

Analisis Sentimen Pada Berita Online Terhadap Coklit Pemilu 2024 Dengan Metode Naïve Bayes

Sentiment Analysis In Online News Regarding 2024 Election Matching And Research Using The Naïve Bayes Method

Sarah Nikensia Mawikere¹ Irene R.H.T. Tangkawarow^{1*}

¹Program studi Teknik Informatika, Fakultas Teknik, Universitas Negeri Manado

Article Info	ABSTRAK
Article history: Received: Des 09, 2024 Revised: Jan 10, 2025 Accepted: Jan 28, 2025	Indonesia adalah salah satu negara yang menganut sistem demokrasi. Pemilihan umum dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil setiap lima tahun sekali. Pengawasan dari Bawaslu adalah bentuk pengawasan yang terlembaga dari suatu organ negara. "Coklit" atau pencocokkan dan penelitian merupakan istilah yang digunakan dalam pencocokkan dan penelitian data pemilih dalam pemilihan umum di Indonesia. Pada penelitian ini penulis akan melakukan analisis terhadap data coklit pada berita online menggunakan metode Naive Bayes agar dapat mengklasifikasikan opini dan sentimen yang diperoleh dalam beberapa kategori, seperti positif, dan negative. Pada penelitian ini didapatkan komentar positif sebanyak 34 komentar, negatif sebanyak 60 komentar, dan netral sebanyak 17 komentar.
Kata kunci analisis sentimen, berita online, coklit 2024, naïve bayes	
Keywords Coklit 2024 , naïve bayes, online news, sentiment analysis	ABSTRACT <i>Indonesia is a country that adheres to a democratic system. General elections are held directly, publicly, freely, secretly, honestly and fairly every five years. Supervision from Bawaslu is an institutionalized form of supervision from a state organ. "Coklit" or matching and research is a term used in matching and researching voter data in general elections in Indonesia. In this research, the author will carry out an analysis of real data on online news using the Naive Bayes method in order to classify the opinions and sentiments obtained into several categories, such as positive and negative. In this research, there were 34 positive comments, 60 negative comments, and neutral as many as 17 comments.</i>
Corresponding Author: Sarah Nikensia Mawikere, Program Studi Teknik Informatika, Fakultas Teknik Universitas Negeri Manado, 001035. Jalan Kampus Unima. M. Kec. Tondano Selatan Email: nikenziamawikere@gmail.com	

PENDAHULUAN

Indonesia adalah salah satu negara yang menganut sistem demokrasi. Dalam keberjalanan sistem demokrasi ditandai dengan diadakannya pemilihan umum secara periodik. Undang-Undang Nomor 7 Tahun 2017 Tentang pemilihan umum Pasal I ayat (2) Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 menyatakan bahwa rakyat memiliki kedaulatan, tanggung jawab, hak dan kewajiban untuk secara demokratis memilih pemimpin yang akan membentuk pemerintahan guna mengurus dan melayani seluruh lapisan masyarakat, serta memilih wakil rakyat untuk mengawasi jalannya pemerintahan. Dalam konteks pengawasan pemilu di Indonesia, pengawasan terhadap proses pemilu dilembagakan dengan adanya lembaga Badan Pengawas Pemilu (Bawaslu). Pengawasan dari Bawaslu adalah bentuk pengawasan yang terlembaga dari suatu organ negara. Di samping pengawasan oleh Bawaslu, terdapat juga pengawasan yang dilakukan oleh masyarakat terhadap proses penyelenggaraan pemilu yang disebut dengan kegiatan pemantauan pemilu. Berdasarkan Perbawaslu No. 2 Tahun 2023 Tentang Pengawasan Partisipatif pada pasal 1 ayat (8) tentang Pengawasan Partisipatif adalah tugas Bawaslu, Bawaslu Provinsi, Bawaslu Kabupaten/Kota, dan Panwaslu 2 Kecamatan yang diselenggarakan untuk meningkatkan partisipasi masyarakat dalam pengawasan Pemilu dan/atau Pemilihan[1]. "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 yang telah dikumpulkan, untuk memastikan keakuratan dan keabsahan informasi yang diperoleh[2]. Pada penelitian ini penulis akan melakukan analisis sentimen[3] pada berita online[4] dengan kata kunci COKLIT menggunakan metode Naive Bayes agar dapat mengklasifikasikan sentimen yang diperoleh dalam beberapa kategori, seperti positif, dan negative[5]. Dengan demikian, petugas sensus dapat lebih memahami preferensi dan kebutuhan masyarakat, sehingga dapat membantu dalam perencanaan dan pengambilan keputusan yang lebih tepat dan efektif.

METODE PENELITIAN

1. Data Penelitian

Data yang digunakan adalah data yang berhubungan dengan Pencocokan dan Penelitian (Coklit) pemilu 2024. Pengambilan sampel data dilakukan secara acak pada tanggal 6 maret 2023. Pengambilan sampel data menggunakan kata kunci Pencocokan dan Penelitian (Coklit) pemilu 2024 dan mendapatkan data sebanyak 130 data.

2. Tahapan 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, analisis sentimen menggunakan metode Naïve Bayes.

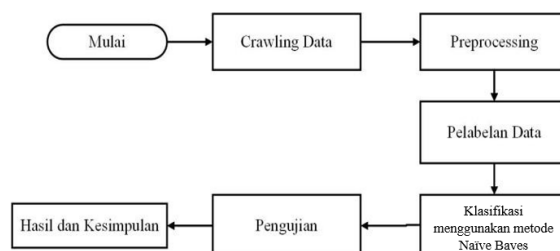


Figure 1 Tahapan Penelitian

3. Scraping Data

Scraping data adalah proses pengumpulan data secara otomatis dari website atau sumber online lainnya. Proses ini dapat dilakukan menggunakan perangkat lunak yang disebut scraper atau bot, yang mengambil data dari situs web secara teratur dan secara otomatis, lalu mengubahnya menjadi format yang dapat diproses atau diakses dengan mudah. Data yang dikumpulkan peneliti adalah data teks.[6]

4. Preprocessing 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. Adapun tahapan yang dilakukan pada penelitian ini adalah Pembersihan Data, Tokenizing, Normalisasi, Remove Stopwords, dan Steming.[7]

5. Pelabelan data

Tahapan pelabelan bertujuan untuk mengidentifikasi data tweet untuk mengetahui polaritas teks, apakah tweet tersebut bersifat positif, negatif, atau netral[8]. Tahapan pelabelan data pada penelitian ini menggunakan kamus *Inset Lexicon* untuk melakukan pelabelan secara otomatis pada setiap tweet[9]. 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.[10]

6. Naïve Bayes

Naive Bayes classifier merupakan sebuah metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan teorema Bayes dengan tidak ketergantungan (independen) yang tinggi. Algoritma Naïve Bayes digunakan sebagai penggolong dalam berbagai masalah dunia nyata seperti analisis sentimen, deteksi spam email, pengelompokan otomatis email, pengurutan email berdasarkan prioritas, dan kategorisasi dokumen. Metode Naive Bayes yang cocok digunakan dalam pengklasifikasian teks atau dokumen adalah multinomial Naive Bayes, dimana tidak hanya melihat kata yang muncul namun juga jumlah kemunculannya[11]. Metode pengklasifikasian tersebut dikemukakan oleh ilmuwan Inggris Thomas Bayes.

Pada persamaan 2.1 menunjukkan persamaan klasifikasi data dengan Naïve Bayes[12]:

$$P(C|X) = \frac{p(X|C)p(C)}{P(X)} \quad (1)$$

Dimana,

X = peubah acak

C = kelas

$P(C | X)$ = probabilitas C didasarkan pada kondisi X

$P(X | C)$ = probabilitas X didasarkan pada kondisi C

Pada persamaan 2.2 menunjukan Perhitungan yang dibutuhkan dalam NBC yaitu probabilitas tiap kelas:

$$P(C) = \frac{N_c}{N} \quad (2)$$

Dimana:

$P(C)$ = probabilitas C

N_c = jumlah total kelas C

N = jumlah total keseluruhan kelas

7. Pengujian dengan confusion matrix

teknik evaluasi yang umum digunakan dalam machine learning, khususnya untuk mengukur kinerja model klasifikasi. Confusion Matrix memberikan gambaran tentang bagaimana model klasifikasi bekerja dengan menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas yang ada[11]. Dalam konteks analisis sentimen (positif, negatif, dan netral), confusion matrix dapat menunjukkan seberapa baik model mengklasifikasikan tweet berdasarkan sentimen yang sebenarnya[13].

HASIL DAN PEMBAHASAN

1. Crawling Data

Proses scraping pada website berita dilakukan tanggal 6 mare t2023 dengan kata kunci “Pencocokkan dan Penelitian (COKLIT) 2024” Dari proses crawling data tersebut didapatkan data sebanyak 130 data yang belumdiketahui label sentimennya.

2. Preprocessing Data

Tahapan selanjutnya adalah preprocessing pada tahapan ini data hasil dari proses crawling data akan diproses hingga menjadi data yang bersih, Berikut tahapan dalam melakukan preprocessing:

- Pembersihan data

Proses ini mulai dari whitespace hingga menghilangkan karakter yang mengurangi kualitas data, seperti link, nama akun userseseorang dan tanda (#,@,r',dll),serta penghapusan stopword. Program akan mencari karakter yang telah tentukan, kemudian karakter tersebut akan dihapus dari data tersebut.

Tabel 1 hasil 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

Tokenizing adalah proses memecah dokumen menjadi kumpulan kata. Tokenization dapat dilakukan dengan menghilangkan tanda baca dan memisahnya per-spasi. Tabel.2 menunjukan hasil tahapan tokenizing:

Tabel 2 hasil tokenizing

Teks	Hasil Tokenizing
pengurangan tps ini dipastikan terlaksana setelah proses pencocokan dan penelitian cokit data pilih pemilu selesai	['pengurangan', '', 'tps', '', 'ini', '', 'dipastikan', '', 'terlaksana', '', 'setelah', '', 'proses', '', 'pencocokan', '', 'dan', '', 'penelitian', '', 'cokit', '', 'data', '', 'pilih', '', 'pemilu', '', 'selesai']

- Normalisasi

Normalisasi adalah proses yang berkaitan dengan model data relational untuk mengorganisasi himpunan data dengan ketergantungan dan keterkaitan yang tinggi atau erat[14]. Tabel.3 menunjukan hasil Normalisasi:

Tabel 3 hasil Normalisasi

Teks	Hasil Normalisasi
['pengurangan', '', 'tps', '', 'ini', '', 'dipastikan', '', 'terlaksana', '', 'setelah', '', 'proses', '', 'pencocokan', '', 'dan', '', 'penelitian', '', 'cokit', '', 'data', '', 'pilih', '', 'pemilu', '', 'selesai']	['pengurangan', '', 'tps', '', 'ini', '', 'dipastikan', '', 'terlaksana', '', 'setelah', '', 'proses', '', 'pencocokan', '', 'dan', '', 'penelitian', '', 'cokit', '', 'data', '', 'pilih', '', 'pemilu', '', 'selesai']

- Penghapusan Stopword

Tahapan ini menghilangkan kata yang tidak deskriptif atau tidak penting karena tidak mengandung atau merepresentasikan data[15]. Tabel.4 menunjukan hasil dari Penghapusan Stopword:

Tabel 4 hasil Penghapusan Stopword

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

- Stemming Data

Proses ini berupa penghilangan imbuhan dan akhiran pada setiap token. Proses ini menggunakan library Pysastrawi berdasarkan algoritme Nazief dan Adriani dengan menghilangkan imbuhan dari sebuah kata menjadi kata dasar[16]. Tabel.5 menunjukan hasil dari proses Stemming:

Tabel 5 Hasil Stemming

Teks	Hasil Steamming
['pengurangan', '', 'tps', '', 'ini', '', 'terlaksana', '', 'proses', '', 'pencocokan', '', 'dan', '', 'penelitian', '', 'cokit', '', 'data', '', 'pilih', '', 'pemilu', '', 'selesai']	['kurang', '', 'tps', '', 'laksana', '', 'proses', '', 'cocok', '', 'teliti', '', 'cokit', '', 'data', '', 'pilih', '', 'pemilu', '', 'selesai']

'penelitian', '', 'coklit', '', 'data', '', 'pilih', 'milu', '', 'selesai'] '', 'pemilu', '', 'selesai']	
---	--

3. Pelabelan data

Data yang sudah melalui tahapan text preprocessing dan dikelompokkan per kata kemudian dicocokkan dengan kamus Inset Lexicon untuk mengetahui polaritas text apakah positif, negatif, atau netral. kamus kata yang digunakan, dapat dilihat pada Tabel 6:

Tabel 6 kamus Inset Lexicon

Positif		Negatif	
acungan jempol	Empati	Absurd	Dicuri
adaptif	ekonomis	Acuh	Dosa
adil	Enak	Alergi	Egois
ahli	Gagah	Ambigu	Gagap
Positif		Negatif	
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

Setelah pencocokan dengan kamus kata, selanjutnya dilakukan training set. Dari hasil training set tersebut menghasilkan polaritas text. Sebuah text/tweet dianggap positif atau negatif jika kata tersebut mempunyai keterkaitan dengan kamus kata, dan dianggap netral jika text/tweet tidak terdapat kata yang terkait dalam kamus kata. tabel.7 merupakan hasil dari tahap pelabelan data yang dilakukan dengan mencocokkan dengan kamus kata:

Tabel 7 hasil dari tahap pelabelan data

No	content_clean_stemm	positive_counts	negative_counts	label
1.	advertisement padang - cocok teliti coklit dat...	3	5	negative
2.	marabahan - giat selenggara jelang milu 2024 j...	3	13	negative
3.	advertisement adil negeri jakarta pusat pn jak...	3	4	negative
4.	advertisement komisi pilih kpu taban lanjut ta...	3	9	negative
5.	bagi baca berita kpu pasti selenggara...	1	1	neutral

Setelah melakukan labeling, kita akan menghitung hasil dari jumlah setiap label yang

dihasilkan, dan hasil yang didapatkan adalah sebagai berikut :

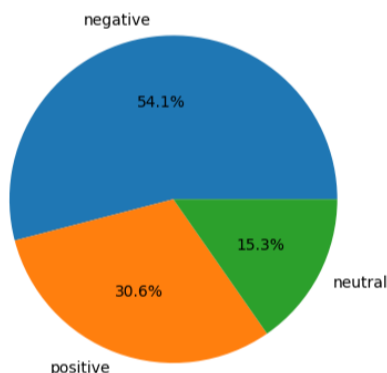


Figure 2Pie Chart Hasil Pelabelan

Gambar diatas merupakan hasil persentase. Dari hasil tersebut dapat dilihat nilai dari data negatif adalah sebanyak 54.1%, data positif sebanyak 30.6%, dan data yang dianggap netral adalah sebanyak 15.3%.

4. Analisis Menggunakan Naïve Bayes

Analisis sentimen merupakan salah satu pendekatan yang banyak digunakan pemrosesan bahasa alami *Natural Language Processing* untuk memahami bagaimana perasaan atau pendapat pengguna terhadap suatu topik. Dalam hal ini, metode *Naïve Bayes* diterapkan untuk mengklasifikasikan sentimen menjadi beberapa kategori, seperti negatif, netral, dan positif. 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).

Tabel 8 Hasil accurasi Analisis Naïve Bayes

	precision	recall	f1-score	support
negative	0.67	0.77	0.71	13
neutral	0.00	0.00	0.00	4
positive	0.33	0.40	0.40	5
accuracy			0.55	22
macro avg	0.33	0.39	0.36	22
weighted avg	0.47	0.47	0.50	22

Nilai keberhasilan sistem dalam penelitian ini dengan akurasi sebesar 54.54%. Selanjutnya dari hasil model confusion matrix dapat dilihat nilai precision dan recall di setiap kelas klasifikasi yang memiliki tingkat kemampuan yang berbeda-beda. Angka precision pada kelas negative sebesar 0.67 untuk kelas netral 0.00, dan untuk kelas positif 0.33. Sedangkan angka recall pada kelas negatif sebesar 0.77, pada kelas netral 0.00 dan untuk kelas positif sebesar 0.40. Selanjutnya menggunakan *Confusion Matrix* menggambarkan performa prediksi model dalam setiap kategori sentimen.

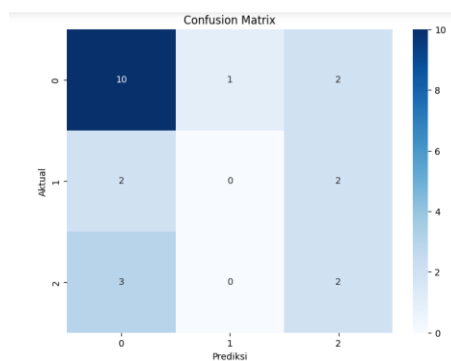


Figure 3 confusion matrix Analisis Naïve Bayes

Gambar 3 Pada gambar diatas Model confusion matrix menunjukkan bahwa secara benar sebanyak 2 data sebagai positif, 10 sebagai data negative dan 2 sebagai data netral.

Perhitungan akurasi manual dari perhitungan matrix sebagai berikut:

$$Akurasi = \frac{TP+TN+TN}{Total\ Data\ Matrix} \times 100\%$$

$$Akurasi = \frac{10+0+2}{22} \times 100\%$$

$$= \frac{12}{22} \times 100\%$$

$$=54.54\%$$

SIMPULAN

Dari penggunaan model prediksi yang dilakukan untuk menganalisis sentimen terhadap data dari berita online tentang covid mendapatkan hasil komentar positif sebanyak 30.6% , hasil negatif sebanyak 54.1% , dan hasil yang netral adalah sebanyak 15.3%, dengan rincian komentar positif sebanyak 34 komentar, negatif sebanyak 60 komentar, dan netral sebanyak 17 komentar. Hasil akurasi score berdasarkan klasifikasi menggunakan Naïve Bayes adalah sebesar 54.54% dengan rincian Angka precision pada kelas negative sebesar 0.67 untuk kelas netral 0.00, dan untuk kelas positif 0.33. Sedangkan angka recall pada kelas negatif sebesar 0.77, pada kelas netral 0.00 dan untuk kelas positif sebesar 0.40

UCAPAN TERIMA KASIH

Ucapan terima kasih saya sampaikan kepada Tuhan Yesus Kristus yang telah memberikan kesempatan serta kemampuan bisa menyusun artikel ini sampai selesai. Terima kasih juga kepada keluarga (mama,papa,kakak,adik) yang selalu memberikan semangat serta dorongan, dan juga selalu membantu dalam banyak hal untuk penyelesaian artikel ini. Ucapan terima kasih juga kepada Universitas Negeri Manado, Program Studi Teknik Informatika yang menyediakan sarana dan prasarana dalam penerbitan artikel ini. Terima kasih juga kepada Ibu Dr. Irene R.H.T. Tangkawarouw selaku dosen yang selalu membimbing,memberikan bantuan ilmu serta tenaga dalam menyusun artikel ini sampai selesai.

DAFTAR PUSTAKA

- [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] “PERATURAN KOMISI PEMILIHAN UMUM REPUBIK INDONESIA NOMOR 2 TAHUN 2017”.
- [3] 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.
- [4] P. S. Statistika, F. Matematika, D. A. N. Ilmu, P. Alam, and U. I. Indonesia, “ANALISIS SENTIMEN PADA REVIEW APLIKASI BERITA,” 2020.
- [5] A. V. Sudiantoro *et al.*, “Analisis Sentimen Twitter Menggunakan Text Mining Dengan,” vol. 10, no. 2, pp. 398–401, 2018.
- [6] D. D. Ayani, H. S. Pratiwi, and H. Muhardi, “Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace,” vol. 7, no. 4, pp. 257–262, 2019.
- [7] 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.
- [8] 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.
- [9] 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>
- [10] 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.
- [11] 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.
- [12] J. Grafika and N. Kampus, “STUDI LITERATUR TENTANG PERBANDINGAN METODE UNTUK PROSES,” vol. 2016, no. Sentika, pp. 18–19, 2016.
- [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.