
ANALYSIS OF SENTIMENT ON TEACHER MARKETPLACE ISSUES USING THE LEXICON AND K-NEAREST NEIGHBOR ALGORITHMS

Addien Anaba¹, Rahmawan Bagus Trianto², Eko Supriyadi

Prodi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas An Nuur Purwodadi, Indonesia

ABSTRACT

The advancement of social media makes it easier for users to express opinions. Twitter has become one of the media that is loved by internet users, users can freely express their thoughts or opinions, apart from that they can also express everything that is being experienced. The busy issue of the Teacher Marketplace initiated by the Minister of Education, Nadiem Makarim, has invited many comments from internet users. Twitter users' tendencies in posting content can be determined by analyzing sentiment. In this research, the Lexicon and K-Nearest Neighbor (KNN) methods are proposed to analyze sentiment towards the education minister's discourse on Twitter social media on the topic of Teacher Marketplace Issue Sentiment by classifying it into positive, neutral and negative. The results of this research show that the accuracy value obtained was 91.70%, precision 90.51%, recall 71.95%. By carrying out this sentiment analysis, it is hoped that the problems contained in the Marketplace Guru topic controversy can be identified, used as input and consideration for further research.

Keywords: Twitter, K-Nearest Neighbor, sentiment analysis.

Correspondence:

Addien Anaba

Universitas An Nuur Purwodadi, Email; addien.id@gmail.com

PENDAHULUAN

Kemajuan media sosial saat ini semakin memudahkan pengguna internet dalam mengakses berbagai informasi yang ramai diperbincangkan. Berbagai media sosial seperti Twitter, TikTok, Instagram, Facebook dll. disediakan untuk pengguna Internet. Media sosial ini memungkinkan penggunanya melakukan banyak aktivitas. Tidak hanya untuk memperoleh informasi, tetapi juga untuk media partisipatif seperti pembuatan dan argumentasi konten, pemberian komentar dan penyebaran berbagai informasi seperti tulisan (karya tulis), audio (podcast), gambar (poster) bahkan video (film). Selain itu, media sosial juga digunakan sebagai wadah untuk berkomunikasi, membentuk komunitas, menjalankan bisnis bahkan membentuk organisasi kemasyarakatan serta memperlancar hubungan antara masyarakat dan pemerintah yang membuat segalanya menjadi lebih mudah saat ini (Arsi et al., 2021).

Belakangan ini Twitter digegerkan dengan trendingnya isu dari dunia pendidikan yang sedang dihebohkan dengan isu *Marketplace* Guru yang digagaskan oleh Menteri Pendidikan Nadiem Makarim. Dalam konteks *Marketplace* ini guru jadi pusat yang disponsorshipkan, lalu setiap pendidikan dapat memilih pengajar sesuai dengan

kebutuhan dengan mengkualifikasi yang ada di *marketplace* tersebut. Mendikbud Ristek Nadiem Makarim telah mengungkapkan bahwa Konsep *Marketplace* Guru yang ia sampaikan pada rapat kerja bersama Komisi X DPR RI diharapkan dapat menyelesaikan berbagai proplema dalam dunia pendidikan khususnya tata kelola guru. Pada konsep ini, sekolah dapat memilih sendiri saat rekrut guru via *marketplace* (Jurnalpost, 2023).

Twitter berfokus pada pengembangan metode analisis sentimen yang dirancang khusus untuk tweet. Analisis sentimen Twitter dilakukan untuk mengetahui opini atau tren opini terhadap nilai pastinya. Akurasi dihitung dengan menghitung persentase data latih dan data uji. Penggunaan Analisis sentimen dalam penelitian ini bertujuan untuk menemukan informasi berharga yang dibutuhkan dari data yang tidak terstruktur dari media sosial Twitter, sehingga diharapkan pada penelitian ini dapat diketahui sentimen (Positif, Netral atau Negatif) pengguna twitter terhadap isu *Marketplace* Guru (Giovani et al., 2020).

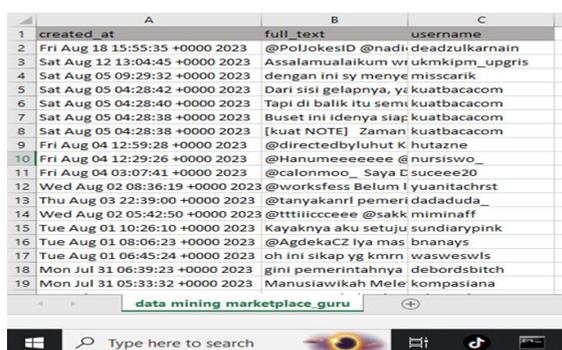
METODE PENELITIAN

Penelitian ini menggunakan algoritma K-NN dikarenakan K-NN dapat memprosedur yang berbasis matematis

untuk mengevaluasi nilai kriteria-kriteria tersebut menjadi sebuah keterangan cklasifikasi. Proses model data mining pada penelitian ini menggunakan metode SEMMA. Banyak proses model data mining yang dapat digunakan diantaranya *Knowledge Discovery Databases (KDD)*, *Cross Industry Standard Process (CRISP-DM)*, dan *Sample, Explore, Modify, Model, Access (SEMMA)* ini sendiri. Setiap model memiliki keunggulannya masing-masing.

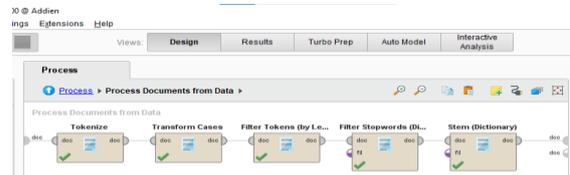
Penelitian ini memilih SEMMA dikarenakan SEMMA memiliki keunggulan lebih terpusat pada aspek pengembangan model dari data mining sehingga pembuatan model lebih terstruktur.

SEMMA memiliki lima tahapan proses yang pertama *sample* merupakan proses pengumpulan data dari twitter atau kerap dikenal dengan istilah *crawling data* menggunakan library *tweet-harvest* kemudian file yang didapatkan disimpan dengan format .csv.



	A	B	C
1	created_at	full_text	username
2	Fri Aug 13 15:55:35 +0000 2023	@POLokesiD	@nadi:deadzulkarnain
3	Sat Aug 12 13:04:45 +0000 2023	Assalamualaikum wr ukmkipm_upgris	
4	Sat Aug 05 09:29:32 +0000 2023	dengan ini sy menyen misscarik	
5	Sat Aug 05 04:28:42 +0000 2023	Dari sisi gelapnya, ya kuatbacacom	
6	Sat Aug 05 04:28:40 +0000 2023	Tapi di balik itu semi kuatbacacom	
7	Sat Aug 05 04:28:38 +0000 2023	Buset ini idenya siap kuatbacacom	
8	Sat Aug 05 04:28:38 +0000 2023	[kuat NOTE] Zaman kuatbacacom	
9	Fri Aug 04 12:59:28 +0000 2023	@directedbyluhut K hutazne	
10	Fri Aug 04 12:29:26 +0000 2023	@Hanumeeeeeee @nursiswo_	
11	Fri Aug 04 03:07:41 +0000 2023	@calonmoo_ Saya E suceee20	
12	Wed Aug 02 08:36:19 +0000 2023	@worksfess Belum l yuantachrst	
13	Thu Aug 03 22:39:00 +0000 2023	@tanyakanrl pemerid dadaduda_	
14	Wed Aug 02 05:42:50 +0000 2023	@tttiicceeee @sakk miminaff	
15	Tue Aug 01 10:26:10 +0000 2023	Kayaknya aku setuju sundiarypink	
16	Tue Aug 01 08:06:23 +0000 2023	@AgdekaCZ Iya mas bnanays	
17	Tue Aug 01 06:45:24 +0000 2023	oh ini sikap yg kmrn wasweswis	
18	Mon Jul 31 06:39:23 +0000 2023	gini pemerintahnya debordsbitch	
19	Mon Jul 31 05:33:32 +0000 2023	Manusiawikah Mele kompasiana	

Langkah yang kedua *explore* dimana data yang telah didapatkan dikumpulkan berdasarkan ide-ide terkait



yang dibangun kemudian divisualisasikan untuk menampilkan informasi data secara visual.

Langkah ketiga *modify* yaitu proses preprocessing data sebelum diklasifikasikan, tahap yang dilakukan yaitu *case folding*, *cleaning*, *tokenize*, *normalize*, *stopword*, dan *stemming*. Proses text preprocessing menggunakan RapidMiner.

Langkah pertama yang dilakukan ialah pembersihan karakter yang tidak diperlukan seperti simbol, angka, mention, tagar. Kemudian pemecahan kata per kata menggunakan *tokenize* isi tweet yang awalnya berupa kalimat dipecah per kata. Selanjutnya semua karakter diubah menjadi huruf kecil menggunakan *transform cases*, selanjutnya menghapus kata hubung, kata ganti kemudian dikembalikan menjadi kata dasar. Dari 1006 data menjadi 1002 data setelah melewati tahap preprocessing. Contoh hasil tahapan preprocessing sebagai berikut:

No.	Data Awal	Data Setelah Preprocessing
1.	@suriresutari_@Guru_PNS Gimana ini ceritanya guru marketplace? Guru di jualkah?	gimana cerita guru marketplace? guru di jualkah

Sentiment	Jumlah
Positif	181
Netral	787
Negatif	33

2.	@PolJokesID @nadiemmakarim Pa Nadiem lg sibuk bikin marketplace buat guru ðŸ«£	pa nadiem lg sibuk bikin marketplace buat guru
3.	Dari sisi gelapnya, yang pertama ide belum menjelaskan masalah status PPPK, yang kedua belum tau siapa yang akan menggaji guru marketplace. Waduh ko bisa ya..ðŸ˜… https://t.co/qJUI94xUkj	dari sisi gelapnya yang pertama ide belum menjelaskan masalah status pppk kedua belum tau siapa yang akan menggaji guru marketplace waduh ko bisa ya
..1002
..

Langkah keempat *model* yaitu tahap pengolahan data menggunakan algoritma *lexicon* untuk pelabelan, *TF-IDF* untuk pembobotan kata dan *K-Nearest Neighbor* untuk klasifikasi.

1. *Lexicon*

Kata-kata yang terdapat dalam tweet akan dibandingkan dengan kamus *lexicon*. Jika kata tersebut terdapat dalam kamus *lexicon* maka akan diberi score, score tersebut yang menjadikan kata tersebut berlabel positif, netral atau negatif. Kamus *lexicon* yang dipakai adalah *Indonesian Sentiment (InSet) Lexicon* yang bersumber dari Koto (2017). InSet ini berisi sekitar 10.225 kata yang diberi nilai dari -5 sampai +5.

Tabel 4.2 menjelaskan hasil dari pelabelan menggunakan metode *lexicon based*. Didapatkan sebanyak 181 tweet yang masuk ke dalam label sentimen positif, 787 tweet masuk ke dalam sentimen netral dan 33 masuk ke dalam label sentimen negatif.

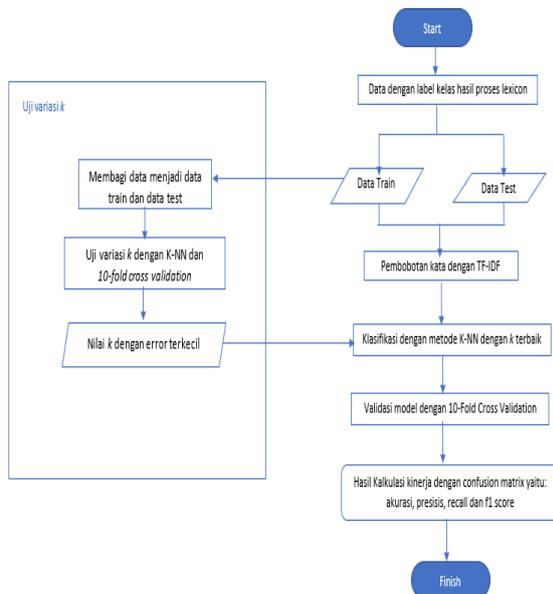
2. *TF-IDF*

Setiap term yang ada pada tweet yang telah melewati tahap preprocessing diberi bobot. Pembobotan tersebut bertujuan untuk memberikan nilai kepada suatu term yang nantinya akan dijadikan sebagai input pada proses klasifikasi. Berikut adalah simulasi perhitungan bobot kata dengan TF-IDF.

$$TF \times IDF(d, t) = TF(d, t) \times \log \frac{N}{df(t)}$$

Untuk :

$TF \times IDF (d,t)$ = bobot term t pada dokumen d



$TF(d,t)$ = frekuensi kemunculan term t pada dokumen d
 N = total seluruh dokumen
 $df(t)$ = jumlah dokumen yang terdapat term t

3. K-Nearest Neighbor

Hal pertama yang dilakukan adalah mencari varian terbaik dari nilai k dengan cara menguji beberapa nilai k , kemudian memilih nilai k dengan nilai error terkecil, dan menguji algoritma. Jadi adanya metode lexicon dalam penelitian ini ialah untuk membantu metode K-NN dalam melabelkan data menjadi kelas positif, netral dan negatif.

Gambar 3.3 flowchart tahapan model

Langkah terakhir dalam SEMMA yaitu asses merupakan tahapan menilai

data dengan mengevaluasi kegunaan dan keandalan hasil proses data mining dan mengevaluasi seberapa baik kerjanya (Azevedo and Santos, 2008).

HASIL

Setelah dataset dibagi menjadi dua bagian, yaitu data latih dan data uji dengan beberapa percobaan rasio test dan train yaitu 90% latih dan 10% data uji, 80% data latih dan 20% data uji, 70% data latih dan 30% data uji. Selanjutnya dilakukan tahap evaluasi dari model penelitian menggunakan 5-fold cross validation untuk mencari nilai variasi k terbaik dan confusion matrix untuk melihat nilai akurasi, presisi dan recall dari data test.

Kemudian dari tahap-tahapan tersebut mendapatkan hasil sebagai berikut:

accuracy: 91.70%				
	true Netral	true Positif	true Negatif	class precision
pred. Netral	687	46	26	90.51%
pred. Positif	0	118	1	99.16%
pred. Negatif	0	0	2	100.00%
class recall	100.00%	71.95%	6.90%	

a. Hasil Uji Variasi k

1. Rasio 90:10

Pada rasio 90:10 didapatkan hasil akurasi sebesar 91.70% dengan presisi pada kelas positif sebesar 99.16%, kelas netral 90.51% dan kelas negatif 100%.

2. Rasio 80:20

Pada rasio 80:20 didapatkan hasil akurasi sebesar 91.14% dengan presisi pada kelas positif sebesar 99.13%, kemudian kelas netral 100% dan kelas negatif sebesar 89.92%.

accuracy: 88.30%				
	true Netral	true Positif	true Negatif	class precision
pred. Netral	681	69	27	87.64%
pred. Positif	6	95	1	93.14%
pred. Negatif	0	0	1	100.00%
class recall	99.13%	57.93%	3.45%	

3. Rasio 70:30

Selanjutnya pada rasio 70:30 didapatkan akurasi sebesar 88.30% dengan tingkat presisi pada kelas positif 93.14, kelas netral 87.64% kelas negatif 100%.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dipaparkan dapat ditarik kesimpulan sebagai berikut:

1. Dari hasil crawling data menggunakan kata kunci Marketplace Guru dengan kurun waktu dari bulan Mei sampai Agustus didapatkan data sebanyak 1006 dan setelah melewati tahap preprocessing tersisa 1001 data. kemudian mendapatkan label positif sebanyak 181 data, netral 787 data dan negatif sebanyak 33 data.

2. Hasil akurasi algoritma K-NN didapatkan k terbaik adalah 7 dengan

tingkat akurasi 91,70%, presisi 90,51%

accuracy: 91.14%				
	true Netral	true Positif	true Negatif	class precision
pred. Netral	687	50	27	89.92%
pred. Positif	0	114	1	99.13%
pred. Negatif	0	0	1	100.00%
class recall	100.00%	69.51%	3.45%	

dan recall 71,95%

Saran

Berdasarkan hasil penelitian, penulis memiliki beberapa saran yang dapat dijadikan masukan dan bahan pertimbangan untuk penelitian selanjutnya sebagai berikut:

1. Berdasarkan hasil analisis yang diperoleh, penelitian ini terbatas dan mengalami kendala pada tahap modify atau preprocessing data, dimana pengolahan data terbatas pada tahap normalisasi kata pada kamus normalisasi bahasa indonesia saja. Sedangkan isi tweet terkadang berisi bahasa daerah atau bahkan terdapat banyak bahasa sehari-hari seperti saya menjadi “gue”, “gw” “aing” dan banyak kata dengan huruf ganda, seperti “sekaliiii”, “semaangaatttt” yang dampaknya kata-kata tersebut tidak ternormalisasi.

2. Data sentimen dalam kamus *lexicon Indonesia Sentimen* (InSet) masih terbatas dan belum lengkap, sehingga untuk penelitian selanjutnya dapat menggunakan metode lain seperti

TextBlop dengan menggunakan translate
terlebih dahulu.

DAFTAR PUSTAKA

- Agusta, L., Teknologi, F., Universitas, I., & Wacana, K. S. (2009). PERBANDINGAN ALGORITMA STEMMING PORTER DENGAN ALGORITMA NAZIEF & ADRIANI UNTUK STEMMING DOKUMEN TEKS BAHASA INDONESIA. In *Konferensi Nasional Sistem dan Informatika*.
- Akhmad Deviyanto. (2018). *PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR*.
- Ali Imron. (2019). *ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DI KABUPATEN REMBANG MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER*.
- Anggi Priliani Yulianto, & Darwis, S. (2021). Penerapan Metode K-Nearest Neighbors (kNN) pada Bearing. *Jurnal Riset Statistika*, 1(1), 10–18. <https://doi.org/10.29313/jrs.v1i1.16>
- Arie Atwa Magriyanti. (n.d.). *ANALISIS PENGEMBANGAN ALGORITMA PORTER STEMMING DALAM BAHASA INDONESIA*. <http://kompas.com>
- Asri, K. N. (2021). Analisis Sentimen Pindah Ibu Kota Berbasis Naive Bayes Classifier. *Jurnal Informatika UPGRIS*, 7, no. 1, 1–6.
- Azevedo, A., & Santos, M. F. (2008a). KDD, semma and CRISP-DM: A parallel overview Assessment in Higher Education View project KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW. In *International Journal of Business Intelligence Research*. <https://www.researchgate.net/publication/220969845>
- Azhar. (2019). *ANALISIS KINERJA ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA SENTIMEN ANALISIS*.
- Budiyanto, A. (2003). *Pengantar Algoritma dan Pemrograman*.
- Deolika, A., & Taufiq Luthfi, E. (2019). ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING. *Jurnal Teknologi Informasi*, 3(2).
- Dwi Putranti, N., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *IJCCS*, 8(1), 91–100.
- Fauziah, S., Nur Sulistyowati, D., & Asra, T. (2019). OPTIMASI ALGORITMA VECTOR SPACE MODEL DENGAN ALGORITMA K-NEAREST NEIGHBOUR PADA PENCARIAN JUDUL ARTIKEL JURNAL. *Maret*, 15(1), 21. www.bsi.ac.id
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Heri Kurniawan. (2017). *Sistem Penentuan Tingkat Kualitas Air pada Depot Air Minum Menggunakan Metode K-Nearest Neighbor (KNN)*.
- Indriani, A. (n.d.-a). *ANALISA PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DAN K-NEAREST NEIGHBOR TERHADAP KLASIFIKASI DATA*.
- Indriani, A. (n.d.-b). Klasifikasi Data Forum dengan menggunakan Metode Naive

- Bayes Classifier. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*. www.bluefame.com,
- Jayani, D. H. (2020). *10-media-sosial-yang-paling-sering-digunakan-di-indonesia*.
- Jurnalpost. (2023). *Rekrutmen Guru lewat Marketplace, Sekolah Pilih Sendiri, Masa Kontrak Kerja PPPK Guru Dihilangkan*.
<https://jurnalpost.com/rekrutmen-guru-lewat-marketplace-sekolah-pilih-sendiri-masa-kontrak-kerja-pppk-guru-dihilangkan/51533/>
- Liu, B. 2010. (n.d.). *Sentiment Analysis and Subjectivity*.
- Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). *ANALISIS SENTIMEN PENGGUNA GOPAY MENGGUNAKAN METODE LEXICON BASED DAN SUPPORT VECTOR MACHINE*.
<http://studentjournal.umpo.ac.id/index.php/komputek>
- Matresya Matulatuwa, F., Sedyono, E., & Iriani, A. (2017). *TEXT MINING DENGAN METODE LEXICON BASED UNTUK SENTIMENT ANALYSIS PELAYANAN PT. POS INDONESIA MELALUI MEDIA SOSIAL TWITTER* (Vol. 2, Issue 3).
- Maulana, H. A., & Witanti, A. (2023). *JURNAL TEKNIK INFORMATIKA AUTOMATIC SUMMARIZING DOKUMEN REPOSITORY DENGAN TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY METHOD*. *JUTEKIN*, 11(1).
<https://doi.org/10.51530/jutekin.v11i1.665>
- MUCHAMAD HENDRIX PRISTYAWAN. (2018). *ANALISIS ALGORITMA SEQUENTIAL SEARCH UNTUK Mencari Data Mahasiswa Teknik Informatika Pada SEKOLAH TINGGI TEKNOLOGI PELITA BANGSA SKRIPSI*.
- NaikPangkat.com. (2023). *Marketplace Guru Menjadi Solusi Atau Justru Marketplace Mempersulit Guru?*
<https://naikpangkat.com/marketplace-guru-menjadi-solusi-atau-justru-mempersulit-guru/>
- Ni'mah, A. T., & Arifin, A. Z. (2020). *Perbandingan Metode Term Weighting terhadap Hasil Klasifikasi Teks pada Dataset Terjemahan Kitab Hadis*. *Rekayasa*, 13(2), 172–180.
<https://doi.org/10.21107/rekayasa.v13i2.6412>
- P. Arsi, B. A. K. A. N. (2021). *Analisis Sentimen Pindah Ibu Kota Berbasis Naive Bayes Classifier*. *Jurnal Informatika UPGRIS*, Vol. 7, No. 1, 1–6.
- Permana, A. Y., & Eendi, D. M. M. (2020). *Analisis Sentimen pada Teks Opini Penilaian Kinerja Dosen dengan Pendekatan Algoritma KNN*. *Jurnal Ilmiah KOMPUTASI*, 19, 1–12.
<https://doi.org/10.32409/jikstik.19.1.2729>
- Poona. (2013). *Knowledge seeker's blog Knowledge seeker's blog CRM and data mining Day 08*.
- Putro, H. F., Vlandari, R. T., & Saptomo, W. L. Y. (2020). *Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan*. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, 8(2).
<https://doi.org/10.30646/tikomsin.v8i2.500>
- Refka Muhammad F. (2020). *Deteksi Berita Rumor pada Sosial Media Twitter Menggunakan Metode Naive Bayes Multinomial dengan Pembobotan TF-IDF*.
- Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023a). *Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes*

-
- dan KNN. *Jurnal KomtekInfo*, 1–7.
<https://doi.org/10.35134/komtekinfo.v10i1.330>
- Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023b). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, 1–7.
<https://doi.org/10.35134/komtekinfo.v10i1.330>
- Syaifudin, Y. W., & Irawan, R. A. (2018). *IMPLEMENTASI ANALISIS CLUSTERING DAN SENTIMEN DATA TWITTER PADA OPINI WISATA PANTAI MENGGUNAKAN METODE K-MEANS* (Vol. 4).
- Tessem, B., Bjørnstad, S., Chen, W., & Nyre, L. (2015). Word cloud visualisation of locative information. *Journal of Location Based Services*, 9(4), 254–272.
<https://doi.org/10.1080/17489725.2015.1118566>
- Tjipta, E., Sekolah, W., Manajemen Informatika, T., Komputer, D., & Malang, A. (2013). 78 Sekolah Tinggi Manajemen Informatika dan Komputer ASIA Malang PERANCANGAN INFORMATION RETRIEVAL (IR) BERBASIS TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK PERINGKASAN TEKS TUGAS KHUSUS BERBAHASA INDONESIA. In *Jurnal Ilmiah Teknologi dan Informasi ASIA* (Vol. 7, Issue 1).
- Widhi Saputro, I., & Wulan Sari, B. (2019). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction. *Citec Journal*, 6(1).