag{1}$$

Keterangan :

A : Hipotesa data dari B yang merupakan suatu kelas spesifik.

B : Data kelas yang belum diketahui.

$P(A|B)$: Probabilitas dari suatu hipotesis A berdasarkan kondisi B (posterior probability).

$P(A)$: Probabilitas hipotesis A (prior probabilitas).

$P(B)$ dan

$P(B|A)$: Probabilitas dari kelas sebelumnya berdasarkan kondisi hipotesis [2].

Persamaan (1) merupakan bentuk umum dari teori naïve bayes.

2.6. Wikipedia

Wikipedia adalah salah satu situs web paling populer di dunia dan salah satu ensiklopedia terbesar yang memiliki berbagai macam informasi dari segala sumber [3].

2.7. COVID-19

COVID-19 merupakan virus yang dapat menginfeksi saluran pernapasan, mulai dari pilek hingga penyakit serius seperti sindrom pernapasan akut yang parah (SARSCoV) [4].

2.8. Sentimen analisis

Sentiment analisis merupakan proses mengekstrak dan mengolah suatu data yang tekstual untuk mendapatkan sentiment dari emosi, pendapat, penilaian dalam suatu kalimat yang mengandung beberapa kelas seperti sentimen positif, negatif dan netral. Sentimen analisis berfokus pada opini yang mengandung kalimat positif, negatif dan netral [5]. Dalam hal ini karakteristik sentimen dalam unit teks menggunakan metode Natural Language Processing [6].

2.9. Klasifikasi

Klasifikasi teks dirancang untuk membantu mengatur sejumlah besar informasi banyak sehingga pengguna dapat mengerti [7].

2.10. Preprocessing data

Sebelum dataset diisi ke dalam model, data diproses terlebih dahulu [8]. Dalam tahap preprocessing terdiri dari proses cleaning, case folding, stopwords, tokenizing dan stemming.

2.10.1. Cleaning Data

Tahap pertama hal yang dilakukan yaitu membersihkan data awal dengan mengurangi karakter dan data yang tidak relevan [9].

2.10.2. Case Folding

Setelah melakukan tahap pembersihan dalam suatu data, dilakukan case folding untuk mengubah suatu kalimat menjadi huruf kecil pada dokumen [10].

2.10.3. Tokenizing

Pada tahap ini dilakukan proses pemisahan suatu kalimat menjadi deretan kata, paragraph atau halaman menjadi potongan kata tunggal.

2.10.4. Stopwords removal

Pada tahap ini dilakukan proses filtering untuk memisahkan kata – kata yang bermakna atau penting dan menghilangkan kata imbuhan serta kata sambung.

2.10.5. Stemming

Mengubah kata imbuhan menjadi kata dasar.

2.11. Pemberian label

Setelah melakukan proses preprocessing, proses yang dilakukan selanjutnya pemberian label.

```
In [5]: def polarity_to_label(x):
        if(x >= -1 and x < 0):
            return 'negatif'
        if(x == 0):
            return 'netral'
        if(x > 0 and x <= 1):
            return 'positif'
        df.label = df.label.apply(polarity_to_label)
```

Gambar 4. Pemberian Polarity

Pada gambar 4 merupakan code untuk tiap polarity setiap kalimat yaitu negatif, netral dan positif.

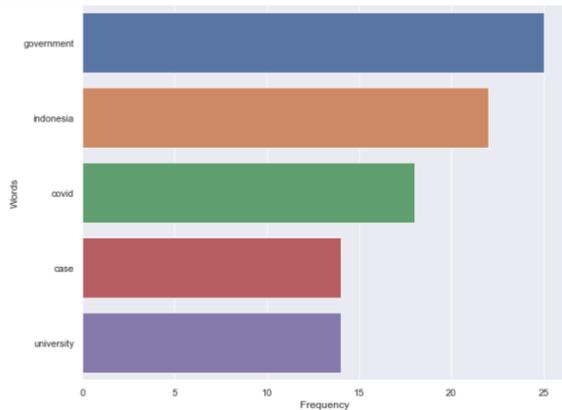
2.12. Evaluasi data

Evaluasi model dilakukan dengan melihat tingkat akurasi metode melalui confusion matrix dan tabel akurasi dan presisi untuk masing-masing model [11].

		Hasil Prediksi		
		A	B	C
Hasil Aktual	A	A	FB1	FC1
	B	FA1	TB	FC2
	C	FA2	FB2	TC

Gambar 5. Confusion matrix

Hasil actual yang didapat pada gambar 5 terdiri dari tiga kelas yaitu A, B dan C. Sedangkan hasil prediksi merupakan hasil yang diperoleh dari model klasifikasi terdiri dari tiga kelas yaitu A, B dan C. Studi kasus tersebut dikelompokkan menjadi sembilan score : TA, FA1, FA2, FB1, TB, FB2, FC1, FC2, dan TC.



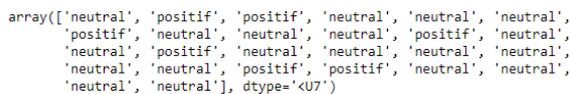
Gambar 10. Grafik batang

3.5. Probabilitas kata dan tf-idf

Berdasarkan data yang berasal dari salah satu situs open source yaitu wikipedia diketahui jumlah data training sebanyak 233 data dan 26 data testing dapat dilihat pada Gambar 11.



Gambar 11. TFIDF features



Gambar 12. Probabilitas kata

Pada gambar 12 adalah probabilitas kata yang didapat melalui label sentimen positif, negatif dan netral.

Tabel 1. Total kalimat berdasarkan sentimen

Total : 259		
Positif	Negatif	Netral
86	55	118

Tabel 2. Bagan pengujian data

Data latih	Data Uji
90%	10%
80%	20%
70%	30%
60%	40%

Tabel 3. Akurasi pengujian data

Pembagian Data	Akurasi
90% : 10%	81%
80% : 20%	67%
70% : 30%	67%
60% : 40%	60%

Dapat dilihat pengujian ini memiliki jumlah kalimat yang berdasarkan sentimen positif, negatif dan netral dapat dilihat pada tabel 1. Pengujian dilakukan menggunakan 4 tahap data uji yang dapat dilihat pada tabel 2 dan hasil akurasi pengujian tiap data dilihat pada tabel 3.

Perbandingan pengujian yang dilakukan dapat dilihat pada gambar 13, gambar 14, gambar 15 dan gambar 16.

Accuracy of the classifier is 0.8076923076923077

classification report is:				
	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	2
neutral	0.84	1.00	0.91	16
positif	0.71	0.62	0.67	8
accuracy			0.81	26
macro avg	0.52	0.54	0.53	26
weighted avg	0.74	0.81	0.77	26

Gambar 13. Hasil pengujian 90:10

Accuracy of the classifier is 0.6730769230769231

classification report is:				
	precision	recall	f1-score	support
negatif	0.50	0.25	0.33	4
neutral	0.72	0.84	0.78	31
positif	0.57	0.47	0.52	17
accuracy			0.67	52
macro avg	0.60	0.52	0.54	52
weighted avg	0.66	0.67	0.66	52

Gambar 14. Hasil pengujian 80:20

Accuracy of the classifier is 0.6666666666666666

classification report is:				
	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	11
neutral	0.67	0.88	0.76	43
positif	0.70	0.58	0.64	24
accuracy			0.67	78
macro avg	0.46	0.49	0.47	78
weighted avg	0.58	0.67	0.61	78

Gambar 15. Hasil pengujian 70:30

Accuracy of the classifier is 0.5961538461538461

classification report is:				
	precision	recall	f1-score	support
negatif	0.50	0.15	0.24	13
neutral	0.60	0.91	0.73	54
positif	0.58	0.30	0.39	37
accuracy			0.60	104
macro avg	0.56	0.45	0.45	104
weighted avg	0.58	0.60	0.55	104

Gambar 16. Hasil pengujian 60:40

Berdasarkan hasil pengujian akurasi untuk data tes yang tertinggi pada pengujian persentase 90%:10% dengan akurasi sebanyak 81%, diikuti pada persentase 80%:20% dengan akurasi sebanyak 67%, persentase 70%:30% dengan akurasi sebanyak 67% dan akurasi terendah pada saat pengujian persentase 60%:40% dengan akurasi yaitu sebanyak 60% dengan rata-rata akurasi dalam pengujian keempat tahap adalah 68,75%.

Precision tertinggi dalam pengujian pada persentase 90:10 yaitu 84%, diikuti pengujian 80:20 sebanyak 72%, pengujian 70:30 sebanyak 70% dan

precision paling rendah pada pengujian 60:40 sebanyak 50% dengan rata-rata precision keempat pengujian adalah 69%.

Recall tertinggi dalam pengujian pada persentase 90:10 sebanyak 100%, diikuti pengujian 60:40 sebanyak 91%, pengujian 70:30 sebanyak 88% dan recall yang terendah pada pengujian 80:20 sebanyak 84% dengan rata-rata recall dalam pengujian keempat tahap adalah 90,75%.

4. DISKUSI

Menurut penelitian (Mesut, 2012) dalam pengklasifikasian berita politik Turki yang menentukan apakah berita tersebut dikategorikan sebagai sentimen positif atau negatif. Penelitian menghasilkan model klasifikasi yang mendapatkan sebesar 72,05% menggunakan Naïve Bayes Classifier, Akurasi Maximum Entropy sebesar 69,44% dan menggunakan SVM dalam bigram sebesar 66,81%[14].

Menurut penelitian yang dilakukan oleh Nanang Ruhjana, 2019 yang mendapatkan klasifikasi berbentuk teks dalam kategori positif dan negatif untuk penerapan lalu lintas ganjil genap, pada penelitian ini mendapatkan akurasi sebesar 86,67%, dengan precision sebesar 71,43% dan recall 80% (Ruhjana, 2019) dan Dwi Suci Ariska Yanti, Indriati, Putra Pandu Adikara, 2019 dalam proses pengujian mendapatkan hasil uji yang memiliki nilai F-Measure tertinggi sebesar 66,1% dan akurasi sebesar 66,5% [15].

Berdasarkan penelitian Sipayung, Maharani dan Zefanya yang membuat analisis sentimen dengan menggunakan naïve bayes classifier. Pada komentar yang dibagi dengan kategori label positif dan negatif. Penelitian ini menghasilkan enam kategori dengan banyak keyword yaitu 55 kata benda, 120 sentimen berdasarkan sentimen positif sebanyak 66 dan 54 sentimen negatif. Hasil dari klasifikasi ini dalam menentukan berdasarkan kategori mendapatkan akurasi tertinggi sebanyak 77,14% dan 75,42% dan penentuan sentimen pada precision sebesar 99,12% dan 72,9%[16].

Pada penelitian terdahulu oleh Adhi Viki Sudiantoro, Eri Zuliarso pada tahun 2018 dengan judul “Analisis sentimen twitter menggunakan text mining dengan algoritma Naïve Bayes Classifier”. Pada penelitian ini membahas opini masyarakat Indonesia terhadap pilkada Jawa Barat, dengan jumlah tweet yang didapat sebanyak 300. Dalam menganalisis data yang didapat sebelumnya dikelompokkan berdasarkan tingkat emosi (positif, negatif, netral). Dalam proses analisis tersebut algoritma Naïve Bayes Classifier mendapatkan hasil pengujian akurasi sebesar 84%. Sehingga disimpulkan bahwa metode yang digunakan berhasil memprediksi sentimen berdasarkan kategori yang bersifat true memiliki akurasi yang tinggi sebanyak 84%. Aplikasi yang digunakan untuk mengklasifikasi data tersebut dengan RStudio[5].

Menurut penelitian yang dilakukan oleh Samsir, Ambiyar, Unung Verawardina, Firman Edi dan Ronal Watrianthos dengan judul “ Analisis sentiment pembelajaran daring pada twitter di masa pandemic COVID-19 menggunakan metode naïve bayes pada tahun 2021. Pada penelitian ini membahas pembelajaran daring dalam masa pandemic COVID-19 . Data yang diambil sebanyak 12,906 yang berkaitan dengan keyword yang dipilih pada minggu pertama November 2020. Penelitian ini diuji pada minggu pertama bulan november 2020 sebanyak 30% berupa data sentimen positif , 69% berupa sentiment negatif dan 1% berupa sentiment netral. Berdasarkan hasil yang didapat bahwa pembelajaran daring belum diimplementasikan secara maksimal di Indonesia pada masa pandemi[2].

Berdasarkan penelitian lainnya adalah “Analisis Sentimen Hatespeech pada Twitter dengan Metode Naïve Bayes Classifier dan Support Vector Machine” dalam penelitian ini dihasilkan akurasi sebesar 66,6% , precision 67,1%, recall sebesar 66,7%, TP rate sebesar 66,7% dan nilai TN sebesar 75,8%[5].

Berdasarkan penelitian yang pernah dilakukan terkait penggunaan algoritma Naïve Bayes Classifier dalam analisis sentimen terhadap ketujuh penelitian ini terdapat persamaan dari jurnal ini yaitu menggunakan algoritma yang sama dan sama – sama mengklasifikasi data sesuai dengan metode algoritma Naïve Bayes Classifier. Perbedaan yang mendasar yaitu terletak pada penggunaan data yang tidak sama dan tools yang digunakan untuk mendapatkan akurasi yang terbaik.

5. KESIMPULAN

Pada penelitian ini data training didapat berjumlah 233 data dan 26 data testing. Pengklasifikasian yang diperoleh berdasarkan tiap sentimen dengan nilai precision sebesar 71%, 0% dan 84%. Recall sebesar 62%, 0% dan 100% berdasarkan sentimen positif, negatif dan netral serta akurasi yang didapat juga tinggi sebesar 81% pada presentase 90: 10 atau bisa dikatakan sebagai 90% data training dan 10% data testing. Pada data wikipedia mengenai covid-19 di Indonesia kata yang paling banyak dicari saat masa covid-19 adalah government dan kata yang jarang dicari yaitu mengenai university. Dalam hal ini pengklasifikasian naïve bayes mampu menghasilkan akurasi tertinggi sebesar 81%. Namun untuk mendapatkan hasil akurasi yang maksimal dapat menggunakan data yang lebih baik atau bersih. Dapat juga melakukan dengan penggabungan jumlah data yang lebih banyak maupun dengan menggunakan metode lain agar dapat memperoleh hasil akurasi yang lebih baik.

DAFTAR PUSTAKA

- [1] G. Colavizza, “COVID-19 research in Wikipedia,” *Quant. Sci. Stud.*, vol. 1, no. 4, pp. 1349–1380, 2020, doi:

- 10.1162/qss_a_00080.
- [2] Prabowo, W. A. and Wiguna, C. 2021. 'Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM', *Jurnal Media Informatika Budidarma*, 5(1), p. 149. doi: 10.30865/mib.v5i1.2604.
- [3] P. Konieczny, "From Adversaries to Allies? The Uneasy Relationship between Experts and the Wikipedia Community," *She Ji*, vol. 7, no. 2, pp. 151–170, 2021, doi: 10.1016/j.sheji.2020.12.003.
- [4] utra, I. and Dana, I, 'Pengaruh Profitabilitas, Leverage, Likuiditas Dan Ukuran Perusahaan Terhadap Return Saham Perusahaan Farmasi Di Bei'. *E-Jurnal Manajemen Universitas Udayana*, 5(11), 2016, p. 249101.
- [5] A. Muzaki and A. Witanti, "Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier," *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [6] A. V. Sudiantoro et al., "Analisis Sentimen Twitter Menggunakan Text Mining Dengan," vol. 10, no. 2, pp. 398–401, 2018.
- [7] H. N. Irmada and Ria Astriratma, "Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 915–922, 2020, doi: 10.29207/resti.v4i5.2313.
- [8] I. Wahyudi, S. Bahri, and P. Handayani, "Aplikasi Pembelajaran Pengenalan Budaya Indonesia," vol. V, no. 1, pp. 135–138, 2019, doi: 10.31294/jtk.v4i2.
- [9] A. K. Fauziyyah, "Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python," *J. Ilm. SINUS*, vol. 18, no. 2, p. 31, 2020, doi: 10.30646/sinus.v18i2.491.
- [10] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier," *Pros. SNST*, vol. 7, pp. 37–42, 2016, [Online]. Available: https://www.publikasiilmiah.unwahas.ac.id/index.php/PROSIDING_SNST_FT/article/view/1506/1589.
- [11] D. Iskandar and Y. K. Suprpto, "Perbandingan Akurasi Klasifikasi Tingkat," *Netw. Eng. Res. Oper.*, vol. 2, no. 1, pp. 37–43, 2015, [Online]. Available: <http://nero.trunojoyo.ac.id/index.php/nero/article/view/42>.
- [12] Ratnawati, F. (2018) 'Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter', *INOVTEK Polbeng - Seri Informatika*, 3(1), p. 50. doi: 10.35314/isi.v3i1.335.
- [13] M. Martin and L. Nilawati, "Recall dan Precision Pada Sistem Temu Kembali Informasi Online Public Access Catalogue (OPAC) di Perpustakaan," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 77–84, 2019, doi: 10.31294/p.v21i1.5064.
- [14] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017, [Online]. Available: <https://t.co/jrvaMsgBdH>.
- [15] H. S. Utama, D. Rosiyadi, D. Aridarma, and B. S. Prakoso, "Sentimen Analisis Kebijakan Ganjil Genap Di Tol Bekasi Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 247–254, 2019, doi: 10.33480/pilar.v15i2.705.
- [16] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.