

Analisis Media *Text Clustering* pada Twitter Akan Kasus Selebriti Menggunakan *Orange Data Mining*

¹*Andaru Liken Anggoro, ²Lulasnov Viola Prameswari Ken, ³Majid Gigih Setiawan.

^{1,2,3}Institut Teknologi Telkom Purwokerto
Purwokerto, Indonesia

¹18103097@ittelkom-pwt.ac.id, ²218103110@ittelkom-pwt.ac.id, ³318103111@ittelkom-pwt.ac.id

*Penulis Korespondensi

Diajukan : 14/12/2022

Diterima : 24/12/2022

Dipublikasi : 02/01/2023

ABSTRAK

Media sosial saat ini tidak hanya digunakan sebagai alat bantu komunikasi jarak jauh namun juga menjadi salah satu wadah yang digunakan oleh masyarakat untuk berbagi, mengakses, dan diskusi seputar isu – isu publik saat ini. Salah satunya media sosial yang dapat digunakan sebagai wadah tersebut adalah twitter yang merupakan platform media *online real time* yang sering digunakan untuk berkomunikasi dengan sebutan ciutan – ciutan antar satu pengguna dengan yang lain. Trending topik pada twitter berisi komentar masyarakat Indonesia yang mengandung statement opini. Salah satu contoh kecilnya yaitu kasus artis selebriti yang tersandung permasalahan mengenai “kabur karantina” menjadi salah satu topik yang sangat menarik untuk dibahas. Penelitian ini mencoba menganalisis tanggapan masyarakat melalui komentar di twitter dengan metode analisis vader kemudian dari analisis tersebut menghasilkan *tweet profiler* dan juga *visualition distribution*. Penelitian ini menggunakan aplikasi penambangan data *orange data mining* dengan mengaplikasikan *preprocess text* yang meliputi *transformation, tokenization, normalization, dan filtering* yang bertujuan agar *text* dapat dianalisis. Melalui analisis sentimen dengan metode *text clustering* di twitter kita dapat memahami bagaimana orang menggambarkan dan mengungkapkan persepsi mereka terhadap kasus yang akan di bahas baik secara positif, negatif maupun netral. Analisis tersebut penting untuk melihat sejauh mana media sosial seperti twitter digunakan oleh masyarakat saat ini sebagai salah satu instrumen diseminasi informasi publik di Indonesia.

Kata Kunci: *Orange Data Mining*, Penambangan Data, Visualisasi Data, Analisis, dan Sentimen Analisis.

I. PENDAHULUAN

Media sosial saat ini menjadi salah satu kebutuhan utama yang tidak bisa dilepaskan dari kehidupan sehari-hari. Banyak sekali media sosial yang kini diminati masyarakat, salah satunya yaitu Twitter. Twitter merupakan jenis media sosial microblogging yang memfasilitasi pengguna untuk menulis dan mempublikasikan aktivitas atau pendapatnya. Melalui Twitter, pengguna bisa berbagi keseharian seperti memposting foto ataupun menyampaikan opini mengenai satu hal. Trending topik di Twitter selalu menjadi perbincangan hangat di masyarakat. Penyampaian opini di Twitter dapat menjadi salah satu media untuk di jadikan objek penelitian (Alizah et al., 2020). Platform Twitter kini diramaikan dengan beredarnya berita bahwa seorang *public figure* melarikan diri dan tidak mengikuti prosedur karantina covid-19 usai berpergian dari luar negeri. Hal tersebut memincu banyak presepsi di kalangan masyarakat, pasalnya dalam karantina covid-19 merupakan salah satu kepentingan untuk bersama.

Beberapa penelitian terdahulu yang dilaksanakan oleh Dony Jacarria Pangestu1 dan Achmad

Koda pada tahun 2022, penelitian tersebut meneliti tentang memprediksi sentimen pelayanan perusahaan otobus menggunakan algoritma *Multinomial Naive Bayes* dengan teks *mining* data facebook. Hasil dari penelitian tersebut menghasilkan tingkat akurasi prediksi untuk *precision* dan *recall* sebesar 95%, dengan hasil tersebut algoritma *Multinomial Naive Bayes* memiliki tingkat *fair classification* dan nilai *diagnostic* yang baik (Jacarria Pangestu & Kodar, 2022).

Penelitian lainnya yang dilaksanakan oleh Dedi Darwis, Nery Siskawati, dan Zaenal Abidin meneliti tentang klasifikasi *tweet* tentang BMKG berdasarkan komentar negatif, positif, dan netral dengan menggunakan algoritma *naive bayes*. Hasil yang didapat pada penelitian ini yaitu didapatkan tingkat akurasi klasifikasi *tweet* sebesar 68,97% (Darwis et al., 2021).

Penelitian selanjutnya yang dilaksanakan oleh Lila Setiyani, Mokhammad Wahidin, Dudi Awaludin, dan Sri Purwan meneliti mengenai prediksi jumlah kelulusan tepat waktu mahasiswa menggunakan metode *naive bayes*. Hasil penelitian ini mendapatkan hasil akurasi prediksi sebesar 90% (Setiyani et al., 2020). Pada dasarnya, penelitian ini mencoba menganalisis komentar dengan metode Vader, yaitu metode analisis lexicon-based berbasis rule – based sentiment analysis. Vader akan menganalisis text yang menghasilkan class sentiment berupa positif, negatif, neutral dan compound (Indah Nurhafida & Sembiring, 2021). Dalam penelitian ini memanfaatkan preprocess text dalam Orange Data Mining untuk mengenali tipe konten dari beberapa informasi *tweet*. dengan mengaplikasikan teknik clustering.

II. STUDI LITERATUR

Penelitian Terdahulu

Beberapa penelitian terdahulu yang dilaksanakan oleh Dony Jacarria Pangestu¹ dan Achmad Koda pada tahun 2022, penelitian tersebut meneliti tentang memprediksi sentimen pelayanan perusahaan otobus menggunakan algoritma *Multinomial Naive Bayes* dengan teks *mining* data facebook. Hasil dari penelitian tersebut menghasilkan tingkat akurasi prediksi untuk *precision* dan *recall* sebesar 95%, dengan hasil tersebut algoritma *Multinomial Naive Bayes* memiliki tingkat *fair classification* dan nilai *diagnostic* yang baik (Jacarria Pangestu & Kodar, 2022).

Penelitian lainnya yang dilaksanakan oleh Dedi Darwis, Nery Siskawati, dan Zaenal Abidin meneliti tentang klasifikasi *tweet* tentang BMKG berdasarkan komentar negatif, positif, dan netral dengan menggunakan algoritma *naive bayes*. Hasil yang didapat pada penelitian ini yaitu didapatkan tingkat akurasi klasifikasi *tweet* sebesar 68,97% (Darwis et al., 2021).

Penelitian selanjutnya yang dilaksanakan oleh Lila Setiyani, Mokhammad Wahidin, Dudi Awaludin, dan Sri Purwan meneliti mengenai prediksi jumlah kelulusan tepat waktu mahasiswa menggunakan metode *naive bayes*. Hasil penelitian ini mendapatkan hasil akurasi prediksi sebesar 90% (Setiyani et al., 2020).

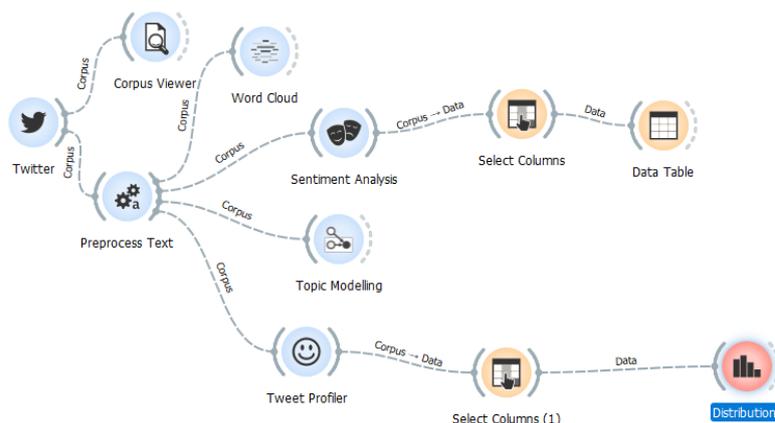
III. METODE

Metode yang digunakan dalam penelitian ini yaitu metode *Preprocess Text* yang meliputi *transformation*, *tokenization*, *normalization*, dan *filtering* yang bertujuan agar *text* bisa dianalisis dapat dilakukan dengan mengintegrasikan *Website Scraping* serta *tools Orange Information Mining*. Untuk memudahkan mengenali tipe konten dari beberapa informasi *tweet*, Dalam penelitian ini memanfaatkan *tools preprocess text* dalam Orange Data Mining untuk mengenali tipe konten dari beberapa informasi *tweet* dengan mengaplikasikan teknik *clustering* (Rifai et al., 2021).

IV. HASIL DAN PEMBAHASAN

Skenario Penelitian

Berikut tampilan design widget text clustering dengan menggunakan Orange Data Mining seperti gambar di bawah ini.



Gambar 1 Skenario Penelitian

Data yang dicrawling dari twitter kemudian diinput dan dianalisis satu persatu berdasarkan objek. Kemudian akan dihubungkan pada widget sesuai dengan apa yang akan digunakan pada penelitian sehingga menghasilkan widget seperti gambar tersebut (Hayuningtyas & Sari, 2019).

Crawling Data (Twitter)

Data yang diambil pada penelitian ini adalah *trending topic tweet* masyarakat Indonesia di twitter. Diambil dari 1000 *tweet* dengan kata kunci Rachel Venny, karantina, dan kabur dari API dan *token* twitter. Dalam Orange Data mining data inputan bisa disebut dengan *corpus*. *Widget corpus* yaitu kumpulan dokumen yang bisa menyajikan jumlah baris kalimat, serta memastikan mana fitur yang akan dan tidak akan diinput untuk keperluan analisis.

Preprocess Text

Sebelum melaksanakan analisis text, pelaksanaan *text mining* untuk mengolah *text* dalam perihal ini merupakan *preprocess text*. *Text* dipisahkan menjadi unit yang lebih kecil (*token*), kemudian *transformation*, *tokenization*, *normalization*, serta *filtering*. Tahapan *preprocess text* telah dilakukan, kemudian data telah berbentuk *text-text* terpisah dan dapat dilihat dalam bentuk *word cloud* pada Orange Data Mining dalam *widget Preprocess Text*. Di bawah ini tahapan dalam *preprocess text* dilakukan sebelum *text* dianalisis memakai Orange Data Mining (Yulianita et al., 2020) :

Transformation

Pertama yang dilakukan dari *preprocess text* adalah *transformation* merupakan proses mengubah data input untuk transformasi huruf kecil secara *default*.

Tokenization

Tokenisasi adalah metode memecah teks menjadi komponen yang lebih kecil (kata, kalimat, bigrams). Meliputi: *Word* dan *Punctuation* akan membagi teks berdasarkan kata dan tetap membiarkan *symbol* tanda baca (tidak menghilangkannya); contoh: *This Sample. (This), (sample), (.)*. *Tweet*, yang akan membagi teks dengan model Twitter *pra-trained*, yang memuat *hashtag*, *emoticons* dan simbol khusus lainnya. Contoh: *This words. :-) #simple* → *(This), (Words), (.), (:-), (#simple)*. Pada dasarnya *Word* dan *Punctuation* serta *tweet* memiliki kesamaan sifat proses, akan tetapi *word* dan *punctuation* menjadi proses utama dalam *tokenization*. *Word* dan *punctuation* juga digunakan untuk analisis tren.

Normalization

Proses selanjutnya adalah *normalization* yang berlaku *stemming* dan *lemmatization text*. *Text* yang telah dipisahkan kata demi kata akan menghasilkan sebuah *text* yang berdiri sendiri dalam sebuah kalimat. Konten status serta pendapat umumnya memiliki penyusunan yang kurang sempurna (*typo*). Diharapkan dengan proses normalisasi ini *text* tersebut akan dikenal maknanya dengan memakai *WordNet Lemmatizer* mempraktikkan jaringan persamaan kata (*sinonim*) kognitif buat *token* (kata) bersumber pada pada basis informasi *lexicon* (kamus) bahasa Indonesia yang besar dari NLTK (*Alami Language Toolkit*).

yang terdaftar dalam *lexicon vader*. Hasil akhir dari evaluasi merupakan skor total ialah *compound*. Skor total inilah yang hendak direkapitulasi serta dibanding hasilnya. Dalam menganalisis terdapat sebagian atribut *variable* yang hendak difokuskan dimana dari fitur yang jadi atribut yang digunakan untuk dianalisis (*used features*) dari *widget corpus* ialah kolom Texttrans serta ialah *text* hasil *translate* yang ada pada tiap informasi kementar, yang tujuannya buat memperoleh hasil berbentuk atribut positif negatif, neutral, serta skor total (*compound*).

Tabel Data

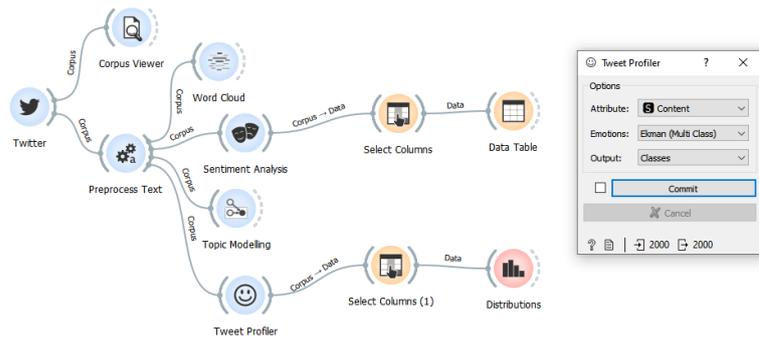
Tabel data juga merupakan reaksi hasil untuk menampilkan atribut mana saja yang dipilih untuk ditampilkan sebagai *output* dalam *widget select column* pada langkah sebelumnya. Dengan bantuan *data crawling twitter* menggunakan API hasil analisis pada tabel data hasil analisis sentimen akan dihitung seberapa positif, negatif, dan neutral dengan melihat dari skor total (*Compound*) dengan formula perhitungan format data numerik.

ID	Content	Date	Language	Location	Number of Likes	Number of Retweets	In Reply To	Author Name	Author Description	Number of Statuses	Number of Favourites	Number of Friends	Number of Followers
94	"Aku akan men..."	2022-01-17 05:3...	in	?	0	0	?	One Piece man...	Walaupun Hanu...	134385	1	17	
163	"Kam-kaminyaa..."	2022-01-17 05:1...	in	?	0	0	?	神楽 (Hatus)	KAMIREI's RP a...	121990	0	73	
144	"Kegiatan ini m..."	2022-01-17 05:2...	in	?	0	0	?	@karantinacleg	Karantina Perla...	2137	1614	41	1
1348	"Nanti aku ma..."	2022-01-16 16:4...	in	?	0	0	?	@conanbook	we are Conaner...	169547	14	24	
1611	"Random hug" ...	2022-01-16 15:2...	in	?	0	0	?	ate k/via	was @taeatelo...	4001	449	931	1
282	"hahaha gak du..."	2022-01-17 04:3...	in	?	1	0	?	* 0	KEKEY : The y...	10088	332	215	
129	"karantina hanu..."	2022-01-17 05:2...	in	?	0	0	?	@karantinacleg	Karantina Perla...	2137	1614	41	1
1212	"udah sekarang..."	2022-01-16 18:0...	in	?	0	0	?	98liner	I've found a nat...	14080	10960	301	
1345	"# klo ketemu ..."	2022-01-16 16:4...	in	?	4	2	?	@zoeyms	Zoey_oung	8775	284	1655	1
259	"#ChenZheyuan..."	2022-01-17 04:4...	in	?	16	5	?	ig: @dramachi...	MANDARIN PO...	3556	3764	110	3
145	"#logga Sapka..."	2022-01-17 05:2...	in	?	0	0	?	den Mitro	Hai & follow aj...	17507	29	6	
134	"#Robingya #Eh..."	2022-01-17 05:2...	in	?	1	1	?	Serambi Indone...	Official Twitte...	271423	27	945	95
481	"-- Pate pur..."	2022-01-17 03:2...	in	?	0	0	?	ꦒꦏ꧀	Make your hear...	21371	12454	308	
749	"(Panggilan lisan...)"	2022-01-17 01:2...	in	?	0	0	?	Jieqiong	(RP) A beauty q...	11265	19	221	
925	"(karantina 10 h..."	2022-01-17 00:0...	in	?	0	0	?	cici, S.Kd	#staykids #eve...	27694	32382	1624	
1405	"(nahan nangisi)"	2022-01-16 16:2...	in	?	0	0	?	@aculterasi	mia	46967	79639	331	
432	"Anggota Unit ..."	2022-01-17 03:4...	in	?	1	0	?	HUMAS PULOG...	humas Polsek P...	3713	1049	2	
663	"KABUR DRI RIL..."	2022-01-17 02:1...	in	?	0	0	?	@calontoyool	syie	1406	126	30	
185	"#busaha me..."	2022-01-17 05:0...	in	?	0	0	?	Ha 0 000	#TREASURESEC...	16693	17169	1154	1
261	"#MY bundaaa ..."	2022-01-17 04:4...	in	?	63	6	?	ꦒꦁꦤꦶꦁꦢꦤ...	// not personal ...	226006	13	1517	86
1985	"#myb downoon ..."	2022-01-16 14:2...	in	?	41	1	?	ꦒꦁꦤꦶꦁꦢꦤ...	// not personal ...	226006	13	1517	86
923	"-tk @Stray_Kid..."	2022-01-17 00:0...	in	?	137	12	?	Story Kids	Indonesian STA...	493556	126	4199	95
654	"-tk bracha kaba..."	2022-01-17 02:1...	in	?	43	0	?	Story Kids	Indonesian STA...	493556	126	4199	95
976	"-tk @Stray_Kid..."	2022-01-16 23:5...	in	?	390	31	?	Story Kids	Indonesian STA...	493556	126	4199	95
930	"-tk @Stray_Kid..."	2022-01-17 00:0...	in	?	38	0	?	Story Kids	Indonesian STA...	493556	126	4199	95
23	"-tk mau nanya..."	2022-01-17 06:0...	in	?	5	0	?	Story Kids	Indonesian STA...	493556	126	4199	95
310	"-tk stay anak a..."	2022-01-17 04:2...	in	?	20	0	?	Story Kids	Indonesian STA...	493556	126	4199	95
646	"-jangan ngasal..."	2022-01-17 02:1...	in	?	0	0	?	Feddy Goenawan	Your Financial ...	994	6	1074	

Gambar 4 Tabel Hasil Data

Tweet Profiler

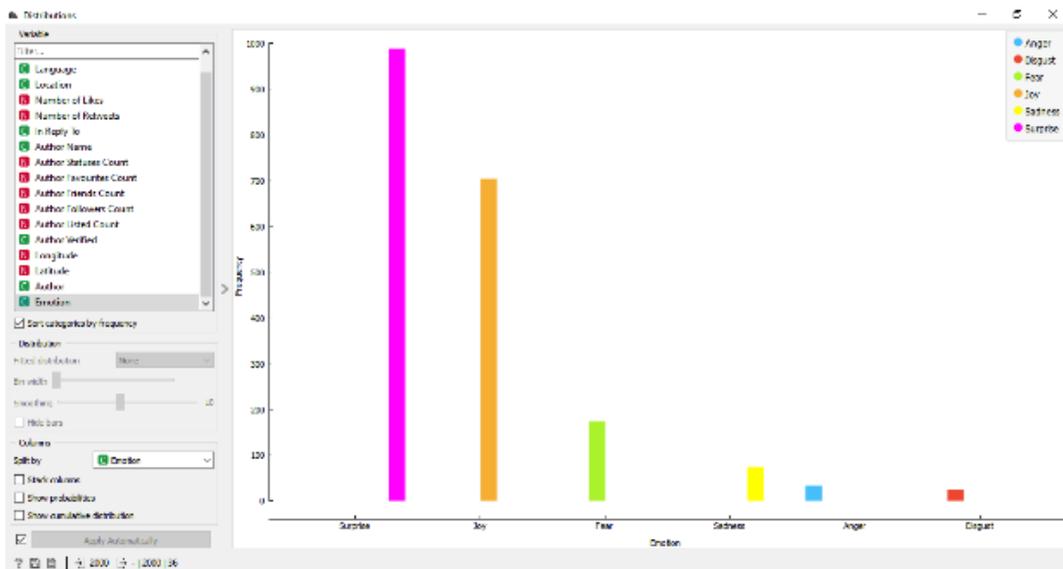
Tweet Profiler mengambil data tentang sentimen dari *server* untuk setiap *tweet* (ataupun dokumen) yang diberikan. *Widget* mengirimkan informasi ke *server*, tempat model menghitung probabilitas serta/ ataupun skor emosi. *Widget* menunjang 3 klasifikasi emosi, ialah Ekman, Plutchik serta *Profile of Mood States* (POMS). Klasifikasi emosi terdapat 3 kategori yaitu Ekman, Plutchik ataupun *Profile of Mood States*. Klasifikasi kelas jamak akan menghasilkan satu emosi yang sangat bisa jadi per dokumen, sedangkan banyak label akan menciptakan nilai dalam kolom untuk tiap emosi. Penelitian ini akan menggunakan atribut konten untuk analisis, klasifikasi emosi *Ekman* dengan opsi multi kelas dan memilih untuk mengamati variabel Emosi yang telah dikelompokkan dengan orange data mining. Pada penelitian ini menggunakan data 2000 tweets mengenai Rachel Vennya, karantina, dan kabur. Data yang telah di *crawling* menggunakan *widget* dari orange data mining dengan *Corpus* dan dihubungkan ke *Tweet Profiler*.



Gambar 5 Tweet Profiler

Distribution

Widget Distribusi menampilkan distribusi nilai atribut diskrit atau kontinu. Jika data berisi variabel kelas, distribusi dapat dikondisikan di kelas. Setelah melakukan tweet profiler pada widget. Langkah selanjutnya menghubungkan *Corpus* pada *Distribution*. Hasil akan terlihat 6 bentuk emosi dari data twitter yang telah diinput. Pada penelitian ini widget menampilkan emosi para pengguna twitter dengan *hashtag* Rachel Vennya, Karantina, dan Kabur. Dari hasil 6 emosi ini data menunjukkan bahwa respon dari twitter per tanggal 17 Januari 2022 adalah *Joy* dan *Surprise*.



Gambar 6 Distibution nilai atribut

V. KESIMPULAN

Berdasarkan hasil uraian analisis yang dilakukan diatas maka dapat ditarik kesimpulan yaitu beberapa trending topik yang menjadi fokus pembicaraan dimana yang sekiranya menjadi isu yang marak dibicarakan dan untuk melihat respon masyarakat terhadap hal tersebut. Selanjutnya Metode analisis menggunakan *tweet profiler* dapat mengetahui *mood* atau emosi para pengguna twitter dengan trending topik yang sedang terjadi di beberapa negara khususnya mengenai isu tersebut. Dengan melakukan analisis clustering distribution kita dapat mengetahui klasifikasi pengguna twitter dengan visualisasi emosi yang telah diinput ke dalam setiap *Corpus* dalam *Orange Data Mining*.

VI. REFERENSI

Alizah, M. D., Nugroho, A., Radiyah, U., & Gata, W. (2020). Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter. *Indonesian Journal on Software Engineering (IJSE)*,

6(2), 223–229. <https://doi.org/10.31294/ijse.v6i2.8991>

- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Hayuningtyas, R. Y., & Sari, R. (2019). Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naïve Bayes Dan Pso. *Jurnal Techno Nusa Mandiri*, 16(1), 37–42. <https://doi.org/10.33480/techno.v16i1.115>
- Indah Nurhafida, S., & Sembiring, F. (2021). Analisis Text Clustering Masyarakat Di Twiter Mengenai Mcdonald’Sxbts Menggunakan Orange Data Mining. *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 28–35.
- Jacarria Pangestu, D., & Kodar, A. (2022). Implementasi Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Terhadap Pelayanan Perusahaan Otobus Menggunakan Data Facebook (Studi Kasus: Grup Facebook Murni Jaya Lovers). *Jurnal Informatika: Jurnal Pengembangan IT*, 7(3), 156–160.
- Rifai, B., Normah, Febryanto, B. D., Yulianto, F., & Reflianah, N. (2021). Analisis Sentimen Opini Publik Terhadap Penerapan Kebijakan Social Distancing Dalam Pencegahan Covid-19. *Paradigma – Jurnal Informatika Dan Komputer*, 23(1), 55–62. <https://doi.org/https://doi.org/0.31294/p.v%vi%i.8756>
- Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review. *Faktor Exacta*, 13(1), 35. <https://doi.org/10.30998/faktorexacta.v13i1.5548>
- Yulianita, T., Utami, T. W., & Haris, M. Al. (2020). Analisis Sentimen Dalam Penanganan Covid-19 di Indonesia Menggunakan Naive Bayes Classifier. *Seminar Nasional Variansi (Venue Artikulasi-Riset, Inovasi, Resonansi-Teori, Dan Aplikasi Statistika)*, 235–243.