

Studi Komparatif: Performansi Akurasi Algoritma Klasifikasi untuk Analisis Sentimen Pada Kandidat Presiden RI di Pemilu 2024

Ryan Putranda Kristianto*¹, Hendra², Stephanus Surijadarma Tandjung³

^{1,2,3}Prodi Ilmu Informatika, Fakultas Teknik, Universitas Katolik Darma Cendika (UKDC) Surabaya

E-mail: ¹ryan@ukdc.ac.id, ²hendra@student.ukdc.ac.id, ³stephanusst@ukdc.ac.id

(Naskah masuk: 13 Desember 2023, diterima untuk diterbitkan: 20 Oktober 2024)

Abstrak: Penelitian ini mengulas analisis sentiment terkait pemilihan presiden dan wakil presiden Republik Indonesia tahun 2024 yang menjadi sorotan utama di platform sosial media dan berita nasional. Fokus pada Anies Baswedan, Prabowo Subianto, dan Ganjar Pranowo, ulasan masyarakat menunjukkan variasi sentimen positif dan negatif. Dalam menggunakan algoritma klasifikasi seperti K-Nearest Neighbors (KNN), Multinomial Naive Bayes, Decision Tree dan Stochastic Gradient Descent dalam supervised learning, tujuannya adalah memberikan wawasan mendalam tentang pandangan masyarakat terhadap calon pemimpin. Dari hasil analisis, skenario terbaik adalah skenario 1 dengan akurasi rata-rata 75%, sedangkan Multinomial Naive Bayes menjadi algoritma terbaik dengan akurasi rata-rata 79%. Penelitian ini tidak hanya memberikan kontribusi ilmiah, tetapi juga menyoroti peran opini publik dalam proses demokratisasi dan pemilihan pemimpin negara, memberikan pemahaman lebih mendalam tentang preferensi masyarakat terhadap calon pemimpin.

Kata Kunci – K-Nearest Neighbors (KNN); Multinomial Naive Bayes; Decision Tree; Stochastic Gradient Descent; Pemilu 2024

Comparative Study: Accuracy Performance of Classification Algorithms for Sentiment Analysis on Indonesian Presidential Candidates in the 2024 Election

Abstract: This research reviews sentiment analysis related to the 2024 presidential and vice-presidential election of the Republic of Indonesia, which has become a major focus on social media platforms and national news. Focusing on Anies Baswedan, Prabowo Subianto, and Ganjar Pranowo, public reviews show variations in positive and negative sentiments. In using classification algorithms such as K-Nearest Neighbors (KNN), Multinomial Naive Bayes, Decision Tree, and Stochastic Gradient Descent in supervised learning, the aim is to provide deep insights into public views towards potential leaders. From the analysis results, the best scenario is scenario 1 with an average accuracy of 75%, while Multinomial Naive Bayes becomes the best algorithm with an average accuracy of 79%. This research not only provides scientific contributions but also highlights the role of public opinion in the process of democratization and selection of national leaders, providing a deeper understanding of public preferences towards potential leaders.

Keywords – K-Nearest Neighbors (KNN); Multinomial Naive Bayes; Decision Tree; Stochastic Gradient Descent; 2024 Election

1. PENDAHULUAN

2024 merupakan waktu pemilihan presiden dan wakil presiden republik indonesia, di mana presiden yang akan bertanding pada pemilihan tersebut yaitu Anies Baswedan, Prabowo Subianto dan Ganjar Pranowo menjadi bahan perbincangan pada sosial media di Indonesia seperti Facebook Instagram, bahkan media-media berita nasional memanas membicarakan elektabilitas dan popularitas bakal calon presiden tersebut. Perbincangan dan ulasan atau review user atau masyarakat dalam membicarakan elektabilitas dan popularitas bakal calon presiden tersebut sangat bervariasi dimulai dari sentimen positif dan negatif, oleh sebab itu, hal ini merupakan topik yang cukup panas pada permulaan menjelang pemilihan presiden republik Indonesia.

Ulasan-ulasan review user terkait topik tersebut dapat dianalisa untuk mendapatkan dan memahami pola patternnya, selain itu analisa terkait ulasan review user tersebut dapat digunakan untuk mengetahui proporsi antara sentimen negatif dan positif antara masing masing bakal calon presiden, teknik analisis ini dinamakan analisis sentimen [8].

Analisis sentimen dapat digunakan untuk menganalisa kecenderungan user apakah ulasannya merupakan sentimen negatif atau sentimen positif untuk melakukan analisis sentimen digunakan model algoritma klasifikasi seperti contohnya KNN, Multinomial Naive Bayes, Decision Tree dan lain - lain algoritma tersebut merupakan algoritma yang tergolong ke dalam supervised learning menggunakan algoritma klasifikasi, sebuah ulasan yang berasal dari sosial media dapat diketahui kecenderungan polarisasinya. Penting untuk mengetahui proporsi polarisasi ulasan pengguna sosial media terhadap pemilu presiden 2024 yang akan datang dikarenakan topik permasalahan yang sedang trending topik di Indonesia

Analisis Sentimen terhadap bakal calon kandidat presiden 2024 telah banyak dilakukan oleh peneliti - peneliti sebelumnya dengan menggunakan banyak pendekatan metode dan algoritma, seperti Naïve Bayes [1][3][4], SVM [2][5], bahkan kombinasi algoritma Naïve Bayes dan PSO juga digunakan oleh peneliti dalam melakukan analisis sentiment [3][6][7]. Pada penelitian tersebut, metode evaluasi yang digunakan adalah seperti: Confusion Matrix dan AUC.

Tujuan khusus dari penelitian ini adalah selain memberikan kontribusi informasi kemasyarakatan terkait pemilihan pemilu presiden 2024 dan proporsi polarisasi user indonesia dalam menilai bakal calon presiden yang akan bertanding selain itu memberikan kontribusi ilmu pengetahuan dan teknologi terkait studi komparatif algoritma klasifikasi untuk sentimen analisis terhadap permasalahan ini, dimana pada penelitian studi komparatif ini digunakan algoritma: KNN, Multinomial Naïve Bayes, Decision Tree dan Stochastic Gradient Descent.

Banyaknya bertebaran ulasan ulasan dan berita yang ditujukan untuk bakal calon presiden yang akan bertanding pada pemilu 2024 dianggap pro dan kontra atau mendukung dan sebaliknya anggapan subjektivitas proporsi ulasan yang mengarah ke negatif bukan hanya dari 1 atau 2 sumber suara saja oleh karena itu penting untuk memvalidasi ulasan ulasan yang bertebaran di sosial media dan proporsinya terkait pemilu 2024

2. METODE PENELITIAN

Pada penelitian ini, digunakan data yang bersumber dari Sosial Media Twitter, dengan kata kunci "prabowo", "prabowo subianto", "ganjar", "ganjar pranowo", "anies", "anies baswedan". Selanjutnya dengan berdasar kata kunci tersebut, dilakukan crawling ulasan - ulasan tweet pengguna menggunakan aplikasi add ons Google Chrome: No Coding Data Scraper. Dari ulasan tersebut, selanjutnya data akan di lakukan preprocessing data menggunakan Teknik: Case Folding, Tokenizing, Normalization, Stopwords, Stemming. Pada penelitian eksperimen ini, akan digunakan sederet algoritma klasifikasi seperti: Stochastic Gradient Descent, KNearest Neighbour (KNN), Random Forest (Ensemble Method), Naïve Bayes dan Logistic Regression. Ke 5 algoritma tersebut merupakan algoritma tradisional [9] (kecuali Stochastic Gradient Descent merupakan algoritma optimisasi) yang sangat populer digunakan untuk keperluan klasifikasi data. Digunakan metode evaluasi Confusion Matrixs untuk mengukur tiap model algoritma. Digunakan skenario pengujian eksperimen seperti berikut:

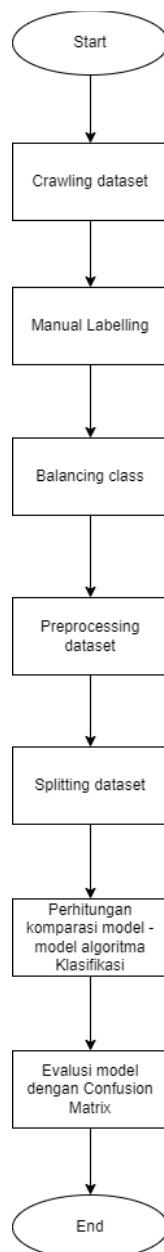
1. Splitting data training 80% : data testing 20%
2. Splitting data training 70% : data testing 30%
3. Splitting data training 60% : data testing 40%

Skenario pengujian tersebut digunakan untuk mendapatkan perspektif peneliti terkait evaluasi model yang dilakukan. Metode penelitian yang ditunjukkan pada gambar 1, dimulai dengan melakukan crawling dataset, kemudian selanjutnya dilakukan pelabelan manual, kemudian dilakukan penyeimbangan proporsi dataset tiap kelas, hal ini dilakukan untuk menghindari ketidakseimbangan antar kelas, dimana mengakibatkan menurunnya tingkat akurasi pada tiap model algoritma yang dibangun. Sesudah itu dilakukan preprocessing dataset, dimana pada

tahapan ini dilakukan pembersihan dataset, dimana seperti yang telah disebutkan pada paragraf sebelumnya, bahwa teknik preprocessing yang digunakan yaitu [10]:

1. Case Folding: proses merubah seluruh karakter teks menjadi huruf kecil
2. Tokenizing: proses merubah teks menjadi serangkaian teks yang lebih kecil
3. Normalization: proses merubah teks ke dalam bentuk standar dan normal atau baku
4. Stopwords: proses menghilangkan teks yang mengandung nilai gain yang rendah (tidak informatif)
5. Stemming : proses menghilangkan imbuhan awalan dan akhiran dari sebuah teks

Setelah dataset bersih, langkah selanjutnya, adalah melakukan splitting data, pada tahap splitting data ini, skenario pengujian diterapkan untuk mendapatkan perspektif akurasi model algoritma dari setiap skenario, setelah itu baru dilakukan implementasi algoritma. Tahapan terakhir adalah melakukan evaluasi model menggunakan metode Confusion Matrix.



Gambar 1. Flowchart metode penelitian

Dalam melakukan proses crawling dataset, peneliti menggunakan addOns Chrome No Coding Data Scraper, untuk labelling dataset dan balancing class dilakukan secara manual dengan

menganalisa dataset yang sudah dicrawling. Selanjutnya proses PreProcessing, splitting, perhitungan algoritma hingga evaluasi model algoritma dilakukan dengan membuat coding program berupa model Machine Learning di Google Collaboratory.

Mengacu pada gambar 1, yaitu pada tahapan preprocessing dataset, peneliti akan melakukan visualisasi dataset, guna mencari pola dan profiling dari dataset tersebut, dengan teknik ini, peneliti setidaknya akan memperoleh gambaran mengenai dataset terkait ulasan masyarakat mengenai bakal calon kandidat presiden. Pada tahapan perhitungan algoritma, peneliti akan mendapatkan hasil perhitungan algoritma dalam mempelajari pola dataset yang digunakan, sehingga ketika dilakukan uji testing data baru pada model algoritma, dapat dikenali apakah data ulasan tersebut merupakan Negatif atau Positif.

3. HASIL DAN PEMBAHASAN

Penulis telah melakukan eksperimen pencarian model di Google Collab, dengan mengacu kepada skenario pengujian berikut:

1. Splitting data training 80% : data testing 20%
2. Splitting data training 70% : data testing 30%
3. Splitting data training 60% : data testing 40%

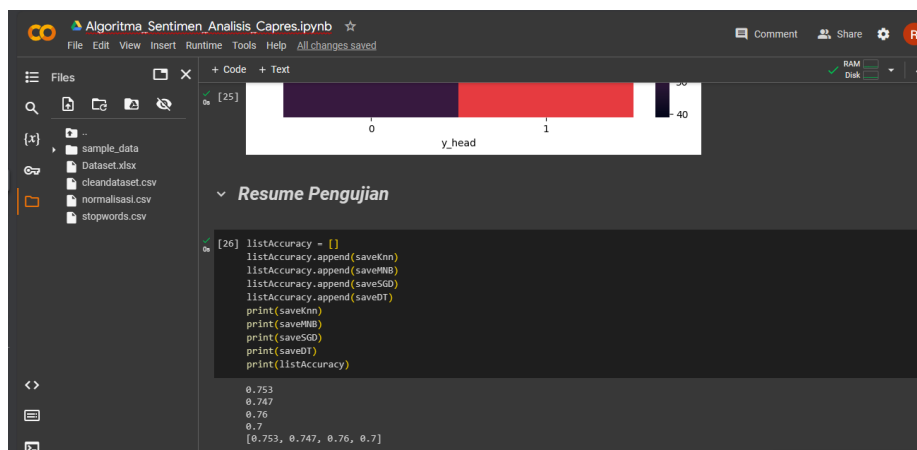
241	Satu kata untuk pak Anies baswedan : Belum Layak R	NEGATIF	Anies Baswedan
242	Rekam jejaknya di Jkt saja menyedihkan begini sok2	NEGATIF	Anies Baswedan
243	Bisa ngomong doang nih si Anis. Gak bisa kerja. Bagu	NEGATIF	Anies Baswedan
244	Pendidikan tinggi, omong jago, tapi ngga bisa kerja c	NEGATIF	Anies Baswedan
245	Loe waktu di Jakarta aja nga ada peningkatan lapang	NEGATIF	Anies Baswedan
246	Sepertinya anda kurang banyak mendengar, karena i	NEGATIF	Anies Baswedan
247	Ngurusin pengangguran di DKI aja ga becus!!!!	NEGATIF	Anies Baswedan
248	Halah sok peduli...	NEGATIF	Anies Baswedan
249	Kaya bisa kerja aja loe, cuma bisa nata kata	NEGATIF	Anies Baswedan
250	Bicara rekam jejak @aniesbaswedan Yg menonjol ha	NEGATIF	Anies Baswedan
251	Emangnya ma rakyat yg ada di dlm negeri sdh g laku	NEGATIF	Anies Baswedan
252	Semangat pak Prabowo SubiantoPemimpin Berkelas	POSITIF	Prabowo Subianto
253	Saya sangat support apa yang dikerjakan SubiantoPe	POSITIF	Prabowo Subianto
254	Tapi kok 3x kalah ya...sungguh mengherankan	NEGATIF	Prabowo Subianto
255	Bijaksananya pesonality pak Prabowo BetulSubiantc	POSITIF	Prabowo Subianto
256	Kematangan semangat BetulSubianto YangTerbaik it	POSITIF	Prabowo Subianto
257	momen ini mengandung bawang air mata beliau bu	POSITIF	Prabowo Subianto
258	Fix prabowo next president	POSITIF	Prabowo Subianto
259	prabowo presiden Tetap menjadi presiden pastinya	POSITIF	Prabowo Subianto
260	Prabowo presiden pemersatu.	POSITIF	Prabowo Subianto
261	PILPRES 2024 1x putaran ! PRABOWO PRESIDEN	POSITIF	Prabowo Subianto
262	sehat selalu bapak prabowo mudah mudahan bapa s	POSITIF	Prabowo Subianto

Gambar 2. Screenshot dataset #1

Penulis telah mengumpulkan data crawling dari Twitter, dengan keyword “Prabowo Subianto”, “Anies Baswedan” dan “Ganjar Pranowo”, seperti ditunjukkan pada gambar 2 dan 3. Setelah itu penulis mulai mengimplementasikannya dengan menggunakan code Python pada Google Collab, seperti ditunjukkan pada gambar 4.

491	Prabowo cuma itu itu saja yg dibicarakan.	NEGATIF	Prabowo Subianto
492	semua orang indonesia juga bakal bela negaranya, p	NEGATIF	Prabowo Subianto
493	Sudah tua seharusnya pensiun bos. Mendingan men	NEGATIF	Prabowo Subianto
494	Lu udah terlalu tue memimpin negeri ini... Lu punya	NEGATIF	Prabowo Subianto
495	Kalau membela bangsa Indonesia. Kenapa fadlizon y	NEGATIF	Prabowo Subianto
496	SUDAH TERLIHAT DARI WAJAH DAN BAHASA NYA....E	NEGATIF	Prabowo Subianto
497	Asli sama prabowo gw kecewa dan gak akan lagi per	NEGATIF	Prabowo Subianto
498	"Say NO to PENCULIK !!!"	NEGATIF	Prabowo Subianto
499	Adili Prabowo kalau dia penculik....	NEGATIF	Prabowo Subianto
500	Om Wowo terlalu ambisi padahal sdh gagal 2kali dar	NEGATIF	Prabowo Subianto
501	Wowo ga usah teriak2 ttg pengkhianatan...emang el	NEGATIF	Prabowo Subianto
502	Ganjar Pranowo itu jelas jelas melekat sebagai Bapa	NEGATIF	Ganjar Pranowo
503	Tambah yakin dan mantab pilih p Ganjar jadi preside	POSITIF	Ganjar Pranowo
504	Bukti orang Jawa Tengah cinta pak Ganjar	POSITIF	Ganjar Pranowo
505	Ganjar...ganjar...klo gini tu munafik nama ny	NEGATIF	Ganjar Pranowo
506	Seneng setiap kali Liat Pak Ganjar menyapa rakyat. G	POSITIF	Ganjar Pranowo
507	Dalam kajian ilmu PSIKOLOGI, sifat asli Ganjar itu bu	NEGATIF	Ganjar Pranowo
508	Mantab Pak Ganjar, lanjutkan terus tanpa akhir, untu	POSITIF	Ganjar Pranowo
509	Bapak ganjar. Mending jadi artis dah.. Udah ada bibit	NEGATIF	Ganjar Pranowo
510	GANJAR PRANOWO YES MANTAP PRESIDEN INDONE	POSITIF	Ganjar Pranowo
511	Terimakasih dan orang Jawa Tengah pasti sllau cinta	POSITIF	Ganjar Pranowo
512	ga cuma anak muda doang yang respect sama pak Ga	POSITIF	Ganjar Pranowo

Gambar 3. Screenshot dataset #2



Gambar 4. Screenshot implementasi code di Google Collab #1

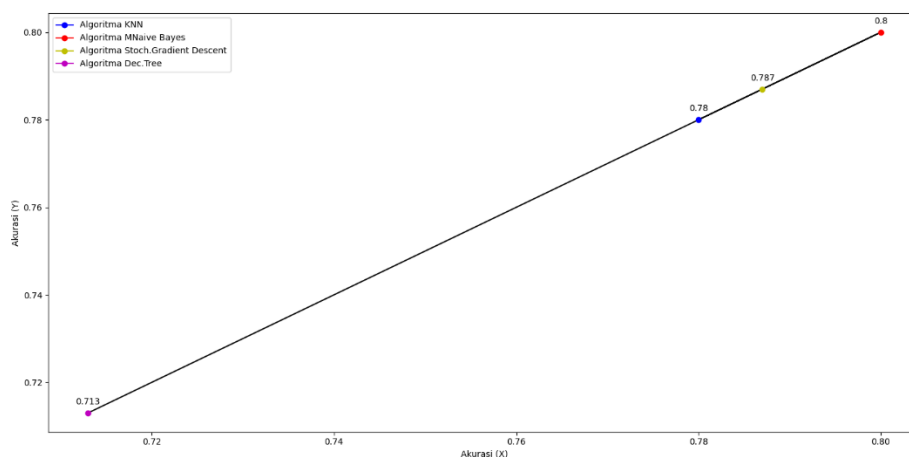
Dari hasil eksperimen dan pengujian didapatkan hasil komparasi algoritma – algoritma seperti ditunjukkan pada table 1. Berdasarkan hasil pengujian diketahui skenario terbaik ada pada skenario 1, dimana rata – rata akurasinya mencapai 75%, selisih 2% dibanding skenario ke 2 dan 3. Sedangkan algoritma terbaik adalah Multinomial Naïve Bayes yang memiliki rata – rata 79% dalam semua skenario pengujian. Pengujian ini menggunakan Confusion Matrix. Visualisasi pengujian antar algoritma dapat dilihat pada gambar 6 - 8



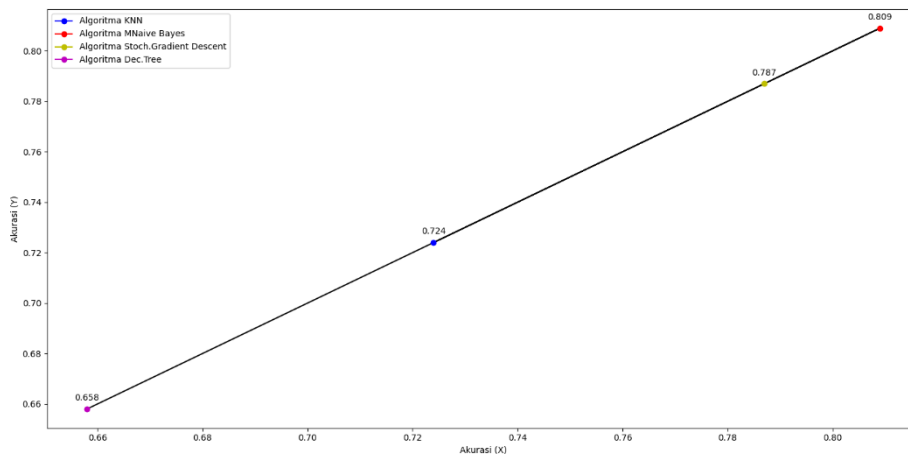
Gambar 5. Screenshot implementasi code di Google Collab #2

Tabel 1. Perbandingan Peformansi Antar Algoritma

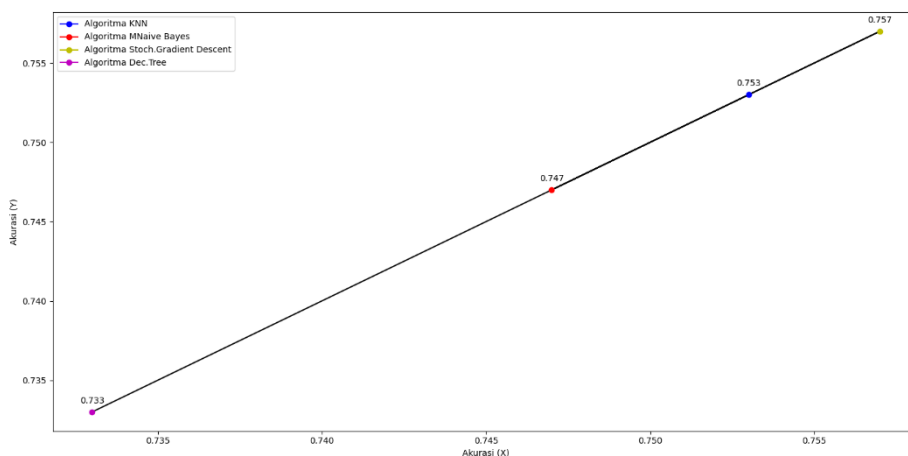
Skenario Pengujian	Akurasi KNN	Akurasi Multinomial Naïve Bayes	Akurasi Stochastic Gradient Descent	Akurasi Decision Tree	Rata – rata akurasi tiap skenario
1; data training 80% : data testing 20%	78%	80%	79%	71%	77%
2; data training 70% : data testing 30%	72%	81%	79%	66%	75%
3; data training 60% : data testing 40%	75%	75%	76%	73%	75%
Rata – rata algoritma dalam keseluruhan skenario pengujian	75%	79%	78%	70%	



Gambar 6. Visualisasi perbandingan algoritma skenario ke 1



Gambar 7. Visualisasi perbandingan algoritma skenario ke 2



Gambar 8. Visualisasi perbandingan algoritma skenario ke 3

4. KESIMPULAN

Berdasarkan pengujian yang dipaparkan pada subbab sebelumnya, menunjukkan hasil bahwa, dalam semua skenario pengujian, skenario pengujian 1: data training 80% : data testing 20% adalah skenario terbaik, dengan nilai akurasi rata - rata mencapai 75%. Sedangkan algoritma terbaik adalah Multinomial Naïve Bayes, dimana memiliki nilai rata - rata akurasi lebih tinggi dibanding yang lain dalam semua skenario pengujian yaitu 79%

UCAPAN TERIMA KASIH

Terimakasih kepada biro unit LPPM Universitas Katlik Darma Cendika (UKDC) Surabaya yang telah mendukung publikasi ini lewat dana hibah penelitian UKDC.

DAFTAR PUSTAKA

- [1] Sya'bani, M. R. F., Enri, U., & Padilah, T. N. (2022). Analisis sentimen terhadap bakal calon presiden 2024 dengan algoritme naive bayes. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 265-273.
- [2] Nardilasari, A. P., Hananto, A. L., Hilabi, S. S., Tukino, T., & Priyatna, B. (2023). Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter. *JOINTECS (Journal of Information Technology and Computer Science)*, 8(1), 11-18.

- [3] Nugroho, A., & Zy, A. T. (2023). Penerapan Algoritma Naive Bayes dan PSO pada Analisis Sentimen Kandidat Calon Presiden 2024. REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer, 7(3), 1367-1380.
- [4] Misrun, C. A., Haerani, E., Fikry, M., & Budianita, E. (2023). Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier. Jurnal CoSciTech (Computer Science and Information Technology), 4(1), 207-215.
- [5] Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis sentimen terhadap layanan indihome berdasarkan twitter dengan metode klasifikasi support vector machine (SVM). Jurnal Media Informatika Budidarma, 4(3), 650-658.
- [6] Jatmiko, H. B., Kurniadi, N. T., & Maulana, D. (2022). Optimasi Naïve Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta. Journal Automation Computer Information System, 2(1), 22-30.
- [7] Arsi, P., Wahyudi, R., & Waluyo, R. (2021). Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 5(2), 231-237.
- [8] Pintoko, B. M., & Lhaksana, K. M. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier. eProceedings of Engineering, 5(3).
- [9] Sutoyo, I. (1). IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI DATA PESERTA DIDIK. Jurnal Pilar Nusa Mandiri, 14(2), 217-224. <https://doi.org/10.33480/pilar.v14i2.70>
- [10] Charibaldi, N., Harfiani, A., & Simanjuntak, O. S. (2024). Comparison of the Effect of Word Normalization on Naïve Bayes Classifier and K-Nearest Neighbor Methods for Sentiment Analysis. Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, 9(1), 25-31.