



# Penerapan Algoritma Klasifikasi Naïve Bayes Untuk Analisis Sentimen Tentang Pemilu 2024

Yusuf Ramadhan Nasution<sup>1</sup>, Suhardi<sup>2</sup>, Ilham Hafiz Satrio<sup>3</sup>

ramadhannst@uinsu.ac.id<sup>1</sup>, suhardi@uinsu.ac.id<sup>2</sup>, ilham.hafiz@uinsu.ac.id<sup>3</sup>

Universitas Islam Negeri Sumatera Utara

Ilmu Komputer

Jl. Lapangan Golf, Desa Durian Jangak, Kec. Pancur Batu, Kab. Deli Sserdang, Telp. (061) 6615683

## ARTICLE INFO

Article history:

Received : 29 Agustus 2024

Received in revised : 18 November 2024

Accepted : 2 Desember 2024

Available online : 12 Desember 2024

## ABSTRACT

*The news about the proposal of the government of the Republic of Indonesia regarding the postponement of the 2024 elections is certainly an interesting discussion. In this research, sentiment analysis will be carried out on the issue of postponing the election. In this study, a dataset obtained using the crawling technique was obtained in the amount of 1280 tweet data about the postponement of the 2024 election. Data labeling in this study uses lexicon-based techniques with Indonesian dictionaries. By applying this technique, the details of the data in the positive class are 67.7%, namely 157 opinion data, and 32.3% negative, namely 75 opinion data. The sentiment classification system's training and test data yield a 9:1 ratio when the Naïve Bayes Classifier method is applied, and word weighting using TF-IDF yields an accuracy value of 91.67%, precision of 90.91%, recall of 100%, and f1-score of 95.24%.*

*Keywords: elections, sentiment analysis, naïve bayes classifier, lexicon-base, TF-IDF*

## 1. Pendahuluan

Dalam Negara Kesatuan Republik Indonesia (NKRI), pemilu adalah proses yang dilakukan oleh rakyat secara langsung, terbuka, bebas, rahasia, dan jujur untuk memilih anggota Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah, Presiden dan Wakil Presiden, dan Dewan Perwakilan Rakyat Daerah.

Pemilu memungkinkan pemilihan pemimpin dan wakil rakyat yang memenuhi harapan rakyat. Melalui wakil rakyat yang terpilih ini, Indonesia menggantungkan harapan besar dan percaya untuk mewujudkannya. Diharapkan bahwa pemimpin dan wakil yang dipilih melalui pemilu akan mampu mengubah Indonesia menjadi negara yang besar, maju, dan mampu sejajar dengan negara-negara hebat lainnya di dunia. Rakyat Indonesia semakin bersemangat. Mereka berharap pemimpin dan wakil rakyat yang terpilih akan menjalankan amanah dengan cara yang tepat dan dapat diandalkan. Diharapkan wakil dan pemimpin yang menang dalam pemilihan dapat mengurangi disparitas di antara rakyat Indonesia dan berhasil meningkatkan ekonomi, pendidikan, dan kesehatan. Kemudian dapat mensejahterakan rakyat secara adil dan merata [1]

Menurut detikJateng, Menko Polhukam Mahfud MD berkomentar tentang usulan penundaan pemilu yang diajukan oleh Bawaslu pada hari Minggu, 16 Juli 2023. Ia menyatakan bahwa pemilihan 2024 tidak akan ditunda dan akan berlangsung lebih aman dan damai daripada yang sebelumnya. "Menurut saya, pemilu sekarang ini relatif lebih damai, karena tinggal 4 bulan lagi

penentuan calon, dan alhamdulillah kita tenang, tidak ada kekerasan-kekerasan fisik kekerasan politik", katanya. [2].

Sebelumnya, topik penelitian mengenai Pilkada Serentak 2020 sudah pernah dilakukan oleh peneliti terdahulu dengan menggunakan metode Naïve Bayes. Pada tahun 2021 oleh Novrisal, Marthasari, dan Aditya melakukan penelitian dengan judul Sentimen Analisis Tweet Berbahasa Indonesia Pada Pilkada serentak 2020 Menggunakan Metode Naïve Bayes Bebas Particle Swarm Optimization dengan jumlah data sebanyak 1000 Dataset yang menghasilkan tingkat akurasi sebesar 81% [3].

Pada topik penelitian mengenai Pilkada Medan pada sosial media yang dilakukan pada tahun 2021 oleh Anam, Firdaus, Fitri, Lusiana, Agustin, dan Agustin melakukan penelitian dengan judul Analisis Pilkada Medan pada Sosial Media Menggunakan Analisis Sentimen dan Social Network Analysis dengan jumlah data sebanyak 10271 data yang menghasilkan tingkat akurasi sebesar 81.72% [4].

Berdasarkan latar belakang di atas, tujuan dari penelitian ini adalah untuk mengetahui pendapat masyarakat tentang penundaan pemilu 2024 melalui media sosial. Untuk mencapai tujuan ini, penulis menggunakan metode Naïve Bayes untuk menganalisis sentimen data opini masyarakat mengenai subjek pemilu 2024.

## 2. Metode penelitian

Penelitian ini menggunakan pendekatan kualitatif dengan pendekatan deskriptif. Penelitian kualitatif menggunakan pengumpulan data, analisis, dan interpretasi daripada statistik. Berfokus pada multimetode, naturalistik, dan interpretatif (dalam pengumpulan data, paradigma, dan interpretasi), biasanya berhubungan dengan masalah sosial dan manusia yang bersifat interdisipliner. Penelitian kualitatif ini berfokus pada pemahaman tentang masalah kehidupan sosial dalam konteks realitas atau setting alam yang luas, kompleks, dan rinci. Penelitian yang menggunakan paradigma kualitatif adalah jenis penelitian yang menggunakan pendekatan induksi, yang bertujuan untuk menghasilkan konstruksi teori atau hipotesis melalui pengungkapan fakta. [5].

### 2.1 Metode Pengumpulan Data

#### 1. Studi Pustaka

Studi pustaka merupakan langkah penting yang membantu peneliti dalam membangun pemahaman yang mendalam tentang topik penelitian, mengidentifikasi kekurangan pengetahuan yang ada, dan menetapkan dasar teoretis untuk penelitian yang akan dilakukan. Dalam penelitian ini, penulis mengumpulkan data dari literatur, jurnal, skripsi, buku, e-book, dan kajian teori yang relevan [6].

#### 2. Observasi

Dengan menggunakan RapidMiner untuk mengumpulkan data opini dan pandangan masyarakat tentang Penundaan Pemilu 2024 di media sosial Twitter, penulis melakukan observasi secara langsung. Penelitian ini hanya melihat kata kunci Penundaan Pemilu 2024. Data yang digunakan berupa tweet yang diambil dari Twitter, sebuah platform media sosial. Data diambil menggunakan metode crawling Operator yang tersedia di RapidMiner. Sebanyak 1280 tweet diambil dari Twitter. Data ini kemudian dimasukkan untuk diproses [7].

### 2.2 Klasifikasi

Klasifikasi menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dua tugas utama yang terlibat dalam klasifikasi adalah (1) membuat model sebagai prototipe untuk disimpan sebagai memori dan (2) menggunakan model untuk

menemukan, mengklasifikasikan, atau memprediksi objek data lain untuk mengetahui di kelas mana objek data tersebut dalam model yang sudah disimpannya [8].

Salah satu contoh predictive data mining adalah klasifikasi sendiri. Proses pembagian data adalah proses membagi data menjadi dua bagian: data latihan (data instruksi) untuk mengajar sistem dan data uji (data uji) untuk menguji sistem [9].

### 2.3 Algoritma Naïve Bayes

Naive Bayes adalah algoritma pengklasifikasian yang paling umum dan memiliki tingkat keakuratan yang tinggi. Banyak penelitian telah dilakukan menggunakan algoritma ini.

Untuk memprediksi kemungkinan keanggotaan kelas, algoritma Naïve Bayes berguna. Ini adalah statistik penting dalam data mining. Metode ini didasarkan pada kuantif keuntungan dari berbagai keputusan klasifikasi dengan menggunakan probabilitas. Salah satu karakteristik utama dari Naive Bayes Classifier ini adalah keyakinan yang kuat—atau naif—bahwa tidak terpengaruh oleh setiap situasi atau peristiwa [10].

Salah satu keuntungan menggunakan metode Naive Bayes adalah bahwa itu hanya membutuhkan jumlah data pelatihan (training data) yang kecil saja untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Selain itu, metode ini dapat berfungsi jauh lebih baik dalam situasi dunia nyata yang kompleks [11].

Singkatnya, Algoritma Naïve Bayes adalah yang paling efektif untuk memprediksi hal-hal penting.

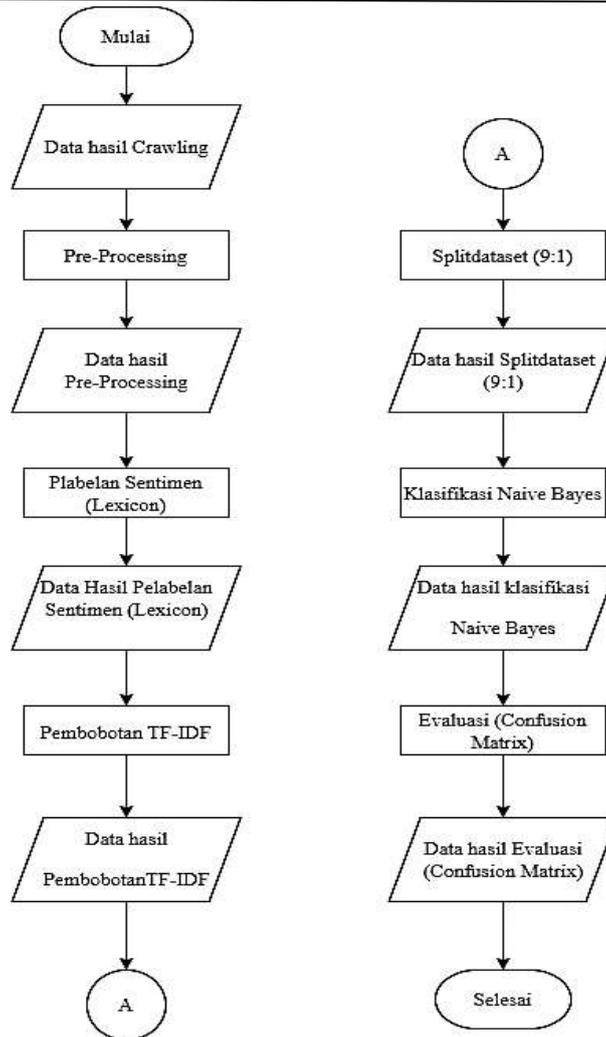
Naive Bayes menggunakan perhitungan probabilitas dan statistik. Formula untuk teorema Bayes dapat ditemukan di bawah ini.

$$P(T|X) = \frac{P(X|T)P(T)}{P(X)}$$

Di mana :

- X = data yang kelasnya belum diketahui
- T = hipotesis dari data X dengan suatu kelas yang spesifik
- P(T|X) = probabilitas bebas bersyarat hipotesis T berdasarkan kondisi bukti X terjadi (probabilitas posterior).
- P(T) = probabilitas awal hipotesis T terjadi tanpa bukti.
- P(X|T) = probabilitas X yang didasarkan pada kondisi hipotesis T
- P(X) = probabilitas kondisional X

Dalam penelitian ini, perancangan dilakukan melalui beberapa tahapan, seperti yang ditunjukkan dalam flowchart berikut :



Gambar 1. Flowchart Klasifikasi Text

### 3. Hasil dan Pembahasan

Hasil dari rancangan tahapan penelitian ini adalah pengumpulan dataset yang mengandung opini atau tanggapan masyarakat terhadap masalah penundaan pemilu 2024. Secara keseluruhan, dataset ini akan diklasifikasikan terdiri dari 1280 tweet dari media sosial *Twitter* yang diperoleh melalui teknik crawling *RapidMiner Studio*.

#### 3.1 Pengumpulan Data

Proses pada *crawling Data* menggunakan *RapidMiner Studio*. Langkah awal yang dilakukan adalah masukkan operator "*Twitter Search*" ke dalam proses *RapidMiner Studio*. Konfigurasi koneksi dengan memasukkan kunci *API* dan token akses yang sudah di dapatkan dari akun *Twitter Developer*. Setelah koneksi dikonfigurasi, tentukan kata kunci atau kriteria pencarian yang ingin digunakan untuk mendapatkan data *Twitter*. Kemudian atur parameter untuk data yang akan diambil, misalnya jumlah *tweet* yang ingin diambil, apakah ingin mengambil *tweet* terbaru, atau *tweet* yang lebih lama, dan sebagainya. Jalankan proses di *RapidMiner* untuk mengambil data *Twitter* berdasarkan kriteria pencarian yang telah ditentukan. Setelah data *Twitter* diambil, data dapat disimpan dalam format *.csv*. Data yang akan digunakan pada proses selanjutnya adalah sebagai berikut :

	Text
0	Pengadilan Tinggi DKI Jakarta 'Batalkan' Putus...
1	RT @tvOneNews: Pengadilan Tinggi DKI Jakarta B...
2	RT @OposisiCerdas: Pengadilan Tinggi DKI Jakar...
3	RT @republikaonline: Partai Prima menghormati ...
4	RT @MediaKomando: Pengadilan Tinggi DKI Jakart...
...	...
1275	PDIP Ungkap ada pihak tempuh berbagai cara dem...
1276	Ahmad Doli Kurnia : Jangan ada lagi gerakan Tu...
1277	Ikuti Jejak Prima, Partai Berkarya Gugat Perda...
1278	PDIP ungkap ada pihak tempuh berbagai cara dem...
1279	Ahmad Doli Kurnia: jangan ada lagi gerakan tun...

1280 rows × 1 columns

Gambar 2. Tampilan *Dataset* yang digunakan

### 3.2 Preprocessing Data

Setelah data hasil *Crawling* diperoleh dan disimpan dalam bentuk *.csv*. Selanjutnya dilakukan *preprocessing data* yaitu serangkaian tindakan yang diambil untuk membersihkan, mengubah, dan menyesuaikan data mentah sebelum digunakan untuk pemodelan atau analisis. *Preprocessing data* bertujuan untuk mempersiapkan data sehingga dapat diolah dengan lebih baik dengan akurasi dan efisiensi. Di antara tahapan *preprocessing data* yang digunakan adalah *case folding* (mengubah semua huruf menjadi huruf kecil), *cleaning* (menghapus nilai yang tidak diperlukan), *removal of duplicates* (menghapus data duplikat), *tokenization* (memisahkan teks menjadi unit kecil yang disebut token), *normalization* (mengubah, menghapus, memperbaiki kata atau istilah singkatan), *stopword removal* (menghapus kata yang tidak memiliki arti), dan *stemming* (menghapus imbuhan pada seluruh kata).

	text
0	adil daerah khusus ibukota jakarta batal putus...
1	adil daerah khusus ibukota jakarta batal putus...
2	adil daerah khusus ibukota jakarta batal putus...
3	partai prima hormat putus batal tunda pemilu
5	adil daerah khusus ibukota jakarta batal putus...
...	...
819	ikut jejak prima partai karya gugat perdata ko...
823	pdip tempuh tunda pemilu windah begal demokrat...
824	ahmad dol kurnia gera tunda pemilu windah bega...
918	april adil daerah khusus ibukota jakarta batal...
923	puji puji jokowi tunda pemilu

232 rows × 1 columns

Gambar 3. Tampilan Hasil *Preprocessing Data*

### 3.3 Labeling

Pada tahapan ini data hasil *Preprocessing* yang sudah diperoleh pada tahap sebelumnya, Label akan diberikan pada *dataset* yang dibagi menjadi dua kelas. *Tweet* dengan sentimen positif akan dilabelkan, dan *tweet* dengan sentimen negatif akan dilabelkan, sehingga label dapat diberikan berdasarkan nilai yang terkandung dalam *tweet*. Proses labelisasi sentimen dilakukan secara otomatis dengan menggunakan *lexicon-based*.

	text	StSc	label
0	pengadilan daerah khusus ibukota jakarta batal...	0.000000	Positive
1	pengadilan daerah khusus ibukota jakarta batal...	0.000000	Positive
2	pengadilan daerah khusus ibukota jakarta batal...	0.000000	Positive
3	partai prima menghormati putusan pembatalan pe...	0.142857	Positive
5	pengadilan daerah khusus ibukota jakarta memba...	0.071429	Positive
...	...	...	...
819	ikuti jejak prima partai berkarya gugat perdat...	0.181818	Positive
823	pdip tempuh tunda pemilu windah begal demokrat...	-0.083333	Negative
824	ahmad doli kurnia gerakan tunda pemilu windah ...	0.000000	Positive
918	april pengadilan daerah khusus ibukota jakarta...	0.000000	Positive
923	puji puji jokowi tunda pemilu	0.400000	Positive

232 rows x 3 columns

Gambar 4. Tampilan Hasil Labelisasi

### 3.4 Pembobotan TF-IDF

Pada tahap ini, kata-kata dalam dokumen dinilai berdasarkan frekuensi di mana mereka muncul. Dimulai dengan menghitung nilai *term frekuensi* (TF) dan nilai *dokumen frekuensi* (DF), nilai *invers dokumen frekuensi* (IDF), dan kemudian mengalikan TF dan IDF.

	0	1	2	3	4	5	6	7	8	9	...	222	223	224	225	226	227	228	229
abang	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000
abam	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000
adi	0.443088	0.164505	0.443088	0.0	0.403701	0.0	0.198875	0.233311	0.322585	0.184595	...	0.0	0.0	0.0	0.0	0.267141	0.0	0.000000	0.000000
advokat	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000
agung	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
waspada	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000
wenang	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000
windah	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.327694	0.299595
yakin	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000
zalmi	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000

512 rows x 232 columns

Gambar 5. Document Term Matrix Proses TF-IDF

### 3.5 Klasifikasi Naïve Bayes Classifier

Selanjutnya, data uji diklasifikasikan menggunakan metode *Naive Bayes Classifier*. Untuk membuat keputusan, metode klasifikasi *Naive Bayes* menggunakan prediksi suatu kasus berdasarkan hasil klasifikasi yang telah diperoleh. Hasil ekstraksi fitur akan digunakan dalam proses klasifikasi.

```

1 # naive bayes classifier
2 mnb = MultinomialNB()
3 mnb.fit(x_train, y_train)
4 mnb_pred = mnb.predict(x_test)
5 mnb_acc = accuracy_score(y_test, mnb_pred)
6 print("Test accuracy: {:.2f}%".format(mnb_acc*100))

```

Test accuracy: 91.67%

Gambar 6. Hasil Proses Klasifikasi Naïve Bayes Classifier

Hasil perhitungan berbentuk matriks berordo 2x2 yang sesuai dengan jumlah kelas sentimen yang menunjukkan nilai positif dan negatif. Hasil *matrix confusion* menghitung nilai *accuracy*, *recall*, *precision*, dan *f1-score* sistem yang dibangun selama proses klasifikasi. Tabel berikut menunjukkan hasil *confusion matrix* penelitian.:

Tabel 1. Hasil Nilai Klasifikasi *Confusion Matrix*

		Kelas Sebenarnya	
		Positif	Negatif
Kelas Prediksi	Positif	20	2
	Negatif	0	2

Dengan menggunakan rumus berikut dari Tabel 1, besaran nilai *accuracy*, *precision*, *recall*, dan *f1-score* dapat dihitung:

$$Accuracy = \frac{20 + 2}{20 + 2 + 0 + 2} * 100\% = 91.67\%$$

$$Recall = \frac{20}{20 + 0} * 100\% = 100\%$$

$$Precision = \frac{20}{20 + 2} * 100\% = 90.91\%$$

$$f1 - score = \frac{2 \times (90 \times 100)}{90 + 100} * 100\% = 95.23\%$$

Hasil perhitungan di atas menunjukkan bahwa 24 data uji memiliki *accuracy* sebesar 91,67%, *recall* sebesar 100.00%, *precision* 90,91%, dan *f1-score* sebesar 95,24%.

#### 4. Conclusion

Hasil penelitian tentang Penggunaan Algoritma *Klasifikasi Naive Bayes* untuk Analisis Sentimen Terkait Pemilu 2024 dapat ditarik kesimpulan sebagai berikut :

1. Penelitian ini menggunakan dataset crawling untuk mendapatkan 1280 data tweet (opini) tentang penundaan pemilu 2024.. Kemudian data tersebut dilakukan preprocessing sehingga data yang tersisa sebanyak 232 data. Data mengalami sejumlah penurunan dikarenakan data terindikasi duplikat dalam proses preprocessing sehingga terjadi penurunan dari 1280 data menjadi 232 data, kemudian dilakukan labelisasi sebanyak 232 data tweet (opini) yang akan dianalisis.
2. Metode berbasis leksik dengan kamus berbahasa Indonesia digunakan untuk melabelkan data penelitian ini. Metode ini akan menghitung seberapa banyak kata yang ditemukan dalam kamus yang berkaitan dengan dokumen (opini) dan juga menghitung skor sentimennya. Metode ini menghasilkan rincian data kelas positif sebanyak 67.7% yaitu 157 data opini, dan negatif sebanyak 32.3% yaitu 75 data opini.
3. Hasil akurasi 232 dataset dengan perbandingan 9:1 antara data latih dan uji menunjukkan bahwa metode *Naive Bayes Classifier* pada sistem klasifikasi sentimen sangat akurat, pembobotan kata dengan TF-IDF dengan nilai akurasi sebesar 91,67%, *precision* 90,91%, *recall* 100%, dan nilai *f1-score* sebesar 95,24% diperoleh.

#### References

- [1] A. Hidayat, "Manfaat Pelaksanaan Pemilu untuk Kesejahteraan Masyarakat," *Jurnal Ilmu Politik*, vol. 2, no. 1, pp. 61–74, 2020.
- [2] T. DetikJateng, "Mahfud Md Klaim Pemilu 2024 Lebih Damai, Tak Perlu Ditunda," *detikSumut*. Accessed: Jul. 25, 2023. [Online]. Available: <https://www.detik.com/sumut/berita/d-6825706/mahful-md-klaim-pemilu-2024-lebih-damai-tak-perlu-ditunda>
- [3] A. Novrisal, G. I. Marthasari, and C. S. K. Aditya, "Sentimen Analisis Tweet Berbahasa Indonesia Pada Pilkada Serentak 2020 Menggunakan Metode *Naive Bayes* Berbasis Particle Swarm Optimization," *Jurnal REPOSITOR*, vol. 3, no. 2, pp. 1–7, 2021.
- [4] M. K. Anam, M. B. Firdaus, T. A. Fitri, W. Agustin, and Agustin, "Analisis Pilkada Medan pada Sosial Media Menggunakan Analisis Sentimen dan Social Network Analysis," *Indonesian Journal of Computer Science*, vol. 11, no. 1, pp. 101–114, 2022.

- 
- [5] A. Anggito and J. Setiawan, *Metodologi Penelitian Kualitatif*, 1st ed. Sukabumi: CV Jejak, 2018.
- [6] C. Casro, Y. Purwanti, G. Setyaningsih, and A. P. Kuncoro, "Rancang Bangun Aplikasi Pengaduan Pelanggan Berbasis Web Menggunakan Framework Codeigniter Di Indotechno Purwokerto," *Jurnal Sains dan Informatika*, vol. 6, no. 2, pp. 166–174, 2020.
- [7] I. Irwanto, "Perancangan Sistem Informasi Sekolah Kejuruan dengan Menggunakan Metode Waterfall (Studi Kasus SMK PGRI 1 Kota Serang-Banten)," *Lectura : Jurnal Pendidikan*, vol. 12, no. 1, pp. 86–107, 2021.
- [8] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*, 1st ed. Yogyakarta: C.V ANDI OFFSET, 2012.
- [9] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika Jurnal*, vol. 10, no. 02, pp. 71–76, 2020.
- [10] A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*, 1st ed. Medan: Yayasan Kita Menulis, 2020.
- [11] E. A. Novia, W. I. Rahayu, and C. Prianto, *Sistem Perbandingan Algoritma K-Means Dan Naïve Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan*. Bandung: Kreatif Industri Nusantara, 2020.