

SENTIMEN ANALISIS TERHADAP PUAN MAHARANI SEBAGAI KANDIDAT CALON PRESIDEN 2024 BERDASARKAN OPINI TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN ADABOOST

Anggi Riantika Dewi¹, Sri Diana², Moch Alvi Fakhrezi³, Nana Awang⁴, Helmi Ma'arif⁵, Dedi Dwi Saputra⁶

^{1,2,3,4,5,6} *Jurusan Sistem Sistem Informasi Fakultas Teknologi Informasi Universitas Nusa Mandiri
Jl. Kramat Raya No.18, RW.7, Kwitang, Kec, Senen. Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450*

¹11220472@nusamandiri.ac.id

²11220542@nusamandiri.ac.id

³11220356@nusamandiri.ac.id

⁴11220580@nusamandiri.ac.id

⁵11220437@nusamandiri.ac.id

⁶dedi.eis@nusamandiri.ac.id

Abstrak

Indonesia merupakan negara yang menganut sistem demokrasi, hal ini ditandai dengan diselenggarakannya pemilihan umum presiden dan wakil presiden. Publik sangat antusias dengan pencalonan presiden 2024, tidak hanya di dunia nyata, bahkan di dunia maya seperti Twitter pun masyarakat sangat antusias. Jumlah pengguna Twitter bisa digunakan untuk mengetahui sentimen masyarakat terkait pencalonan presiden 2024, untuk menentukan sentimen positif dan negatif sebuah tweet bisa dilakukan secara manual, namun jika dilihat dari jumlah penggunaannya, opini yang dihasilkan juga besar. Oleh karena itu diperlukan suatu mesin yang dapat menganalisis tweet dan mengklasifikasikan tweet menjadi sentimen positif dan negatif secara otomatis. Dalam penelitian ini, penulis melakukan analisis sentimen terhadap tanggapan warganet di media sosial Twitter terhadap ketua DPR Puan Maharani yang akan mencalonkan diri pada pemilihan wakil presiden Indonesia tahun 2024. Penelitian ini menggunakan metode klasifikasi algoritma Naïve Bayes dengan dan AdaBoost untuk mengklasifikasikan perspektif opini publik Twitter terkait pencalonan presiden Puan Maharani. Akurasi yang dihasilkan dari penelitian ini adalah Naïve Bayes 70.50% dan AdaBoost 68.40%.

Kata Kunci: AdaBoost, Naïve Bayes, Pemilihan Umum, Puan Maharani, Sentiment Analysis, Twitter.

I. PENDAHULUAN

Indonesia adalah salah satu negara dengan sistem demokrasi, yang ditandai dengan pemilihan umum untuk presiden dan wakil presiden. Pada tahun 2024 akan menjadi tahun partai Demokrat bagi seluruh rakyat Indonesia karena masa jabatan Presiden dan Wakil presiden Indonesia pada periode 2019 – 2024 berakhir pada pada tahun 2024 [1]. Masyarakat sangat berantusias terhadap pencalonan presiden 2024 tidak hanya di dunia nyata, tetapi didalam dunia maya seperti Twitter pun masyarakat sangat berantusias.

Twitter dapat dilihat sebagai platform yang menciptakan peluang besar bagi kandidat dan pengguna untuk mengeluarkan suara mereka ke dunia. Oleh karena itu, penting untuk

menganalisis dan memahami peran twitter dalam mengukur *opini* masyarakat terhadap pertanyaan kritis siapa kandidat terbaik untuk memimpin indonesia dalam lima tahun kedepan [1]. Salah satu grup *opini* media sosial yang paling banyak adalah Twitter. Didalam twitter setiap ide atau pendapat diungkapkan melalui tweet. Jumlah pengguna dan pengikut karena kemudahan akses, singkatnya twitter mengkomunikasikan tujuan dan sasaran pengguna. Salah satu media sosial paling populer, Twitter mampu mengungkapkan opini objektif tentang topik lain. Twitter adalah media sosial yang populer karena banyak pengguna media sosial memiliki Twitter, dan jumlah total pengguna Twitter adalah 328 juta pengguna [2]. Oleh karena itu peneliti mengambil informasi di Twitter karena pengguna Twitter bisa dimanfaatkan untuk

mengetahui opini publik tentang pencalonan presiden 2024, menentukan opini positif dan negatif sebuah tweet bisa dilakukan secara manual, namun dilihat dari banyaknya pengguna, opini juga terbentuk besar Jadi butuh lebih banyak waktu dan tenaga. Oleh karena itu, diperlukan suatu mesin yang dapat menganalisis tweet dan secara otomatis mengkategorikan banyak tweet menjadi sentimen positif dan negatif, karena menganalisis tweet secara manual sudah tidak efisien lagi. Untuk membuat mesin yang dapat menganalisa tweet dapat.

Text mining adalah varian bidang penambangan data untuk menemukan pola ekstraksi data dalam *database* besar. Pengertian *text mining* hampir sama dengan data *mining*, tetapi letak perbedaan diantara kedua tersebut yang bisa dilihat jelas data yang digunakan. Untuk *text mining*, pengambilan pola informasi didapat dari *database* tekstual dan data bersifat *Natural Laguange Processing* atau dengan kata lain data yang akan diambil dalam bentuk teks yang dapat dibaca oleh manusia, namun untuk data *mining* pengambilan pola menarik dilakukan didata yang terstruktur [3].

Pada penelitian ini menggunakan algoritma klasifikasi yang disebut algoritma Naive Bayes dan Adaboost. Karena algoritma Naive Bayes adalah cara yang cepat dan akurat untuk memproses data dalam jumlah besar. Sedangkan algoritma AdaBoost dipilih karena dapat menghasilkan model prediksi dengan hasil tingkat AUC yang lebih tinggi.

Melihat dari latar belakang diatas, perlu dilakukan perancangan analisis sistem sentimen pada pemilihan calon presiden 2024. Analisis sentimen ini akan digolongkan menjadi 2 kelas yaitu positif dan negatif. Pada penelitian ini penulis melakukan analisis sentimen tweet warganet media sosial terhadap ketua DPR Puan Maharani yang akan mencalonkan diri pada pemilu capres cawapres Indonesia 2024, dan melakukan klasifikasi dari tweet yang berhashtag @puanmaharani_ri. Penelitian ini menggunakan metode klasifikasi algoritma Naive Bayes dan AdaBoost untuk mengklasifikasikan perspektif *opini* masyarakat twitter terkait pencalonan pilpres Puan Maharani.

II. METODE PENELITIAN

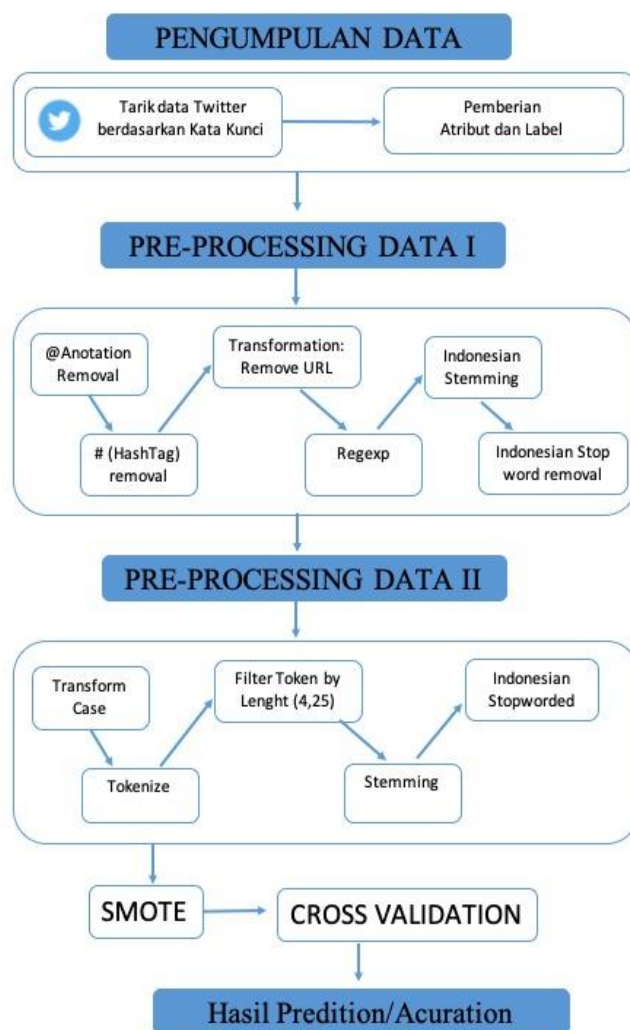
Pada penelitian ini algoritma yang akan digunakan adalah algoritma Naive Bayes & AdaBoost. Algoritma Naive Bayes dan Adaboost dirasa cocok untuk menganalisis sentimen dikarenakan algoritma ini bertujuan sebagai metode klasifikasi dalam kategori positif dan negatif.

Naive Bayes adalah pengklasifikasi probabilistik yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data yang diberikan. [4]. Naive Bayes adalah metode yang memanfaatkan probabilitas dan metode statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes. Naive Bayes merupakan metode klasifikasi yang sangat sederhana dengan mengasumsikan klasifikasi atribut, dengan metode Naive Bayes yang terlebih

dahulu mencari nilai probabilitas dan kemungkinan maksimum untuk setiap atribut pada setiap kelas. [5].

AdaBoost adalah algoritma yang menggunakan bobot berbeda untuk melatih model pada set data yang berbeda. Ini membantu meningkatkan akurasi model dengan menyesuaikan seberapa besar pengaruh yang dimiliki setiap kumpulan data.[6]. *AdaBoost adalah algoritma yang digunakan dalam algoritme boosting untuk meningkatkan performa model pembelajaran mesin. Algoritma mengambil sekumpulan data pelatihan dan menghasilkan model yang lebih efektif dengan menggunakan sejumlah model yang berbeda.*[7].

Berikut diagram proses klasifikasi Sentiment Analysis yang akan dibahas dalam metode penelitian ini:



Gambar 1. Diagram Proses Klasifikasi *Sentiment Analysis*

1. Pengumpulan Data

Hal pertama yang dilakukan pada penelitian ini adalah mengumpulkan data dengan cara *Crawling* data. *Crawling*

$$R = \frac{(TP)}{(TP + FN)}$$

AUC memiliki nilai dengan antara yaitu 0,0 samapi 1,0 dengan tingkat keakuratan klasifikasi sebagai berikut [11]:

- 0,90 – 1,00 = *Excellent Classification*
- 0,80 – 0,90 = *Good Classification*
- 0,70 – 0,80 = *Fair Classification*
- 0,60 – 0,70 = *Poor Classification*
- 0,50 – 0,60 = *Failur*

III. HASIL DAN PEMBAHASAN

1. Pembahasan

Pada penelitian ini menggunakan dataset dari akun twitter @puanmaharani_ri dengan jumlah 752 ulasan yang terdiri dari 358 ulasan negatif dan 394 ulasan positif. Setelah dilakukan labeling data maka akan lanjut dengan menggunakan *Gata Framework* dengan melakukan pengaturan pada teknik seperti gambar dibawah ini:

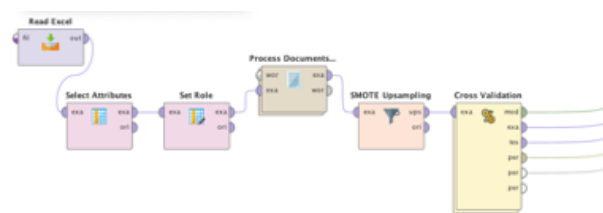


Gambar 3. Tampilan Tools *Gata Framework Text Mining*

Tabel 3. Hasil Preprocessing dengan menggunakan *Gata Framework*

No	Detail	Deskripsi	Detail	Deskripsi	Detail	Deskripsi	Detail	Deskripsi
1	Preprocessing: Remove HTML tags	Remove HTML tags from the text	Preprocessing: Remove URL	Remove URL from the text	Preprocessing: Lowercase	Convert the text to lowercase	Preprocessing: Tokenize	Split the text into tokens
2	Preprocessing: Remove stopwords	Remove stopwords from the text	Preprocessing: Filter tokens	Filter tokens based on length	Preprocessing: Stemming	Apply stemming to the tokens	Preprocessing: Lemmatization	Apply lemmatization to the tokens
3	Preprocessing: Remove special characters	Remove special characters from the text	Preprocessing: Stop words removal	Remove stop words from the text	Preprocessing: Snowball stemmer	Apply Snowball stemmer to the tokens	Preprocessing: Porter stemming	Apply Porter stemming to the tokens
4	Preprocessing: Remove punctuation	Remove punctuation from the text	Preprocessing: Filter tokens by length	Filter tokens based on length	Preprocessing: Lemmatization	Apply lemmatization to the tokens	Preprocessing: Porter stemming	Apply Porter stemming to the tokens
5	Preprocessing: Remove numbers	Remove numbers from the text	Preprocessing: Filter tokens by length	Filter tokens based on length	Preprocessing: Lemmatization	Apply lemmatization to the tokens	Preprocessing: Porter stemming	Apply Porter stemming to the tokens
6	Preprocessing: Remove symbols	Remove symbols from the text	Preprocessing: Filter tokens by length	Filter tokens based on length	Preprocessing: Lemmatization	Apply lemmatization to the tokens	Preprocessing: Porter stemming	Apply Porter stemming to the tokens
7	Preprocessing: Remove digits	Remove digits from the text	Preprocessing: Filter tokens by length	Filter tokens based on length	Preprocessing: Lemmatization	Apply lemmatization to the tokens	Preprocessing: Porter stemming	Apply Porter stemming to the tokens
8	Preprocessing: Remove whitespace	Remove whitespace from the text	Preprocessing: Filter tokens by length	Filter tokens based on length	Preprocessing: Lemmatization	Apply lemmatization to the tokens	Preprocessing: Porter stemming	Apply Porter stemming to the tokens

Gata Framework digunakan karena tweet text yang diambil text berbahasa Indonesia dan hasil keluarannya akan digunakan di RapidMiner untuk pengujian



Gambar 4. Proses Utama RapidMiner

Pada gambar 5 operator "Read Excel" berfungsi untuk membaca file excel hasil dari keluaran twitter yang telah dipreprocessing dengan menggunakan *Gata Framework* sebelumnya, dan dilakukan "Process Documents" untuk melakukan preprocessing tahap kedua agar data benar – benar bisa digunakan untuk ke tahap selanjutnya, dan ditahap selanjutnya menggunakan "Cross Validation" dan ditambahkan didalamnya operator untuk klasifikasi dan evaluasi dari *Sentiment Analysis*.

a.) Process Documents



Gambar 5. Tahap Preprocessing I

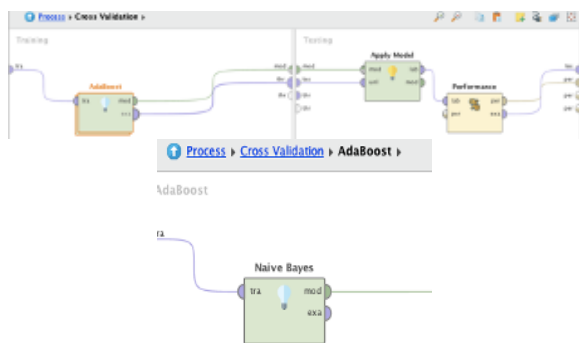
Transform Cases berfungsi untuk menyeragamkan tulisan ke bentuk *lower case* atau *upper case*. Proses *Tokenize* adalah cara memisahkan kata dalam dokumen, memotong setiap kata menjadi bagian yang lebih kecil, dan mengubah huruf menjadi huruf kecil. Proses *Filter Tokens (by Length 4,25)* adalah untuk mengambil kata – kata penting atau menghapus kata dengan jumlah huruf tertentu melalui fungsi *tokens (by length)* dengan parameter min chars 4 dan max chars 25. *Stemming* digunakan untuk pengubahan kata kedalam bentuk *root words* (kata dasar) yang menghilangkan kata imbuhan pada masing – masing kata sehingga menjadi kata dasar dan menghapus kata yang pengejaan nya kurang tepat [12]. *Stopwords* untuk menghapus semua *stopword* atau kata yang kurang berpengaruh dalam penentuan *sentiment* didalam kata.

b.) Cross Validation

Cross validation memiliki dua kolom yaitu training dan testing. Pada kolom training terdapat algoritma klasifikasi yang diterapkan yaitu Naïve Bayes, sedangkan dikolom testing terdapat *Apply Model* untuk menjalankan dan terdapat *performance* sebagai pengukur performa dari model Naïve Bayes seperti pada gambar dibawah ini:



Gambar 6. Validasi pengujian metode Naive Bayes



Gambar 7. Validasi pengujian metode Naive Bayes + AdaBoost

Pada gambar 7 pada pengujian ini, data yang digunakan menggunakan metode Naive Bayes dan Adaboost. Pada pengujian ini menggunakan data yang sudah melalui tahap preprocessing. Data tersebut diambil dari "Read Excel" dan "Process Documents" untuk mengkonversi files menjadi dokumen dan "Cross Validation" yang terdiri dari data Training dan data Testing dan Operator AdaBoost yang didalamnya terdapat operator Naive Bayes.

2. Hasil

Dalam menerapkan metode Naive Bayes dan AdaBoost diperlukan tuning dan percobaan agar dapat mengoptimisasi hasil menjadi lebih baik.

Tabel 4. Confusion Matrix Hasil Klasifikasi

Algoritma	True Positif	True Negatif	False Positif	False Negatif
Naive Bayes	222	279	172	80
Naive Bayes + SMOTE	230	324	164	70
NB + AdaBoost + SMOTE	249	290	145	104

Tabel 5. Hasil Accuracy, Percision, Recall dan AUC

Algoritma	Accuracy	Percision	Recall	AUC
Naive Bayes	70.50%	73.34%	68.79%	0.625
Naive Bayes + SMOTE	73.10%	76.11%	68.03%	0.677
NB + AdaBoost + SMOTE	68.40%	71.18%	63.22%	0.733

Dari hasil yang ditampilkan pada table 5, hasil dari accuracy menggunakan algoritma Naive Bayes yaitu 70.50%, Naive Bayes dan SMOTE yaitu 73,10%, Naive Bayes, SMOTE

dan AdaBoost 68.40%. Hasil dari precision menggunakan algoritma Naive Bayes yaitu 73.34%, Naive bayes dan SMOTE 76.11%, Naive bayes, SMOTE dan AdaBoost 71.18%. Hasil dari Recall menggunakan Algoritma Naive bayes yaitu 68.79%, Naive Bayes dan SMOTE 68.03%, Naive bayes, SMOTE dan AdaBoost 63.22%. Hasil dari AUC menggunakan Algoritma Naive Bayes yaitu 0.625, Naive bayes dan SMOTE 0.677, Naive bayes, SMOTE dan AdaBoost 0.733. Jadi dapat dilihat bahwa pengklasifikasian menggunakan algoritma Naive bayes, SMOTE dan AdaBoost memiliki kinerja baik dengan peningkatan AUC. Yang dimana dengan akurasi 68.40%, persisi 71.18%, Recall 63.22% dan AUC 0.733.

IV. KESIMPULAN

Pada penelitian ini, upaya untuk mengetahui pendapat publik, maka dilakukan pengamatan berupa sentiment positif dan negatif yang terdapat pada media sosial twitter terhadap pencalonan presiden ketua DPR Puan Maharani. Untuk mempresentasikannya maka dilakukan text mining dengan menggunakan metode algoritma Naive Bayes dan AdaBoost untuk mengklasifikasikan label sentimen dari dataset, dari hasil pengujian menunjukkan accuracy menggunakan algoritma Naive Bayes yaitu 70.50%, Naive Bayes, SMOTE dan AdaBoost 68.40%. Hasil dari precision menggunakan algoritma Naive Bayes yaitu 73.34%, Naive bayes, SMOTE dan AdaBoost 71.18%. Hasil dari Recall menggunakan Algoritma Naive bayes yaitu 68.79%, Naive bayes, SMOTE dan AdaBoost 63.22%. Hasil dari AUC menggunakan Algoritma Naive Bayes yaitu 0.625, Naive bayes, SMOTE dan AdaBoost 0.733. dapat disimpulkan bahwa menggunakan algoritma Naive bayes, SMOTE dan AdaBoost memiliki kinerja baik dengan peningkatan AUC. Yang dimana dengan akurasi 68.40%, persisi 71.18%, Recall 63.22% dan AUC 0.733.

V. SARAN

Untuk memberikan rekomendasi yang lebih bermanfaat untuk penelitian lebih lanjut, disarankan agar di masa mendatang kami menggunakan kumpulan data yang lebih kompleks serta meningkatkan teknik preprocessing pada stemming yang menghilangkan kata imbuhan lebih teliti dan lebih baik lagi.

REFERENSI

- [1] D. A. Vonega, A. Fadila, and D. E. Kurniawan, "Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024," vol. 6, no. 2, 2022.
- [2] S. Nurul, J. Fitriyyah, N. Safriadi, E. Esyudha, and P. #3, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," (Jurnal Edukasi dan Penelit. Inform., vol. 5, no. 3, pp. 279–285, 2019, [Online]. Available:

- <http://dev.twitter.com>.
- [3] M. Kuliah, S. K. S. K. Jml, and M. Kuliah, "Program Studi Teknik Industri Fakultas Teknologi Industri - Universitas Ahmad Dahlan," *Dspace.Uii.Ac.Id*, pp. 2019–2020, 2019, [Online]. Available: [https://dspace.uui.ac.id/bitstream/handle/123456789/5744/05.1 bab 1.pdf?sequence=5&isAllowed=y](https://dspace.uui.ac.id/bitstream/handle/123456789/5744/05.1%20bab%201.pdf?sequence=5&isAllowed=y)
- [4] E. Manalu, F. A. Sianturi, and M. R. Manalu, "Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries," *J. Mantik Penusa*, vol. 1, no. 2, pp. 16–21, 2017, [Online]. Available: <https://ezp.lib.unimelb.edu.au/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=ffh&AN=2008-10-Aa4022&site=eds-live&scope=site>
- [5] R. Y. Hayuningtyas, "Penerapan Algoritma Naive Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019, doi: 10.31311/ji.v6i1.4685.
- [6] A. Novandya *et al.*, "Klasifikasi Data Cuaca Dengan Optimasi Decision Tree Berbasis Particle Swarm Optimization Dan Adaboost Klasifikasi Data Cuaca Dengan Optimasi Decision Tree Berbasis Particle Swarm," 2017.
- [7] E. Listiana and M. A. Muslim, "Prosiding Seminar Nasional Teknologi dan Informatika, 2017 : Kudus, 25 Juli 2017," *Pros. SNATIF*, no. 2015, pp. 875–881, 2017.
- [8] F. Ridho, "Rancang bangun aplikasi web crawling untuk mencari harga barang termurah dari berbagai e-marketplace studi kasus: tokopedia, bukalapak, shopee," *Repository.Uinjkt.Ac.Id*, 2020, [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/56138%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/56138/1/FARHAN RIDHO-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/56138%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/56138/1/FARHAN%20RIDHO-FST.pdf)
- [9] S. Hikmawan, A. Pardamean, and S. N. Khasanah, "Sentimen Analisis Publik Terhadap Joko Widodo Terhadap Wabah Covid-19 Menggunakan Metode Machine Learning," vol. 20, no. 2, pp. 167–176, 2020.
- [10] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [11] A. R. Rozzaqi, "Naive Bayes dan Filtering Feature Selection Information Gain untuk Prediksi Ketepatan Kelulusan Mahasiswa," *J. Inform. UPGRIS*, vol. 1, pp. 30–41, 2015.
- [12] N. R. Wardani and A. Erfina, "Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi dokter Menggunakan Algoritma Naive Bayes," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, pp. 12–18, 2021.