
Perbandingan Akurasi SVM & Naive Bayes Pada Analisis Sentimen Data Berita Online Terhadap COKLIT Pemilu 2024

Sarah Nikensia Mawikere¹, Irene R.H.T Tangkaarouw

nikenziamawikeregmail.com¹, irene.tangkawarow@unima.ac.id²

^{1,2} Universitas Negeri Manado

Informasi Artikel

Diterima : 14 Mar 2025

Direvisi : 12 Apr 2025

Disetujui : 30 Apr 2025

Kata Kunci

COKLIT Pemilu 2024,
Analisis Sentimen, SVM,
Naive Bayes, Berita
Online

Abstrak

Sebelum melakukan pemilihan umum, perlu melakukan pencocokkan data dari para pemilih. Untuk itu, untuk membantu Bawaslu dalam mengawasi proses pencocokkan dan penelitian tersebut kita perlu tanggapan masyarakat dalam proses tersebut. Tanggapan masyarakat diabil dari kutipan masyarakat pada Berita Online. Untuk mencari klasifikasi yang lebih tepat, pada penelitian ini penulis akan melakukan perbandingan hasil analisis sentimen pada berita online tentang tahapan coklit menggunakan metode Support Vector Mechine (SVM) dan metode Naive Bayes agar dapat mengklasifikasikan opini dan sentimen yang diperoleh dalam beberapa kategori, seperti positif, dan negative. Dengan menggunakan confussion matrix, hasil akurasi score yang didapatkan berdasarkan klasifikasi menggunakan SVM adalah 63.33%. Sedangkan hasil akurasi score berdasarkan klasifikasi menggunakan Naive Bayes lebih besar yaitu 60.0%.

Keywords

*COKLIT Pemilu 2024,
sentiment analysis, SVM,
Naive Bayes, online news.*

Abstract

Before conducting a general election, it is necessary to match data from voters. For this reason, to assist Bawaslu in supervising the matching and research process, we need public feedback in the process. The public's response is dibilical from public quotes on Online News. To find a more precise classification, in this study, the author will compare the results of sentiment analysis on online news about the coklit stage using the Support Vector Mechine (SVM) method and the Naive Bayes method in order to classify the opinions and sentiments obtained in several categories, such as positive, and negative. Using a confusion matrix, the accuracy score obtained based on classification using SVM is 63.33%. Meanwhile, the accuracy of the score based on classification using Naive Bayes is greater at 60.0%.

A. Pendahuluan

Indonesia adalah salah satu negara yang menganut sistem demokrasi. Dalam keberjalanan sistem demokrasi ditandai dengan diadakannya pemilihan umum secara periodik. “Coklit” atau pencocokkan dan penelitian merupakan istilah yang digunakan dalam pencocokkan dan penelitian data pemilih dalam pemilihan umum di Indonesia. Pencocokkan dan Penelitian (coklit) juga dapat merujuk pada proses verifikasi dan validasi data para pemilih yang telah dikumpulkan, untuk memastikan keakuratan dan keabsahan informasi yang diperoleh[1].

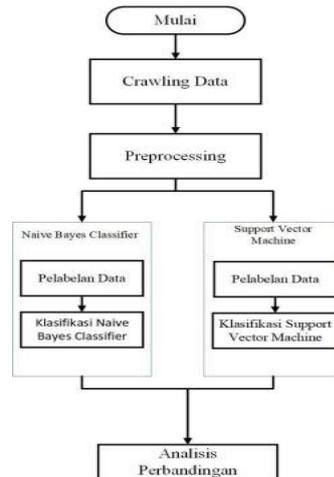
Analisis sentimen merupakan suatu bidang ilmu dari data mining yang berguna untuk menganalisis, mengolah, dan mengekstrak data tekstual pada entitas, seperti layanan, produk, individu, organisasi, peristiwa, atau masalah dan topik tertentu. Pada analisis sentimen, untuk mengedintifikasi kelas positif, negatif, dan netral banyak metode yang bisa digunakan. SVM (Support Vector Machine) dan Naïve Bayes adalah dua algoritma yang sering digunakan dalam analisis sentimen, yang bertujuan untuk mengklasifikasikan teks (seperti ulasan produk atau opini) menjadi kategori tertentu, misalnya positif, negatif, atau netral[2].

Berita online telah menjadi salah satu media massa yang paling sering dikonsumsi oleh masyarakat karena dapat diakses dimanapun dan kapanpun. Portal berita sebagai penyaji konten berita memiliki peranan besar dalam menyampaikan berita di media sosial[3].

Oleh karena itu, pentingnya melakukan analisis sentimen pada berita online terkait coklit pemilu 2024 karena menjadi salah satu bentuk partisipasi masyarakat dalam menanggapi bagaimana tahapan coklit yang dilakukan, serta dapat mendukung Bawaslu dalam mengawasi pemilihan umum yang akan berlangsung. Pada penelitian ini penulis akan melakukan perbandingan hasil analisis sentimen pada berita online tentang tahapan coklit menggunakan metode Support Vector Mechine (SVM) dan metode Naive Bayes agar dapat mengklasifikasikan opini dan sentimen yang diperoleh dalam beberapa kategori, seperti positif, dan negative[4]. Penelitian ini dilakukan sebagai bentuk partisipasi masyarakat dalam mengawasi jalannya pemilihan umum di Indonesia tahun 2024.

B. Metode Penelitian

Proses sentimen analisis terhadap Pencocokan dan Penelitian (Coklit) pemilu 2024 berdasarkan data yang didapat dari situs berita peneliti melakukan scraping data, tahap processing data, pelabelan data, lalu analisis sentimen menggunakan metode Support Vector Machine (SVM) Naïve Bayes, dan yang terakhir adalah membandingkan nilai akurasi dari kedua metode tersebut.



Gambar 1. Tahapan Penelitian

1. Scraping Data

Scraping data adalah proses mengumpulkan data dari sumber online, seperti situs web, media sosial, atau forum, dengan menggunakan teknik-teknik otomatisasi, seperti penggunaan bot atau perangkat lunak khusus.[5]

2. Processing Data

Data yang terkumpul masih berbentuk data yang belum terstruktur dengan isi dari setiap tweet masih dalam bahasa yang tidak baku. Tahap ini bertujuan untuk menghilangkan karakter yang tidak relevan dan mengurangi kualitas model selain itu juga dapat meningkatkan kualitas data dengan tujuan supaya data digunakan dapat digunakan pada tahap selanjutnya. Adapaun tahapan yang dilakukan pada penelitian ini adalah Pembersihan Data, Tokenizing, Normalisasi, Remove Stopwords, dan Steming.[6]

3. Pelabelan Data

Tahapan pelabelan bertujuan untuk mengidentifikasi data tweet untuk mengetahui polaritas teks, apakah tweet tersebut bersifat positif, negatif, atau netral[7]. Tahapan pelabelan data pada penelitian ini menggunakan kamus Inset Lexicon untuk melakukan pelabelan secara otomatis pada setiap tweet[8]. Kamus Inset Lexicon ini adalah kumpulan kata atau frasa yang telah diberikan nilai sentimen tertentu, sehingga dapat membantu mengklasifikasikan teks dengan cepat dan akurat. Setiap kata dalam tweet dicocokkan dengan entri yang ada di dalam kamus, di mana setiap kata diberi nilai sentimen berdasarkan kategori positif, negatif, atau netral.[9]

4. Naïve Bayes

Naïve Bayes adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi data. Algoritma ini berdasarkan pada teori probabilitas Bayes dan mengasumsikan bahwa fitur-fitur dalam data adalah independen satu sama lain[10].

Rumus Naïve Bayes[11]:

$$P(C|X) = P(X|C) * P(C) / P(X)$$

Keterangan:

- $P(C|X)$ = probabilitas bahwa suatu data termasuk dalam kelas C, diberikan fitur-fitur X.
- $P(X|C)$ = probabilitas bahwa fitur-fitur X muncul, diberikan kelas C.
- $P(C)$ = probabilitas prior kelas C.
- $P(X)$ = probabilitas fitur-fitur X.

5. Support Vector Machine

SVM (Support Vector Machine) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM berusaha memisahkan data menjadi dua kelas dengan menggunakan hyperplane yang optimal[12].

Rumus SVM:

Fungsi kerugian SVM:

$$L(w, b) = 1/2 ||w||^2 + C * \sum[\max(0, 1 - y_i(w * x_i + b))]^2$$

Keterangan:

- w: vektor bobot
- b: bias
- x_i : fitur-fitur data
- y_i : label kelas data
- C: konstanta regularization
- $||w||$: norma vektor w

Fungsi keputusan SVM:

$$f(x) = \text{sign}(w * x + b)$$

Keterangan:

- sign: fungsi signum yang mengembalikan 1 jika argumen positif, -1 jika argumen negatif

6. Perbandingan Nilai Akurasi

Setelah pengujian pada kedua metode tersebut, selanjutnya membandingkan nilai akurasi dari kedua metode tersebut yakni metode Support Vector Machine (SVM) dan metode Naïve Bayes. Dengan menggunakan confusion matrix, dalam konteks analisis sentimen (positif, negatif, dan netral), confusion

matrix dapat menunjukkan seberapa baik model mengklasifikasikan data berdasarkan sentimen yang sebenarnya. Dimana metode yang nilai akurasinya lebih tinggi, metode tersebutlah yang lebih tepat.

C. Hasil dan Pembahasan

1. Scraping Data

Pengambilan data atau scraping yang dilakukan didapatkan data yang berjumlah 309 data.

2. Preprocessing Data

Pada proses ini ada beberapa langkah yang harus dilakukan.

- **Pembersihan Data**

Proses yang pertama adalah membersihkan data, pada proses membersihkan data ini menghapus tanda baca yang tidak diperlukan, menghapus link, menghapus simbol, dan lain sebagainya. Tabel satu adalah contoh dari proses pembersihan data.

Table 1. Pembersihan Data

Teks	Hasil pembersihan data
Pengurangan TPS ini dipastikan terlaksana setelah proses pencocokan dan penelitian (coklit) data pemilih Pemilu 2024 selesai.	pengurangan tps ini dipastikan terlaksana setelah proses pencocokan dan penelitian coklit data pemilih pemilu selesai

- **Tokenizing**

Pada proses yang kedua yaitu Tokenizing, pada proses ini data tersebut terpecah menjadi kata-kata. Tabel 2 adalah hasil dari proses Tokenizing.

Table 2. Tokenizing

Teks	Tokenizing
pengurangan tps ini dipastikan terlaksana setelah proses pencocokan dan penelitian coklit data pemilih pemilu selesai	['pengurangan', ' ', 'tps', ' ', 'ini', ' ', ' ', 'dipastikan', ' ', ' ', 'terlaksana', ' ', ' ', 'setelah', ' ', ' ', 'proses', ' ', ' ', 'pencocokan', ' ', ' ', 'dan', ' ', ' ', 'penelitian', ' ', ' ', 'coklit', ' ', ' ', 'data', ' ', ' ', 'pemilih', ' ', ' ', 'pemilu', ' ', ' ', 'selesai']

- **Normalisasi**

Proses yang ketiga adalah Normalisasi data, pada proses ini mengubah kata-kata yang tidak baku menjadi kata-kata yang baku,

mengubah kata-kata singkatan menjadi kata-kata aslinya yang lebih jelas. Tabel 3 adalah hasil dari proses Normalisasi.

Table 3. Hasil Normalisasi

Teks	Normalisasi
['pengurangan', ' ', 'tps', ' ', 'ini', ' ', ' ', 'dipastikan', ' ', 'terlaksana', ' ', 'setelah', ' ', ' ', 'proses', ' ', 'pencocokan', ' ', ' ', 'dan', ' ', 'penelitian', ' ', ' ', 'coklit', ' ', ' ', 'data', ' ', ' ', 'pemilih', ' ', ' ', 'pemilu', ' ', ' ', 'selesai']	['pengurangan', ' ', 'tps', ' ', 'ini', ' ', ' ', 'dipastikan', ' ', 'terlaksana', ' ', 'setelah', ' ', ' ', 'proses', ' ', 'pencocokan', ' ', ' ', 'dan', ' ', 'penelitian', ' ', ' ', 'coklit', ' ', ' ', 'data', ' ', ' ', 'pemilih', ' ', ' ', 'pemilu', ' ', ' ', 'selesai']

- Penghapusan Stopword

Pada proses yang keempat adalah penghapusan stopwords, pada proses ini menghapus kata-kata yang tidak memiliki arti yang penting, dan meninggalkan kata-kata yang memiliki arti yang kontekstual. Tabel 4 adalah hasil dari Penghapusan Stopword.

Table 4. Penghapusan Stopword

Teks	Penghapusan Stopword
['pengurangan', ' ', 'tps', ' ', 'ini', ' ', ' ', 'dipastikan', ' ', 'terlaksana', ' ', 'setelah', ' ', ' ', 'proses', ' ', 'pencocokan', ' ', ' ', 'dan', ' ', 'penelitian', ' ', ' ', 'coklit', ' ', ' ', 'data', ' ', ' ', 'pemilih', ' ', ' ', 'pemilu', ' ', ' ', 'selesai']	['pengurangan', ' ', 'tps', ' ', ' ', ' ', 'terlaksana', ' ', ' ', ' ', 'proses', ' ', 'pencocokan', ' ', ' ', 'penelitian', ' ', ' ', 'coklit', ' ', ' ', 'data', ' ', ' ', 'pemilih', ' ', ' ', 'pemilu', ' ', ' ', 'selesai']

- Stemming

Selanjutnya adalah proses Stemming, yaitu mengubah sebuah kata menjadi kata dasarnya. Tabel 5 adalah hasil dari proses Stemming.

Table 5. Stemming

Teks	Stemming
['pengurangan', ' ', 'tps', ' ', ' ', ' ', 'terlaksana', ' ', ' ', ' ', 'proses', ' ', ' ', 'pencocokan', ' ', ' ', 'penelitian', ' ', ' ', 'coklit', ' ', ' ', 'data', ' ', ' ', 'pemilih', ' ', ' ', 'pemilu', ' ', ' ', 'selesai']	['kurang', ' ', 'tps', ' ', ' ', ' ', 'laksana', ' ', ' ', ' ', 'cocok', ' ', ' ', 'teliti', ' ', ' ', 'coklit', ' ', ' ', 'data', ' ', ' ', 'pilih', ' ', ' ', 'milu', ' ', ' ', 'selesai']

3. Pelabelan Data

Selanjutnya adalah proses pelabelan data, data-data yang sudah melewati Processing data selanjutnya akan diberi label antara label positif atau label

negatif, dengan mencocokkan setiap kata pada data dengan kata yang ada pada kamus insert lexicon.

Table 6. Kamus Insert Lexicon

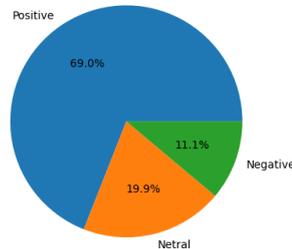
Positif		Negatif	
acungan jempol	Empati	Absurd	Dicuri
adaptif	ekonomis	Acuh	Dosa
adil	Enak	Alergi	Egois
ahli	Gagah	Ambigu	Gagap
akurat	Gembira	Aneh	Gatal
asyik	Gesit	Angkuh	Gila
bagus	Giat	Bajingan	Hama
baik	Halus	Basi	Hancur
banyak	Halal	Bau	ilegal
berani	Hangat	Benci	Ilusi
cantik	Hebat	Berdosa	Iri
cepat	Indah	Berlemak	Ironi
cepat	Idola	Cabul	Jelek
cerdas	Imut	Cebol	Jengkel
cerdik	Janji	Cemburu	Jijik
damai	Jempol	Curang	Kacau
dewasa	Juara	Dendam	Kalah

Setelah mencocokkan kata dengan kamus insert lexicon, maka data tersebut telah mendapatkan label antara positif, negatif atau netral. Pada tabel 7 menunjukkan hasil dari pelabelan tersebut.

Tabel 7. Hasil Pelabelan Data

	tweet	positive_counts	negative_counts	label
0	cocok teliti cokit data pilih milu pelosok de...	2	0	Positive
1	panitia daftar pilih pantarlih cokit lurah ma...	0	0	Netral
2	komisi pilih kpu kota denpasar lanjut tahap co...	2	0	Positive
3	deiyai suarapapua.com komisi pilih kpu kabupate...	0	0	Netral
4	kpu taban lanjut tahap milu warna muncul putus...	0	1	Negative
5	jakarta komisi pilih kpu kembang cocok teliti ...	3	0	Positive
6	tahap cocok teliti cokit data pilih milu sere...	2	0	Positive
7	pekanbaru infopublik tahap pilih milu proses c...	2	0	Positive
8	bawaslu sulbar awas cokit data pilih tahap mu...	1	1	Netral
9	jakarta kompascom komisi pilih kpu proses coco...	2	0	Positive

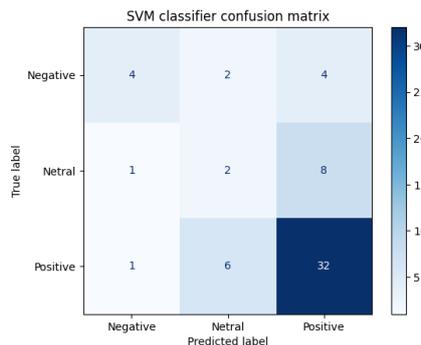
Dari hasil pelabelan data tersebut maka telah didapatkan hasil sebagai berikut, yaitu 205 data positif, 33 data negatif, dan 59 data netral. Gambar 2 menunjukkan visualisasi menggunakan pie chart.



Gambar 2. Visualisasi Pie Chart

4. Klasifikasi Support Vector Machine (SVM)

Setelah melakukan pelabelan data, selanjutnya melakukan klasifikasi pada metode SVM. Dengan menggunakan bahasa pemrograman Python, hasil dari analisis ini menghasilkan laporan klasifikasi yang mencakup berbagai metrik evaluasi, termasuk kemampuan penarikan kembali data (recall), presisi (precision), dan hasil akurasi (accuracy).



Gambar 3. Confusion Matrix SVM

Pada gambar 3 diatas Model confusion matrix menunjukkan bahwa secara benar sebanyak 32 data sebagai positif, 4 sebagai data negative dan 2 sebagai data netral.

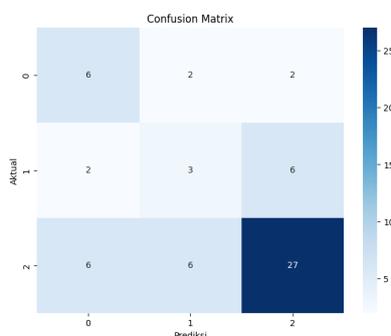
Untuk perhitungan akurasi manual bisa dilihat pada persamaan 1.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{True positif} + \text{True netral} + \text{True Negatif}}{\text{Total data matrix}} \times 100\% \quad (1) \\
 &= \frac{32 + 2 + 4}{60} \times 100\% \\
 &= \frac{38}{60} \times 100\% \\
 &= 63.33\%
 \end{aligned}$$

Jadi, pada klasifikasi menggunakan Metode SVM mendapatkan nilai akurasi sebesar 63.33%.

5. Klasifikasi Naïve Bayes

Selanjutnya melakukan klasifikasi pada metode Naïve Bayes. Dengan menggunakan bahasa pemrograman Python, hasil dari analisis ini menghasilkan laporan klasifikasi yang mencakup berbagai metrik evaluasi, termasuk kemampuan penarikan kembali data (recall), presisi (precision), dan hasil akurasi (accuracy).



Gambar 4. Confusion Matrix Naïve Bayes

Pada gambar 4 diatas Model confusion matrix menunjukkan bahwa secara benar sebanyak 27 data sebagai positif, 6 sebagai data negative dan 3 sebagai data netral.

Untuk perhitungan akurasi manual bisa dilihat pada persamaan 2.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{True positif} + \text{True netral} + \text{True Negatif}}{\text{Total data matrix}} \times 100\% & (2) \\
 &= \frac{27 + 3 + 6}{60} \times 100\% \\
 &= \frac{36}{60} \times 100\% \\
 &= 60.0\%
 \end{aligned}$$

Jadi, pada klasifikasi menggunakan Metode Naïve Bayes mendapatkan nilai akurasi sebesar 60.0%.

6. Perbandingan Nilai Akurasi

Pada proses yang terakhir adalah melakukan perbandingan antara Metode Support Vector Machine (SVM) dan Metode Naïve Bayes. Berdasarkan hasil klasifikasi kedua metode tersebut diatas, metode Support Vector Machine memperoleh hasil nilai akurasi sebesar 63.33%, sedangkan nilai akurasi yang diperoleh dari metode Naïve Bayes lebih rendah yaitu mencapai 60.0%. Jika kedua metode tersebut dibandingkan, maka pada khusus ini nilai akurasi Metode Support Vector Machine lebih unggul atau lebih tinggi daripada metode Naïve Bayes.

D. Simpulan

Kesimpulan dari penggunaan model prediksi yang dilakukan untuk menganalisis sentimen terhadap data dari berita online tentang cokit mendapatkan hasil komentar positif sebanyak 69.0% , hasil negatif sebanyak 11.1% , dan hasil yang netral adalah sebanyak 19.9%, dengan perincian komentar

positif sebanyak 205 komentar, negatif sebanyak 33 komentar, dan netral sebanyak 59 komentar.

Dengan menggunakan confusion matrix, hasil akurasi score yang didapatkan berdasarkan klasifikasi menggunakan SVM adalah 63.33%. Sedangkan hasil akurasi score berdasarkan klasifikasi menggunakan Naïve Bayes lebih besar yaitu 60.0%, dapat disimpulkan bahwa nilai akurasi menggunakan metode SVM lebih besar daripada menggunakan metode Naïve Bayes.

E. Ucapan Terima Kasih

Ucapan terima kasih penulis sampaikan kepada semua pihak yang sudah berkontribusi untuk membantu penyelesaian penelitian ini. Doa, semangat, serta bimbingan dan kontribusi yang diberikan sangatlah penting dalam penyelesaian penelitian ini.

F. Referensi

- [1] P. Abiyasa, "Kewenangan Bawaslu Dalam Penyelenggaraan Pemilu Di Kota Semarang Suatu Kajian Undang- Undang Nomor 7 Tahun 2017 Tentang Pemilu Bawaslu Authority In The Administration Of The Elections In Semarang A Study Act Number 7 Of 2017," Vol. 2, No. 2, Pp. 149–161, 2019.
- [2] D. S. K. D. P. R. (DPR) P. T. M. M. N. B. C. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," J. Inform. dan Tek. Elektro Terap., vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [3] P. S. Statistika, F. Matematika, D. A. N. Ilmu, P. Alam, and U. I. Indonesia, "Analisis Sentimen Pada Review Aplikasi Berita," 2020.
- [4] A. V. Sudiantoro et al., "Analisis Sentimen Twitter Menggunakan Text Mining Dengan," vol. 10, no. 2, pp. 398–401, 2018.
- [5] D. D. Ayani, H. S. Pratiwi, and H. Muhandi, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," vol. 7, no. 4, pp. 257–262, 2019.
- [6] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," vol. 10, pp. 71–76, 2020.
- [7] P. Studi, T. Informatika, F. Teknik, and U. M. Kudus, "Provider Indihome Berdasarkan Pendapat Pelanggan Melalui Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," 2021.
- [8] I. Susianti, S. S. Ningsih, M. Al Haris, and T. W. Utami, "Analisis Sentimen Pada Twitter Terkait New Normal Dengan Metode Naïve Bayes Classifier," Pros. Semin. Edusainstech FMIPA UNIMUS, pp. 354–363, 2020, [Online]. Available: <https://prosiding.unimus.ac.id/index.php/edusaintek/article/view/576/578>
- [9] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 3, no. 3, pp. 2789–2797, 2019.
- [10] C. B. Saputra, A. Muzakir, and D. Udariansyah, "Analisis Sentimen Masyarakat Terhadap #2019Gantipresiden Berdasarkan Opini Dari Twitter

- Menggunakan Metode Naive Bayes Classifier,” Bina Darma Conf. Comput. Sci., pp. 403–413, 2019.
- [11] J. Grafika and N. Kampus, “Studi Literatur Tentang Perbandingan Metode Untuk Proses,” vol. 2016, no. Sentika, pp. 18–19, 2016.
- [12] Wati, R., & Ernawati, S. (2021). Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 06, 240–247. <https://doi.org/10.54367/jtiust.v6i2.1465>
- [13] Y. Rahman and H. Wijayanto, “Klasifikasi Batik Menggunakan Metode K-Nearest Neighbour Berdasarkan Gray Level Co-Occurrence Matrices (GLCM),” *Jur. Tek. Inform. FIK UDINUS*, vol. 244, no. Ecpe, pp. 1–7, 2015.