

PERANCANGAN MODEL SENTIMEN TWEET TERHADAP PILKADA DKI JAKARTA TAHUN 2017 MENGGUNAKAN ALGORITMA NAÏVE BAYES

Muhammad Tegar Lazuardi, Tati Suprpti, Yudhistira Arie Wijaya

Program Studi S1 Teknik Informatika, STMIK IKMI Cirebon

Jln. Perjuangan No. 10B Majasem, Kota Cirebon

tegarlazuardi48@gmail.com, tatisuprpti112004@gmail.com, yudhistira010471@gmail.com

ABSTRAK

Pemilihan Gubernur (Pilgub) DKI Jakarta menjadi peran penting dalam menentukan wilayah politik negara. Selain mencalonkan diri kembali, pasangan calon lainnya adalah Agus-Sylvi dan Anies-Sandiaga. Pilkada DKI Jakarta 2017 merupakan momen penting bagi demokrasi Indonesia. Pada tahun 2017, 7 provinsi, dan 18 kota. Penelitian ini mengkaji jaringan mobilisasi politik yang digunakan pasangan wilayah dalam Pilkada DKI Jakarta 2017. Strategi yang menonjolkan capaian wilayah yang dapat membuat birokrasi lebih efisien dan bersih, serta kepercayaan diri dalam pemberantasan korupsi. Analisis sentimen dan opini mining adalah analisis yang bertujuan untuk melihat perasaan suatu masyarakat atau kelompok terhadap suatu keberadaan tertentu. Realita yang diekspresikan oleh masyarakat dapat bersifat positif maupun negatif. Twitter merupakan salah satu media yang membahas tentang berita seputar Pilkada DKI Jakarta 2017. Tujuan penelitian ini adalah untuk menganalisis tanggapan masyarakat terhadap pencalonan Pilkada DKI Jakarta 2017 di Twitter dan mengklasifikasikan tanggapan positif dan negatif terhadap kalimat dengan menggunakan teknik analisis. Metode yang digunakan dalam penelitian ini menggunakan teknik Naive Bayes untuk mengevaluasi. Dataset tersebut berupa tanggapan masyarakat terhadap pencalonan Pilkada DKI Jakarta di media sosial Twitter sebanyak 1129 dataset, yang kemudian dikelompokkan untuk selanjutnya diklasifikasi berdasarkan opini terkait data tersebut. Hasil penelitian ini menunjukkan akurasi 50.54% menggunakan algoritme Naive Bayes.

Kata kunci: DKI Jakarta, Partisipasi Masyarakat, Pilkada, Strategy, Sentimen, Twitter dan Klasifikasi

1. PENDAHULUAN

Text mining Merupakan algoritma machine learning dan natural language processing untuk melakukan analisis dan ekstraksi informasi dari data teks, dengan tujuan menghasilkan wawasan yang berguna dan dapat diaplikasikan dalam berbagai bidang, seperti bisnis, ilmu sosial, kesehatan, dan sebagainya dan tujuannya untuk menemukan sentimen tweet yang bisa menjawab kalimat positif dan negatif, sehingga Teknik text mining dapat membantu mengidentifikasi topik yang paling sering muncul dalam suatu data teks, seperti topik yang sering dibicarakan dalam tweet atau posting media sosial. [1]. Dengan analisis sentimen sangat cenderung dilakukan untuk melihat opini dari suatu sentimen terhadap sebuah objek dengan mengidentifikasi apakah sentimen tersebut mengandung opini positif atau negatif. Pengaruh tweet terhadap proses klasifikasi sentiment melalui proses Cleaning, selanjutnya melakukan tahapan tokenize untuk menjadikan kalimat menjadi satu potongan kata-kata [2].

Penelitian terdahulu berkaitan dengan media sosial adalah tempat komunitas dapat mengungkapkan pendapatnya. Indonesia memiliki salah satu jejaring sosial terbesar di dunia dan salah satu yang paling banyak digunakan di Indonesia adalah Facebook, twitter dan sebagainya pada tahun 2020, jumlah pengguna aktif Twitter di Indonesia meningkat dari tahun lalu atau 13 juta menjadi 166 juta menjadi pengguna aktif [3]. Penelitian lainnya yang dilakukan Winda dan kawan – kawan adalah melakukan analisa

sentimen pada opini masyarakat terhadap vaksinasi covid – 19 dengan algoritme naïve bayes dengan data yang digunakan berasal dari twitter yang didapatkan total jumlah data sebanyak 3.780 tweet. Hasil evaluasi dari penelitian ini berupa opini masyarakat yang dominan terhadap sentimen positif sebesar 2.278 (60,3%), netral 1.299 (34,4%), dan negatif 203 (5,4%). Kemudian nilai evaluasi yang didapatkan setelah menggunakan Algoritme *Naïve Bayes* yaitu 92,86% Class Recall (93%) [4].

Penelitian ini bertujuan untuk menganalisa tanggapan publik mengenai Pilkada DKI Jakarta Tahun 2017 dan menganalisa Tweet Pilkada DKI Jakarta 2017. Menganalisa tweet Pilkada DKI Jakarta 2017, menggunakan secara online, seperti Twitter dapat membantu menganalisis tweet yang berkaitan dengan Pilkada DKI Jakarta 2017. Sehingga mengklasifikasi hasil kalimat Positif dan Negatif dan mengetahui tingkat akurasi terbaik dalam penggunaan parameteres proses document dalam melakukan proses *Term Frequency – Inverse Document Frequency (TF-IDF)*, dalam menghitung bobot setiap kata akan menghasilkan tingkat akurasi yang didapat pada algoritma klasifikasi yang terbaik.

2. TINJAUAN PUSTAKA

2.1. Text Mining

Text Mining merupakan proses menganalisis dan mengekstrak wawasan dari kumpulan besar data tekstual yang tidak terstruktur. Jenis data ini dapat berasal dari berbagai sumber, seperti platform media

sosial, artikel berita, ulasan pelanggan, atau makalah penelitian. [5].

2.2. RapidMiner

RapidMiner merupakan alat penambangan data yang kuat karena menyertakan plugin yang memungkinkan penggunaan algoritme berbeda dalam alat ini dan alat lainnya dengan operator yang membantu mengembangkan proses untuk membuat model yang sesuai untuk analisis data. *RapidMiner* memakai bermacam teknik deskriptif dan prediksi sehingga dapat membuat suatu keputusan yang terbaik [6].

2.3. Analisis Sentimen

Analisis sentimen merupakan menangkap opini pada dokumen, kalimat, unit, dan aspek berdasarkan orientasi emosionalnya. Pada dokumen, komentar dianalisis dan ditentukan apakah pendapat umum tentang komunikasi kursus itu positif, negatif, atau netral. Pada tingkat kalimat, ini mengekstraksi perasaan setiap kalimat dan membantu menghitung sisi positif dan negatif dari kursus. Ekstraksi entitas tingkat sentimen menggabungkan analisis entitas dan sentimen untuk memberi siswa pemahaman tentang entitas seperti instruktur, kursus, dan tugas [7].

2.4. Media Sosial Twitter

Media Sosial Twitter adalah sebuah platform media sosial yang memungkinkan pengguna untuk berbagi pesan atau "tweet" dalam format teks dengan maksimal 280 karakter, serta foto, video, atau tautan ke situs web lainnya. Pengguna dapat mengikuti akun lain untuk melihat tweet mereka di beranda mereka, dan juga dapat menggunakan fitur "like" dan "retweet" untuk memberikan dukungan pada tweet dari pengguna lain coba dipagrapase[8].

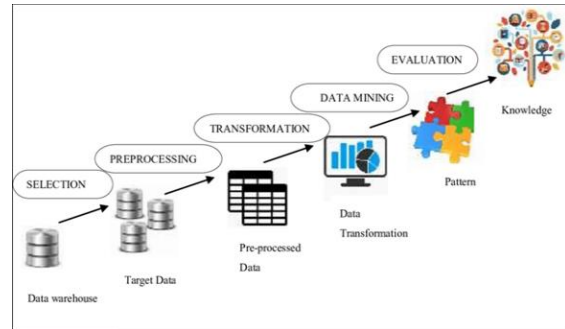
3. METODE PENELITIAN

3.1. Jenis Penelitian

Metode penelitian yang digunakan adalah pendekatan kuantitatif. Kuantitatif ini diolah dan dianalisis untuk menarik kesimpulan. Pendekatan ini mengutamakan pengukuran, pengolahan, dan analisis data secara sistematis dan obyektif untuk mendapatkan kesimpulan yang akurat dan dapat dipercaya.

3.2. Metode Penelitian

Tahapan dalam penelitian ini menggunakan metode analisa *Knowledge Discovery In Database* (KDD). Tahapan analisa data bisa dilihat pada gambar berikut:



Gambar 1. Tahapan Knowledge Discovery in Database (KDD)

Pada penelitian ini, penulis menggunakan metode atau tahapan KDD singkatan dari Knowledge Discovery in Database. Berikut merupakan tahapan – tahapan dan proses yang ada pada Knowledge Discoverey in Database:

3.3. Selection

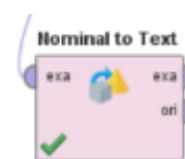
Di awal tahapan terdapat tahap selection, yang mana data selection adalah proses pemilihan subset data dari kumpulan data yang lebih besar berdasarkan kriteria atau parameter tertentu. Ini sering dilakukan untuk fokus pada aspek data tertentu atau untuk mengurangi jumlah data yang perlu dianalisis atau diproses. Pemilihan data mendahului praktik pengumpulan data yang sebenarnya. Data diambil dari Kaggle.com dengan kata kunci data Pilkada DKI Jakarta tahun 2017. dari tanggal 22-23 Desember 2021 sebanyak 1129 twett

3.4. Data Cleaning atau Pre-processing

Pembersihan data untuk proses mengidentifikasi bagian data yang tidak lengkap, salah, benar, tidak akurat atau hilang dan mengoreksi, mengganti atau menghapusnya. Cleansing penelitian ini akan menghapus @, #, karakter URL serta atribut yang hanya menggunakan teks. Operator yang digunakan adalah substitusi.

3.5. Transformation

Dalam analisis sentimen, data transformasi untuk operator nominal to text berguna untuk Operator nominal to text ini biasanya diterapkan pada variabel target, yaitu variabel yang akan dianalisis sentimennya. Dalam kasus analisis sentimen pada teks, variabel target dapat berupa tweet, postingan di media sosial, atau teks lainnya.



Gambar 2. Operator Nominal to text

3.6. Data Mining

Data mining tidak secara khusus mengubah teks yang tidak terstruktur menjadi format terstruktur untuk mengidentifikasi pola. Namun, teknik text mining dapat digunakan untuk mengubah teks yang tidak terstruktur menjadi format terstruktur yang dapat dianalisis oleh teknik data mining yang bermakna dan wawasan baru. Pada tahap Riset ini menerapkan Naïve Bayes.

3.7. Pattern Evaluation

Pattern Evaluation didefinisikan sebagai mengidentifikasi model yang meningkatkan wawasan berdasarkan ukuran yang diberikan. Temukan tempat menarik untuk setiap model menggunakan ringkasan dan visualisasi agar informasi dapat dipahami oleh pengguna

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Membaca dataset dalam bentuk *file excel*, menggunakan operator *Read Excel* seperti tampak pada gambar dibawah ini.



Gambar 3. Parameters Read Excel

Parameter pada operator *Read Excel*, menggunakan parameter *default*. Dari hasil pembacaan operator *Read Excel* didapat informasi sebagai berikut.

Tabel 1. Statik dataset

No.	Uraian	Keterangan
1.	Record	104
2.	Special Attribute	0
3.	Reguler Attribute	0
4.	Attribute :	
	ID	Integer, missing 0
	Sentiment	Polynomial dan label
	Pasangan Caslon	Polynomial
	Texttweet	Polynomial

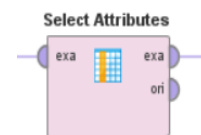
Data Selection ini didapat dari webeset pada *Kaggle.com* dengan operator *search twitter* dan kata kunci data tweet twitter Pilkada DKI Jakarta 2017. Pengambilan data dari tanggal 22 sampai 23 Desember 2022.

Row No.	Sentiment	Id	Pasangan C...	Text Tweet
1	negative	1	Agus-SyM	Banyak akun kloning seolah2 pendukung
2	negative	2	Agus-SyM	?
3	negative	3	Agus-SyM	Kalau situ sih gak nungga hasil ahir QC tp lagi nungga mofit outan pak @SEYudhyono kayak apa... pasca
4	negative	4	Agus-SyM	Kasian ah kasian dengan peluru 1mlir untuk kap RIV
5	negative	5	Agus-SyM	Maaf ya pendukung
6	negative	6	Agus-SyM	aneh deh lebay masa ini di sangkut pautkan sama kandidat Calgub, anda lebay seperti yg anda dukung
7	negative	7	Agus-SyM	Kepada Allah SWT kami ucapkan rasa syukur dan kepada pak @SEYudhyono @presidenSBY
8	negative	8	Agus-SyM	Terima Kasih teruntuk pendukung
9	negative	9	Agus-SyM	Terima Kasih atas Kelidipomahan Relawan
10	negative	10	Agus-SyM	Mengenang pidato kekalahan
11	negative	11	Agus-SyM	dammnn!! politik cantik
12	negative	13	Agus-SyM	sudah boleh ngakak? survey mu jauh panggang dari api!
13	negative	14	Agus-SyM	aku juga kemarin
14	negative	15	Agus-SyM	Karna yang berlebihan itu tidak baik ya bah
15	negative	17	Agus-SyM	kallau pencatutan nama
16	negative	18	Agus-SyM	ya pastilah
17	negative	19	Agus-SyM	dia lah Bambang... dia lah Gusdafi klatra seperti bpk @garabowo genile spt
18	negative	20	Agus-SyM	Garaz e-KTP dikotopi, tobiu yg semua mimp

Gambar 4. Dataset Pilkada DKI Jakarta 2017

4.2. Operator Select Attributes

Operator *select attributes* ini berfungsi untuk memilih data subset atribut yang akan digunakan dari *ExampleSet* dan menghapus atribut lainnya yang tidak digunakan. Tampilan operator Select Attributes dapat dilihat pada gambar 4



Gambar 5. Operator Select Attributes

Pada operator *select attributes* dilakukan proses *selection* terhadap atribut yang akan digunakan dalam penelitian dapat dilihat pada tabel 2

Tabel 2. Tabel Atribut Data

No	Atribut yang digunakan
1	ID
2	Sentiment
3	Pasangan Calon
4	Text Tweet

Semua Atribut ini yang tdigunakan dalam penelitian ini adalah atribut Id, Sentimen, Pasangan Calon, dan Text tweet. Atribut tersebut tidak mempengaruhi proses data mining yang dijadikan nilai untuk menghitung akurasi dan digunakan di dalam penentu pemilihan kepala daerah DKI Jakarta tahun 2017. Hasil yang diperoleh dari proses operator *select attributes* ini dapat dilihat pada gambar 6

Row No.	Sentiment	Id	Pasangan C...	Text Tweet
1	negative	1	Agus-SyM	Banyak akun kloning seolah2 pendukung
2	negative	2	Agus-SyM	?
3	negative	3	Agus-SyM	Kalau situ sih gak nungga hasil ahir QC tp lagi nungga mofit outan pak @SEYudhyono kayak apa... pasca
4	negative	4	Agus-SyM	Kasian ah kasian dengan peluru 1mlir untuk kap RIV
5	negative	5	Agus-SyM	Maaf ya pendukung
6	negative	6	Agus-SyM	aneh deh lebay masa ini di sangkut pautkan sama kandidat Calgub, anda lebay seperti yg anda dukung
7	negative	7	Agus-SyM	Kepada Allah SWT kami ucapkan rasa syukur dan kepada pak @SEYudhyono @presidenSBY
8	negative	8	Agus-SyM	Terima Kasih teruntuk pendukung
9	negative	9	Agus-SyM	Terima Kasih atas Kelidipomahan Relawan
10	negative	10	Agus-SyM	Mengenang pidato kekalahan
11	negative	11	Agus-SyM	dammnn!! politik cantik
12	negative	13	Agus-SyM	sudah boleh ngakak? survey mu jauh panggang dari api!
13	negative	14	Agus-SyM	aku juga kemarin
14	negative	15	Agus-SyM	Karna yang berlebihan itu tidak baik ya bah
15	negative	17	Agus-SyM	kallau pencatutan nama
16	negative	18	Agus-SyM	ya pastilah
17	negative	19	Agus-SyM	dia lah Bambang... dia lah Gusdafi klatra seperti bpk @garabowo genile spt
18	negative	20	Agus-SyM	Garaz e-KTP dikotopi, tobiu yg semua mimp

Gambar 6. Hasil Select Attributes

4.3. Data Labeling

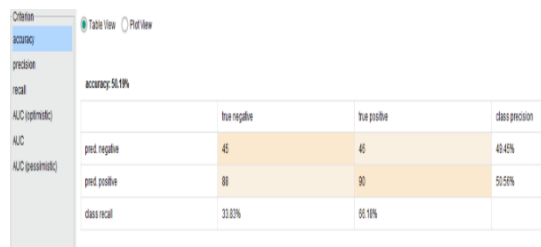
Data diberi label secara manual sebagai sentimen positif dan negatif. Label adalah cara yang bagus untuk mengomunikasikan informasi seperti nama lokasi. Pelabelan adalah bagian dari pembelajaran yang diawasi. Supervised learning adalah model pembelajaran mesin yang memeriksa data dengan tag atau target dimana evaluasi model akan didasarkan pada target tersebut.

Tabel 3. Label Proses

Tweets	Sentimen
memang kalah di pilkada tetapi MENANG di hati rakyat."	Positif
Agus-Sylvi#agussilvy bicara apa kasihan yaa...lap itu air matanya wkwkwkwk"	Negatif

4.4. Naïve Bayes

Penilaian Kinerja Hasil Klasifikasi Algoritme Naive Bayes:



accuracy	50.19%
precision	
recall	
AUC (optimal)	
AUC	
AUC (psymetric)	

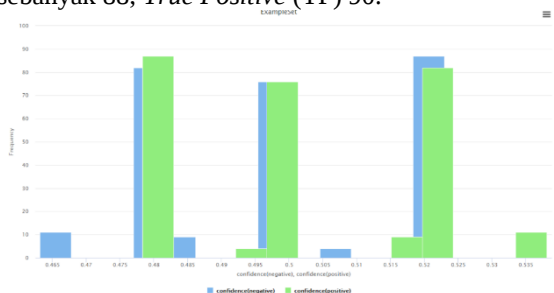
	true negative	true positive	class precision
pred negative	45	45	49.45%
pred positive	88	90	50.56%
class recall	33.83%	90.18%	

Gambar 7. Performance Naïve Bayes

Berdasarkan Hasil *Performance* tentang analisa sentimen Pilkada DKI Jakarta tahun 2017 pada *twitter* yang menggunakan metode *Naïve Bayes* diperoleh hasil akurasi sebesar 50.19% dan *micro average* 49.45% dalam penentuan sentimen memiliki tingkat *precision* sebesar 50.56% dan *recall* sebesar 66.18%, dimana penjelasan tabel sebagai berikut:

Pertama, sebesar 33.83% untuk *predicted class negative* dengan *True Negative* (TN) sebanyak 45, *True Positive* (TP) sebanyak 46.

Kedua, sebesar 66.18% untuk *predicted class positive* dengan *True Negative* (TN) sebanyak 45, *True Positive* (TP) sebanyak 45. *True Negative* (TN) sebanyak 88, *True Positive* (TP) 90.



Gambar 8. Bagan Naïve Bayes

Pembahasan hasil analisis kualitatif dan/atau kuantitatif harus dilakukan dengan tujuan untuk memberikan jawaban atas permasalahan yang diteliti

dalam makalah. Pada bagian ini, seluruh kegiatan yang telah dilakukan dalam penelitian harus diuraikan secara lengkap. Mulai dari konsep penelitian, perancangan penelitian, hipotesis (bila ada), percobaan, pengamatan data, dan hasil dari data pengamatan yang telah dilakukan.

4.5. Pattern Evaluation

Pattern evaluation proses evaluasi pola dimulai dengan mengumpulkan data dari berbagai sumber dan memperhatikan setiap detailnya. Data yang telah dikumpulkan kemudian dianalisis untuk mencari pola dan tren yang menonjol. Setelah itu, pola dan tren tersebut dievaluasi untuk menentukan apakah mereka memiliki nilai yang signifikan atau tidak. Evaluasi dilakukan setelah proses data dilakukan dengan *Naïve Bayes*. menghasilkan berapa banyak sentimen *positive* dan *negative*. *Confusion Matrix* digunakan dalam evaluasi ini karena berfungsi sebagai alat evaluasi dalam memperkirakan objek yang benar (*True*) dan salah (*False*) pada model klasifikasi. *Matrix* sendiri digunakan dalam membandingkan hasil klasifikasi dengan data yang sebenarnya.

Tabel 4. Tabel Confusion Matrix Prediksi Dua Kelas

Predicted Class		Actual Class	
		Class 1	Class 2
	Class 1	True Positive (TP)	False Negative (FN)
	Class 2	False Positive (FP)	True Negative (TN)

Rumus yang digunakan untuk menghitung akurasi adalah:

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

Rumus yang digunakan untuk menghitung nilai precision:

$$Precision = \frac{TP}{TP + FP}$$

Rumus yang digunakan dalam menghitung nilai Recall adalah:

$$Recall = \frac{TP}{TP + FN}$$

Hasil dari algoritme yang digunakan menunjukkan akurasi dari *Naïve Bayes*. Menghasilkan akurasi sebesar 50%, precision 33% dan recall 49% pada algoritme *Naïve Bayes*. Sehingga dapat disimpulkan dari hasil diatas bahwa pada penelitian ini klasifikasi terbaik dalam menghasilkan tingkat akurasi

5. KESIMPULAN DAN SARAN

Melalui metode *Naïve Bayes*, data-data dari *Twitter* tersebut diproses untuk menghasilkan informasi tentang sentimen masyarakat terkait

pemilihan kepala daerah DKI Jakarta. Metode ini didasarkan pada teori probabilitas dan menggunakan klasifikasi untuk menentukan sentimen yang diungkapkan oleh masyarakat dalam setiap tweet. melalui media sosial Twitter yang mengklasifikasikannya menjadi positif dan negatif, hasil akurasi yang didapat pada algoritme Naïve Bayes sebesar 50%, precision 33% dan recall 49%. Peneliti menggunakan metode klasifikasi metode Naïve Bayes karena metode yang tepat digunakan dengan hasil akurasi baik dalam mengklasifikasikan data mining. Dalam studi selanjutnya, lebih banyak set data harus digunakan daripada studi ini untuk meningkatkan informasi tentang pilkada Dki Jakarta tahun 2017 dan meningkatkan presisi dibandingkan studi sebelumnya. Gunakan metode lain untuk memverifikasi bahwa proses penelitian itu akurat. Pencarian tambahan juga dapat mencari sumber data lain seperti Facebook, Instagram, dan YouTube.

DAFTAR PUSTAKA

- [1] H. P. Doloksaribu and Yusran Timur Samuel, "Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi," *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 16, no. 1, pp. 1–11, 2022, doi: 10.47111/jti.v16i1.3747.
- [2] B. Darma, A. Suryani, and Y. Surono, "Pengaruh Fasilitas Kerja Dan Kompetensi Terhadap Kepuasan Kerja Pegawai Dan Dampaknya Terhadap Kinerja Pegawai Kantor Kelurahan Sekecamatan Muara Bulian Kabupaten Batang Hari," *J-MAS (Jurnal Manaj. dan Sains)*, vol. 3, no. 2, p. 170, 2018, doi: 10.33087/jmas.v3i2.54.
- [3] S. R. I. Rezeki, "Penggunaan sosial media twitter dalam komunikasi organisasi (studi kasus pemerintah provinsi dki jakarta dalam penanganan covid-19)," *J. Islam. Law Stud.*, vol. 04, no. 02, pp. 63–78, 2020.
- [4] I. Sholekha, A. Faqih, and A. Bahtiar, "Sentiment Analysis of Public Opinion Covid-19 Vaccine Using Naïve Bayes and Random Forest Methods," *J. Tek. Inform.*, vol. 15, no. 1, pp. 34–43, 2022, doi: 10.15408/jti.v15i1.24847.
- [5] C. Joergensen E Munthe, N. Astuti Hasibuan, and H. Hutabarat, "Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace," *Resolusi Rekayasa Tek. Inform. dan Inf.*, vol. 2, no. 3, pp. 110–115, 2022, doi: 10.30865/resolusi.v2i3.309.
- [6] A. Vilorio *et al.*, "Determinating student interactions in a virtual learning environment using data mining," *Procedia Comput. Sci.*, vol. 155, no. 2018, pp. 587–592, 2019, doi: 10.1016/j.procs.2019.08.082.
- [7] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.