KOMPARASI METODE KLASIFIKASI UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PENERAPAN KURIKULUM MERDEKA

Wachid Darmawan[1)], Eny Jumiati[2)], Riski Sulistiyaningsih[3)]
STMIK Widya Pratama [123)]
wachiddw.stmikwp@gmail.com[1)], enyjumiati@stmik-wp.ac.id[2)], riskiul19@gmail.com[3)]

Abstrak

Di Indonesia, kegiatan belajar mengajar menggunakan kurikulum yang seringkali mengalami perubahan, tujuan perubahan tersebut adalah agar sistem pendidikan yang ada menjadi lebih baik. Kurikulum merdeka merupakan kurikulum yang diterapkan saat ini yang diluncurkan dan diperkenalkan pada tahun ajaran 2022/2023. Dengan adanya kurikulum merdeka banyak pendapat (opini) yang bernada positif ataupun bernada negatif yang muncul di Twitter. Untuk mengetahui akurasi dataset yang dihasilkan dari opini yang ada menggunakan metode klasifikasi. Selain itu metode klasifikasi digunakan untuk mengetahui model klasifikasi terbaik, serta dapat mengetahui analisis sentimen yang dihasilkan dari opini yang ada. Metode yang digunakan untuk klasifikasi adalah k-Nearest Neighbor (K-NN), Naive Bayes (NB) dan Support Vector Machine (SVM). Komparasi metode ini bertujuan untuk menemukan metode terbaik dari dataset yang ada, berdasarkan hasil akurasi yang dihasilkan dari ketiga metode tersebut. Dataset tweet yang dikumpulkan sebanyak 1000 tweet, setelah data dibersihkan diperoleh 220 tweet yang terdiri dari 110 tweet positif dan 110 tweet negatif. Berdasarkan perhitungan ketiga metode dengan menggunakan tools Rapidminer didapatkan akurasi sebagai berikut: untuk metode K-NN adalah 76,36%, akurasi metode NB adalah 70,91%, sedangkan akurasi metode SVM adalah 62,73%. Dari hasil tersebut dapat disimpulkan metode K-NN mendapatkan nilai akurasi terbaik dari metode klasifikasi yang digunakan. Selain itu pada perhitungan machine learning, metode SVM menghasilkan klasifikasi terbaik dalam memprediksi, dengan kategori excellent klasifikasi sumbu Y mendekati 1.00. Tidak hanya itu penerapan kurikulum merdeka oleh masyarakat banyak ditweetkan dengan opini yang bernada negatif. Harapan untuk penelitian kedepan adalah dapat menambahkan jumlah dataset tweet, serta menggunakan algoritma tambahan untuk meningkatkan akurasi dari metode klasifikasi.

Kata kunci: *Kurikulum Merdeka, k-Nearest Neighbor, Naïve Bayes, Support Vector Machine, Analisis Sentimen*

1. Pendahuluan

Kementerian Pendidikan dan Kebudayaan pada awal tahun 2020 mengeluarkan kebijakan Merdeka Belajar. Kebijakan ini tentunya sangat menarik perhatian masyarakat Indonesia karena salah satu program dari kebijkan merdeka belajar adalah munculnya kurikulum baru yaitu kurikulum merdeka (Prasetyo, Pramiyati and Isnainiyah 2021). Dengan adanya kurikulum baru memunculkan banyak diskusi di kehidupan nyata ataupun di media sosial.

Setiap tahunnya penggunaan media sosial di Indonesia memgalami peningkatan, salah satu yang paling populer di Indonesia adalah Twitter. Menurut data yang dihimpun oleh Meltwater dan We Are Social, pengguna Twitter di Indonesia telah menyentuh 24 juta orang per Januari 2023 (Kemp 2023). Dengan semakin bertambahnya pengguna Twitter, tentu semakin banyak pendapat yang sangat cepat memunculkan sebuah trending topik, selain itu banyak pula yang beropini, baik bernada positif ataupun negatif (Irfan, Fauzi and Tibyani 2018).

Banyak pengguna Twitter yang membahas kurikulum merdeka, yang tentunya akan berakibat timbulnya kontroversi. Sering munculnya salah satu argumentasi pada Twitter adalah siswa kurang memiliki waktu istirahat karena banyaknya proyek yang harus dikerjakan secara mandiri atau berkelompok (Darmawan, et al. 2023). Dengan perdebatan tersebut muncul banyak *tweet* tentang kurikulum merdeka yang disampaikan di Twitter.

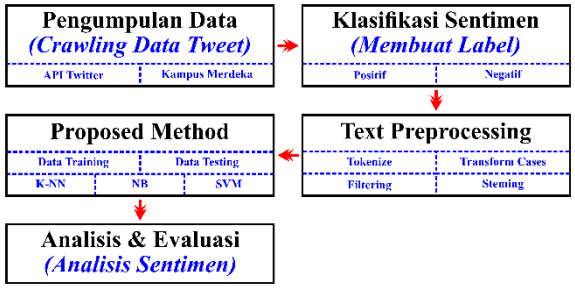
Telah banyak penelitian yang dilakukan mengenai analisis sentimen dengan data pengguna Twitter, seperti: Analisis Sentimen Kampus Pembelajaran Merdeka dengan metode Support Vector Machine (SVM). Tujuan penggunaan metode SVM adalah untuk prediksi arah mood seseorang baik positif maupun negatif dalam kegiatan PMM-BKM dengan nilai akurasi sebesar 95% (Kholila 2021). Sedangkan pada Analisis Sentimen Implementasi Program Pembelajaran Merdeka di Kampus Merdeka dengan metode Naive Bayes, K-Nearest Neighbors dan Decision Tree. 99,2° adalah nilai akurasi dari metode naïve Bayes dan nilai tersebut lebih tinggi dibanding K-Nearrest Neighbors dan Decission Dari ketiga metode yang digunakan, metode Naive Bayes memiliki akurasi yang lebih baik yaitu sebesar 99,2° pada tiga subjek yaitu positif, negatif dan netral (Rozaq, et al. 2022).

Penelitian lain terkait analisis sentimen program pendidikan tahun 2013 di Twitter menggunakan fitur ansambel dan metode K-nearest neighbor. Penerapan fungsi Ensemble Feature pada metode K-Nearest Neighbor akurasinya dapat meningkat sebesar 96%, yang sebelumnya hanya 80% (Irfan, Fauzi and Tibyani 2018). Pada penelitian terkait perbandingan teknik klasifikasi text mining dalam analisis sentimen pada review film, menggunakam metode SVM NBC dan K-NN. Dari metode yang digunakan SVM mendapatkan akurasi terbaik sebesar 72% (Ipmawati, Kusrini and Luthfi 2018).

Pada penelitian analisis sentimen terhadap kurikulum merdeka yang dilakukan adalah komparasi metode klasifikasi. Tujuan komparasi metode ini untuk mencari model terbaik dari metode yang diusulkan. Metode klasifikasi yang diusulkan, adalah: *k-Nearest Neighbor* (K-NN), *Naive Bayes* (NB) dan *Support Vector Machine* (SVM). Dalam penelitian ini *Dataset* diperoleh dari opini masyarakat yang ada di Twitter. Pengambilan *dataset* menggunakan teknik *crawling* data dengan kata kunci Kurikulum Merdeka. Setelah langkah *crawling* data, dilanjutkan langkah *preprocessing*. Dilanjutkan melakukan *proposed method* dan melakukan analisis yang diusulkan menggunakan *tools* Rapidminer. Setelah menyelesaikan tahap *proposed method*, dilanjutkan tahap analisis hasil berdasarkan hasil dari *machine learning* yang dilakukan *tools* Rapidminer.

2. Metode Penelitian

Untuk mengetahui model terbaik dari metode klasifikasi yang diusulkan menggunakan penelitian eksperimen terhadap kurikulum merdeka pada pengguna Twitter. Pada gambar 1 menjelaskan metode penelitian yang digunakan pada penelitian ini.



Gambar 1. Metode Penelitian Eksperimen

2.1 Pengumpulan Data

Rapidminer digunakan untuk mendapatkan dataset *tweet* atau biasa disebut dengan *crawling dataset*. *Crawling dataset* merupakan salah satu teknik yang digunakan untuk mengunpulkan *dataset* dengan menyertakan API (twitter) dalam pengumpulan data tersebut. Dalam proses penentuan sampling menggunakan teknik *purposive sampling* yaitu menentukan sampel berdasarkan tujuan yang diinginkan dengan menggunakan kriteria tertentu dari proses penelitian (Lenaini 2021). Kampus Merdeka adalah kata kunci yang digunakan untuk penentuan sampel dengan jumlah data yang diperoleh sebesar 1000 *tweet*.

2.2 Klasifikasi Sentimen

Setelah melakukan *crawling dataset* dilanjutkan dengan membersihkan *tweet-tweet* yang tidak *relevan*. Setelah itu dilanjutkan penentuan klasifikasi atau memberi label *tweet*, apakah *tweet* bernada positif atau negatif.

2.3 Text Preprocessing

Tujuan pada tahap ini adalah untuk mengetahui informasi apa saja yang ada di dokumen teks. Tahapan yang digunakan adalah: *Tokenizing, Transform Cases, Filtering (Tokens*

dan *Stopwords)* dan *Stemming* (Wibowo, Darmawan and Amalia 2022).

2.4 Proposed Method

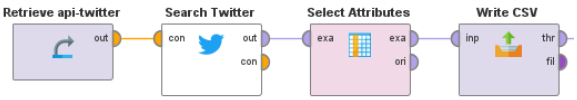
Tahap ini digunakan untuk mengetahui mengetahui model terbaik dari metode yang digunakan, yaitu: *k-Nearest Neighbor* (K-NN), *Naive Bayes* (NB) dan *Support Vector Machine* (SVM). Dari *dataset* yang ada akan dilakukan teknik *cross validation* untuk mengetahui model terbaik dari metode klasifikasi yang digunakan.

2.5 Analisis dan Evaluasi

Tahap terakhir adalah melakukan analisis dan evaluasi hasil dari perhitungan metode yang digunakan. Untuk menghitung akurasi metode klasifikasi pada penelitian ini menggunakan tools Rapidminer.

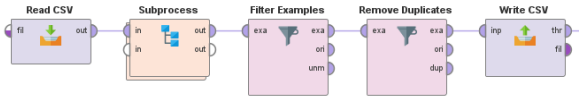
3. Hasil dan Pembahasan

Pada penelitian ini *dataset* yang digunakan sebanyak 1000 tweet dari hasil *crawling dataset* menggunakan tools Rapidminer. Dari *dataset* tersebut tidak semuanya digunakan karena ada *tweet* yang ganda ataupun yang tidak *relevan*. Setelah melakukan pembersihan dan membuat klasifikasi sentimen (label *dataset*) didapatkan *dataset* sebanyak 220 *tweet* yang terdiri dari 110 *tweet* positif dan 110 *tweet* negatif. Berikut tahapan proses *crawling dataset* pada gambar 2.



Gambar 2. *Crawling Dataset Tweet*

Setelah *dataset* dibentuk menjadi dokumen teks, dilanjutkan kedalam proses perbersihan tweet ganda ataupun tidak *relevan*.



Gambar 3. Proses Perbersihan Hasil *Crawling Dataset Tweet*

Setelah *dataset* sudah terbentuk dalam dokumen teks, akan dilanjutkan ketahap *teks preprocessing*, diantaranya:

3.1 *Teks Preprocessing*

3.1.1 *Tokenizing*

Proses *tokenizing* digunakan untuk memisahkan kata atau memotong string input disetiap *tweet* yang ada (Sari 2019) (Flores, Jasa and Linawati 2020), berikut contoh proses *tokenizing* di tabel 1.

Tabel 1. Contoh Proses *Tokenizing*

Sebelum *Tokenizing*	Setelah *Tokenizing*
I love kurikulum merdeka apa itu belajar teori	I love kurikulum merdeka apa itu belajar teori
Kurikulum merdeka ini yang merdeka siapa ya kok aku merasa terjajah	Kurikulum merdeka ini yang merdeka siapa ya kok aku merasa terjajah
aku gak kuat sama tugas projek kurikulum merdeka	aku gak kuat sama tugas projek kurikulum merdeka
kurikulum merdeka ini seru. ga berkutat terus terusan sama buku	kurikulum merdeka ini seru. ga berkutat terus terusan sama buku

3.1.2 *Transform Cases*

Tahap *transform cases* digunakan untuk mengubah huruf besar menjadi huruf kecil semuanya (Fani, Santoso and Suparti 2021), berikut contoh proses *transform cases* di tabel 2.

Tabel 2. Contoh Proses *Transform Cases*

Sebelum *Tokenizing*	Setelah *Tokenizing*
kurikulum merdeka	kurikulum merdeka
MERDEKA DARI	merdeka dari mana
MANA GUA TANYA	gua tanya wargahhhh
WARGAHHHH	
ogah kurikulum	ogah kurikulum
merdeka GWE MAU	merdeka gwe mau
KERJA SENDIRI	kerja sendiri leave
LEAVE ME ALONE	me alone
Bravo mas Menteri	bravo mas menteri
GASSS terus	gasss terus
Kurikulum merdeka	kurikulum merdeka
tuh SERU BGT	tuh seru bgt

3.1.3 *Filtering (Tokens* dan *Stopwords)*

Tahap *filtering* digunakan untuk mengambil kata-kata yang diperlukan dan menghapus kata-kata yang tidak perlu (Wibowo, Darmawan and Amalia 2022), berikut contoh proses *filtering* di tabel 3.

Tabel 3. Contoh Proses *Filtering*

Sebelum *Filtering*	Setelah *Filtering*
~~Saya~~ suka kurikulum	suka kurikulum
merdeka ~~apa itu~~	merdeka belajar teori
belajar teori	
Kurikulum merdeka	Kurikulum merdeka
~~ini yang~~ merdeka ~~siapa~~	merdeka terjajah
~~ya kok aku~~ ~~merasa~~	
terjajah	
~~aku gak~~ kuat ~~sama~~	kuat tugas projek
tugas projek	kurikulum merdeka
kurikulum merdeka	
kurikulum merdeka ~~ini~~	kurikulum merdeka
seru. ~~ga~~ berkutat ~~terus~~	seru. berkutat buku
~~terusan~~ ~~sama~~ buku	

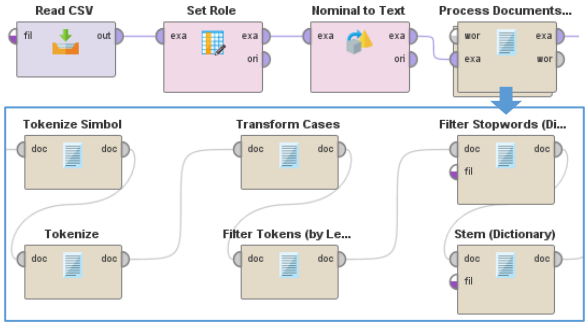
3.1.4 *Stemming*

Tahap *stemming* digunakan sebagai proses konversi semua kata yang menerima imbuhan, baik di awal atau di akhir dan dijadikan sebagai kata dasar (Irfan, Fauzi and Tibyani 2018), berikut contoh proses *stemming* di tabel 4.

Tabel 4. Contoh Proses *Stemming*

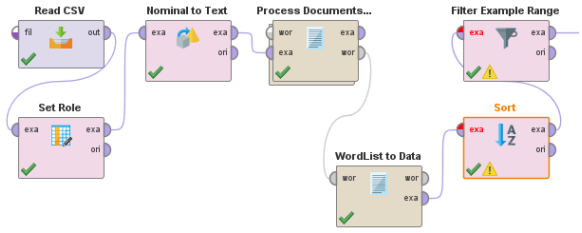
Sebelum *Stemming*	Setelah *Stemming*
tujuan kurikulum	tujuan kurikulum
merdeka tuh bukannya	merdeka tuh bukan
bikin anak pinter tapi	bikin anak pinter tapi
bikin anak sering	bikin anak sering
opname	opname
sumpah kalian yg	sumpah kalian yg
kurikulum merdeka	kurikulum merdeka
projeknya pd susah ga	projek pd susah ga si
si topiknya	topik
Itulah mengapa	Itu mengapa
sekarang kita	sekarang kita laku
berlakukan kurikulum	kurikulum merdeka
merdeka Om. yang	Om. yang upaya
berupaya	kembali ruh
mengembalikan ruh	pendidikan
pendidikan	

Proses *teks preprocessing* pada penelitian ini menggunakan *tools* Rapidminer, gambar 4 menjelaskan proses yang dilakukan.



Gambar 4. Proses *Text Preprocessing*

Pada tahap *text preprocessing* muncul juga kata-kata yang sering ditweetkan oleh masyarakat, pada gambar 5-7 menjelaskan proses dan hasil dari visualisasi *wordcloud* kata-kata yang sering muncul



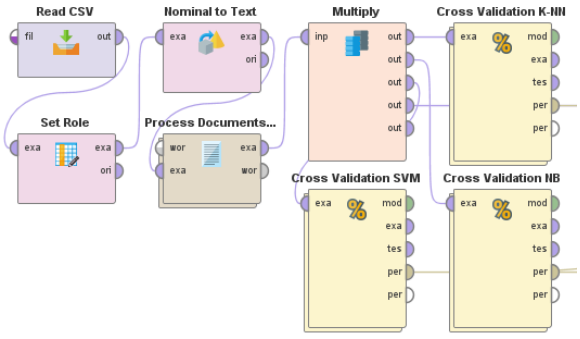
Gambar 5. Proses Menampilkan *Visualisasi Wordcloud*

Gambar 6. *Visualisasi Wordcloud Tweet* Positif

Gambar 7. *Visualisasi Wordcloud Tweet* Negatif

3.2 *Proposed Method*

Untuk melihat model terbaik pada penelitian ini menggunakan *tools* Rapidminer. Dari ketiga metode klasifikasi yang digunakan untuk perhitungan metodenya menggunakan teknik *cross validation*. Pada gambar 8 menjelaskan prosesnya



Gambar 8. Proses Komparasi Metode Klasifikasi

3.3 Analisis dan Evaluasi

Dari perhitungan model klasifikasi dengan menggunakan *dataset tweet* "Kurikulum Merdeka" yang dilakukan dapat dijelaskan pada gambar 9-11. Berikut hasil penelitian ekperimen pada metode K-NN, NB dan SVM.

accuracy: 76.36% +/- 5.59% (micro average: 76.36%)

	true negatif	true positif	class precision
pred. negatif	91	33	73.39%
pred. positif	19	77	80.21%
class recall	82.73%	70.00%	

Gambar 9. Hasil Metode K-NN

Gambar 9 dapat dijelaskan, *tweet* bernada positif, *machine learning* memprediksi 77 *tweet* positif dan ada 33 *tweet* negatif. Sedangkan *tweet* bernada negatif, *machine learning* memprediksi 19 *tweet* positif dan ada 91 *tweet* negatif.

accuracy: 70.91% +/- 9.87% (micro average: 70.91%)

	true negatif	true positif	class precision
pred. negatif	85	39	68.55%
pred. positif	25	71	73.96%
class recall	77.27%	64.55%	

Gambar 10. Hasil Metode NB

Gambar 10 dapat dijelaskan, *tweet* bernada positif, *machine learning* memprediksi 71 *tweet* positif dan ada 39 *tweet* negatif. Sedangkan *tweet* bernada negatif, *machine learning* memprediksi 25 *tweet* positif dan ada 85 *tweet* negatif.

accuracy: 62.73% +/- 5.98% (micro average: 62.73%)

	true negatif	true positif	class precision
pred. negatif	109	81	57.37%
pred. positif	1	29	96.67%
class recall	99.09%	26.36%	

Gambar 11. Hasil Metode SVM

Gambar 11 dapat dijelaskan, *tweet* bernada positif, *machine learning* memprediksi 29 *tweet* positif dan ada 81 *tweet* negatif. Untuk yang *tweet* bernada negatif, *machine learning* diprediksi 1 *tweet* positif dan ada 109 *tweet* negatif.

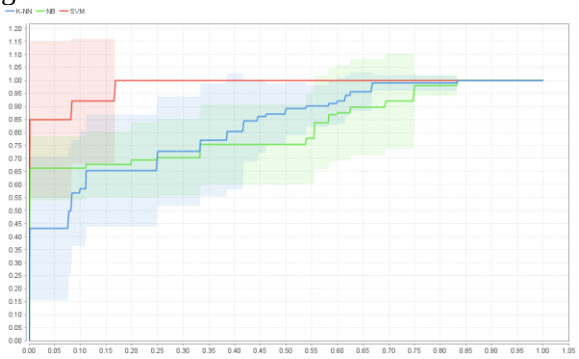
Hasil dari ketiga metode klasifikasi yang digunakan disajikan pada tabel berikut, yang terdiri dari hasil pengukuran *accuracy*, *precision* dan *recall*.

Tabel 5. Hasil Pengukuran Metode Klasifikasi

Metode	*Accuracy*	*Precision*		*Recall*	
		Positif	*Negatif*	*Positif*	*Negatif*
K-NN	76,36%	80,21%	73,39%	70,00%	82,73%
NB	70,91%	73,96%	68,55%	64,55%	77,27%
SVM	62,73%	96,67%	57,37%	26,36%	99,09%

Tabel 5 menjelaskan bahwa tingkat akurasi yang dihasilkan oleh metode klasifikasi pada *dataset tweet* masyarakat hasilnya belum baik (Ayudhitama and Pujianto 2020). Tetapi dari komparasi yang dihasilkan metode K-NN memiliki tingkat akurasi terbaik dari ketiga metode klasifikasi yang digunakan.

Sedangkan untuk memprediksi jumlah klasifikasi *tweet* yang lebih banyak ada di metode SVM dari ketiga metode klasifikasi yang digunakan. Seperti yang tergambarkan di kurva ROCs penggunaan tiga metode klasifikasi pada gambar 12.



Gambar 12. Kurva ROCs

Gambar 12 terlihat metode SVM memiliki grafik ROCs mendekati 1.0 pada sumbu Y pada *class True Positive*. Hal ini membuktikan bahwa metode SVM merupakan metode dengan kualitas klasifikasi "*Excellent*" pada rentang akurasi 0,90-1,00 (Ayudhitama and Pujianto 2020). Metode SVM memiliki hasil klasifikasi yang baik dalam menyelesaikan masalah klasifikasi pada *dataset tweet* yang digunakan.

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian eksperimen dengan melakukan komparasi metode pada *K-Nearest Neighbor*, *Naive Bayes* dan *Support Vector Machine* menggunakan *tools* Rapidminer. Untuk pengambilan *dataset* menggunakan teknik *crawling dataset* dari hasil *tweet* di Twitter. Hasil penelitian menunjukkan bahwa dari ketiga metode yang digunakan dalam komparasi metode klasifikasi, metode *K-Nearest Neighbor* adalah metode dengan nilai akurasi terbaik sebesar 76,36%. Berdasarkan hasil kurva ROCs metode *Support Vector Machine*, grafik sumbu Y mendekati 1,00, masuk dalam kategori "*Excellent*" klasifikasi. Berdasarkan perhitungan *machine learning* yang dilakukan, analisis sentimen pada opini masyarakat terkait implementasi kurikulum merdeka banyak mendapat tanggapan atau opini negatif.

Untuk penelitian lebih lanjut, dapat menambah jumlah *dataset* dan memperbarui *dataset* tersebut dengan teknik *crawling* data *tweet* terbaru. Selain itu, perlu ditambahkan algoritma agar akurasi dari metode klasifikasi dapat meningkat.

DAFTAR PUSTAKA

Ayudhitama, Annisa Putri, dan Utomo Pujianto. 2020. "Analisa 4 Algoritma Dalam Klasifikasi Penyakit Liver Menggunakan Rapidminer." *Jurnal Informatika Polinema (JIP)* 6 (2): 1-9.

Darmawan, Wachid, Muhammad Faizal Kurniawan, Wahyu Setianto, dan Wim Hapsoro. 2023. "Analisis Sentimen Penerapan Kurikulum Merdeka Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor Dengan Forward Selection." *Smart Comp: Jurnalnya Orang Pintar Komputer* 12 (1): 245-253.

Fani, Syiva Multi, Rukun Santoso, dan Suparti. 2021. "Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K-means Clustering." *Jurnal Gaussian* 10 (4): 583-593.

Flores, Veronica Ambassador, Lie Jasa, dan Linawati. 2020. "Analisis Sentimen untuk Mengetahui Kelemahan dan Kelebihan Pesaing Bisnis Rumah Makan Berdasarkan Komentar Positif dan Negatif di Instagram." *Majalah Ilmiah Teknologi Elektro* 66 (1): 49-54.

Ipmawati, Joang, Kusrini, dan Emha Taufiq Luthfi. 2018. "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen." *Indonesian Journal on Networking and Security* 6 (1): 28-36.

Irfan, M. Rizzo, Mochammad Ali Fauzi, dan Tibyani Tibyani. 2018. "Analisis Sentimen Kurikulum 2013 pada Twitter menggunakan Ensemble Feature dan Metode K-Nearest Neighbor." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIK)* 2 (9): 3006-3014.

Kemp, Simon. 2023. *DataReportal: Digital 2023: Global Overview Report.* 26 Januari. https://datareportal.com/reports/digital-2023-global-overview-report.

Kholila, Ni'ma. 2021. "Sentiment Analysis of Program Merdeka Belajar - Kampus Merdeka on Twitter Using Support Vector Machine." *ANTIVIRUS: Jurnal Ilmiah Teknik Informatika* 15 (2): 252-261.

Lenaini, Ika. 2021. "Teknik Pengambilan Sampel Purposive Dan Snowball Sampling." *Jurnal Historis* (Penerbit BPFE) 33-39.

Prasetyo, Herlambang Dwi, Titin Pramiyati, dan Ika Nurlaili Isnainiyah. 2021. "Sentimen Analisis Pengguna Twitter Terhadap Kebijakan Merdeka Belajar Menggunakan Algoritma Naive Bayes." *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)* 2 (1): 559-568.

Rozaq, Abdul, Yessi Yunitasari, Kelik Sussolaikah, Eka Resty Novieta Sari, dan Restyono Ilham Syahputra. 2022. "Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighboars Dan Decision Tree." *Jurnal Media Informatika Budidarma* 6 (2): 746-750.

Sari, Yunita. 2019. *Pengenalan Natural Language Toolkit (NLTK) Bagian 1*. Vol. 11, dalam *Pengenalan Natural Language Toolkit*, oleh Yunita Sari, 59-66. Yogyakarta: Departemen Ilmu Komputer dan Elektronika, UGM.

Wibowo, Ari Putra, Wachid Darmawan, dan Nurul Amalia. 2022. "Komparasi Metode Naïve Bayes dan k-Nearest Neighbor Terhadap Analisis Sentimen Pengguna Aplikasi Pedulilindungi." *Journal of Informatic and Computer Technology (IC-Tech)* 17 (1): 18-23.