



Analisis Sentimen Terhadap Perubahan Rute Krl Commuter Jabodetabek Menggunakan Algoritme Support Vector Machine (Svm)

Fauzapril Duta Sanubari¹, Ultach Enri², Susilawati³

^{1,2}. Universitas Singaperbangsa Karawang

Received: 11 Juni 2023

Revised: 12 Juli 2023

Accepted: 23 Juli 2023

Abstract

KRL Commuter Jabodetabek is one of the modes of public transportation that is an alternative choice for residents of the capital city of Jakarta and its surroundings to reduce congestion in the Jabodetabek area. However, since there was a change in the Jabodetabek Commuter KRL route on May 28, 2022, there have been pro and con opinions from among the public users of the Jabodetabek Commuter KRL public transportation mode. The data used in this research is tweet data from Twitter with the keyword 'krl route changes' with a time span between May 26, 2022 and February 28, 2023. This research uses the Knowledge Discovery in Database (KDD) method. The purpose of this study is to determine public sentiment towards changes in the Jabodetabek Commuter KRL route and to determine the performance evaluation value of Support Vector Machine (SVM) in analyzing public sentiment. This research compares 3 SVM kernels namely RBF kernel, Linear Kernel, and Polynomial Kernel with 3 dataset sharing scenarios (90:10, 80:20, and 70:30) and also compares the effect of using the Synthetic Minority Oversampling Technique (SMOTE) algorithm to handle data imbalance. This research resulted in positive labels totaling 17, neutral labels totaling 184, and negative labels totaling 140. And the best accuracy value was obtained by RBF kernel and Polynomial kernel in scenario 2 (80:20) with the same value of 88.2%.

Keywords: Sentiment Analysis, SVM, KDD, SMOTE

(*) Corresponding Author: fauzapril.duta18144@student.usnika.ac.id, ultach@staff.unsika.ac.id, susilawati.sobur@staff.unsika.ac.id

How to Cite: Sanubari, F. D, Enri, U, & Susilawati, S. (2023). Analisis Sentimen Terhadap Perubahan Rute Krl Commuter Jabodetabek Menggunakan Algoritme Support Vector Machine (Svm). <https://doi.org/10.5281/zenodo.8206986>

PENDAHULUAN

Sejak 19 September 2017, PT KAI Commuter Jabodetabek yang merupakan salah satu anak perusahaan PT Kereta Api Indonesia telah berganti nama menjadi PT Kereta Commuter Indonesia yang mengelola KRL Commuter Jabodetabek dan sekitarnya. KRL Commuter Jabodetabek merupakan salah satu moda transportasi umum yang menjadi pilihan alternatif penduduk ibukota Jakarta dan sekitarnya untuk mengurangi kemacetan di Kawasan Jabodetabek.

Penggunaan KRL Commuter Jabodetabek tentunya memiliki kelebihan dan kekurangannya masing-masing. Salah satu kelebihannya adalah tarif yang murah dibandingkan dengan moda transportasi lain dengan perbandingan jarak tempuh yang lebih jauh. Adapun kekurangan penggunaan moda transportasi ini adalah penuh dan harus

berebut dan berbagi tempat duduk ketika jam berangkat dan pulang kerja serta ketika akhir pekan. (Romadloni, N. T., Santoso, I., & Budilaksono, S., 2019).

Namun semenjak ada perubahan rute KRL Commuter Jabodetabek pada tanggal 28 Mei 2022, Anne Purba selaku *Vice President Corporate Secretary* KAI Commuter menyebut wajar timbul pendapat yang pro dan kontra dari kalangan masyarakat pengguna moda transportasi umum KRL Commuter Jabodetabek terhadap perubahan rute ini.

Penumpang KRL dari Bogor yang akan menuju Tanah Abang merasa kesulitan dengan perubahan rute baru ini karena sebelumnya tidak harus transit di Stasiun Manggarai dan harus menunggu hingga 30 menit. Berbeda dengan penumpang dari bogor, penumpang KRL dari bekasi yang akan menuju ke Tanah Abang merasa diuntungkan dengan perubahan rute ini karena tidak perlu transit seperti sebelumnya.

Analisis sentimen merupakan proses mengekstraksi, mengolah, dan memahami secara otomatis data teks yang tidak terstruktur untuk mendapatkan informasi sentimen yang terkandung dalam opini atau pernyataan opini (Brahimi, Touahria, & Tari, 2019).

Penelitian di bidang analisis sentimen mulai marak dilakukan mulai tahun 2002. Pada penelitian sebelumnya, Aniq & M. Khairul (2019) menghasilkan nilai akurasi sebesar 53,88%, nilai Recall sebesar 49,69%, nilai Precision sebesar 48,77%, Classification Error sebesar 46,12 % dan F-measure adalah 49,23%. Kemudian Arsyah, Imam & Putra (2019) menghasilkan nilai accuracy sebesar 40%, precision sebesar 40%, recall sebesar 100%, dan f-measure sebesar 57,14%.

Penelitian sebelumnya Dedi, Eka, & A. Ferico (2020), menghasilkan nilai akurasi sebesar 82% dan menghasilkan label negatif lebih besar dibandingkan dengan label positif dengan jumlah 77%. Kemudian, Auliya, Adam, Debby, & Nurman (2021) menghasilkan nilai accuracy sebesar 74%, precision sebesar 75%, recall sebesar 92% dan FI-Score sebesar 83%.

Bedasarkan penelitian sebelumnya, penggunaan algoritme Support Vector Machine mampu menghasilkan nilai accuracy yang cukup optimal untuk analisis sentimen tergantung pada banyaknya data dan metode apa yang akan dipakai. Menurut Carventes, Lemont, Mazahua & Lopez (2019) Support Vector Machine (SVM) merupakan algoritme klasifikasi dan juga regresi yang dianggap paling kuat dan memberikan hasil yang optimal terutama untuk klasifikasi pengenalan pola.

LANDASAN TEORI

Analisis Sentimen

Analisis sentimen adalah kombinasi dari data mining dan text mining, atau metode yang digunakan untuk memproses perbedaan pendapat konsumen atau pakar tentang suatu produk, layanan, atau agensi. (Romadloni, N. T., Santoso, I., & Budilaksono, S., 2019).

Imbalance Data

Imbalance data adalah suatu keadaan distribusi kelas data yang tidak seimbang. Jumlah (*instance*) satu kelas data lebih besar atau lebih kecil dari jumlah kelas data lainnya. Kelompok kelas data minoritas (*minority*) adalah kelompok kelas data yang lebih kecil, sedangkan kelompok mayoritas (*majority*) adalah kelompok kelas data yang lebih besar. *Imbalance data* merupakan masalah utama pada klasifikasi dalam bidang *machine learning* dan *data mining* (Siringoringo R, 2018).

Synthetic Minority Oversampling Technique (SMOTE)

Synthetic Minority Oversampling Technique (SMOTE) merupakan salah satu metode modifikasi dari *oversampling* yang dapat digunakan untuk mengatasi

permasalahan imbalanced dataset. Pendekatan ini pertama kali dikemukakan oleh Chawla pada tahun 2002, dan memerlukan "data training sintetis" yang dibuat untuk mengambil sampel berlebih dari kelas minoritas (Arifiyanti, A. A., & Wahyuni, E. D., 2020).

Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu algoritme klasifikasi yang ampuh karena mampu membuat keputusan antara dua kelas yang memungkinkan prediksi label dari satu atau lebih fitur vektor. SVM pertama kali diusulkan oleh Vapnik pada tahun 1993 untuk membangun pengklasifikasian linier (Romadoni, F., Umaidah, Y., & Sari, B. N, 2020).

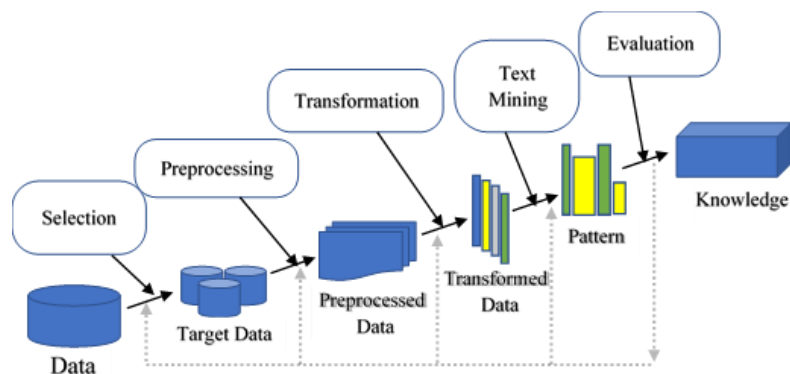
KRL Jabodetabek

KRL Jabodetabek adalah salah satu moda transportasi umum menyediakan rute komuter di wilayah DKI Jakarta, Kota Bogor, Kabupaten Bogor, Kota Bekasi, Kabupaten Bekasi, Kota Tangerang dan Kota Tangerang Selatan. KRL Jabodetabek dinilai mampu mengurangi kemacetan lalu lintas karena mengurangi jumlah kendaraan yang ada (Romadloni, N. T., Santoso, I., & Budilaksono, S., 2019).

METODOLOGI

Metodologi yang digunakan pada penelitian ini adalah Knowledge Discovery in Database (KDD). KDD adalah proses iteratif dan bersifat interaktif yang akan membantu pengguna untuk membuat keputusan. KDD dapat digunakan untuk menggambarkan garis besar proses ekstraksi pada database yang besar (Raka, P. E., Nina S., & M. Jajuli, 2021). Tahapan KDD dapat dilihat pada Gambar 1.

Gambar 1 Tahapan KDD



Data Selection

Proses ini merupakan proses pemilihan data yang diperlukan dengan kebutuhan untuk penggalian informasi pada tahapan KDD. Data hasil seleksi ini digunakan untuk tahap selanjutnya.

Preprocessing

Tahap preprocessing merupakan tahapan yang perlu dilakukan sebelum data mining dapat dilakukan. Tahap preprocessing untuk analisis sentimen diantaranya:

1. Cleansing

proses ini menghapus karakter yang tidak perlu dari data, seperti tanda baca, angka, emoji, dan tautan, untuk mengurangi jumlah karakter yang tidak diperlukan untuk analisis sentimen.

2. Case Folding

Proses ini untuk menyeragamkan keseluruhan teks dalam dokumen menjadi *lowercase* (huruf kecil) karena penggunaan huruf kapital sering tidak konsisten di semua dokumen teks.

3. *Tokenizing*

Tokenizing merupakan proses memecah setiap kata dalam kalimat menjadi satuan kata. Biasanya acuan pembatas antar kata adalah spasi dan tanda baca.

4. *Normalization*

Proses ini untuk memperbaiki kata-kata singkatan dan *slang word* menjadi sebuah kata baku dalam Bahasa Indonesia. Tahap ini penting karena setiap tweet yang dikirim pengguna banyak yang menggunakan kata-kata singkatan *slang word*.

5. *Stopword Removal*

Proses ini merupakan proses untuk menghapus kata-kata yang dianggap tidak mempunyai arti. Proses ini berfungsi untuk mengurangi dimensi teks masukan agar proses lebih ringan.

6. *Stemming*

Proses ini bertujuan untuk mengembalikan berbagai kata menjadi bentuk kata dasar dengan menghilangkan imbuhan.

Transformation

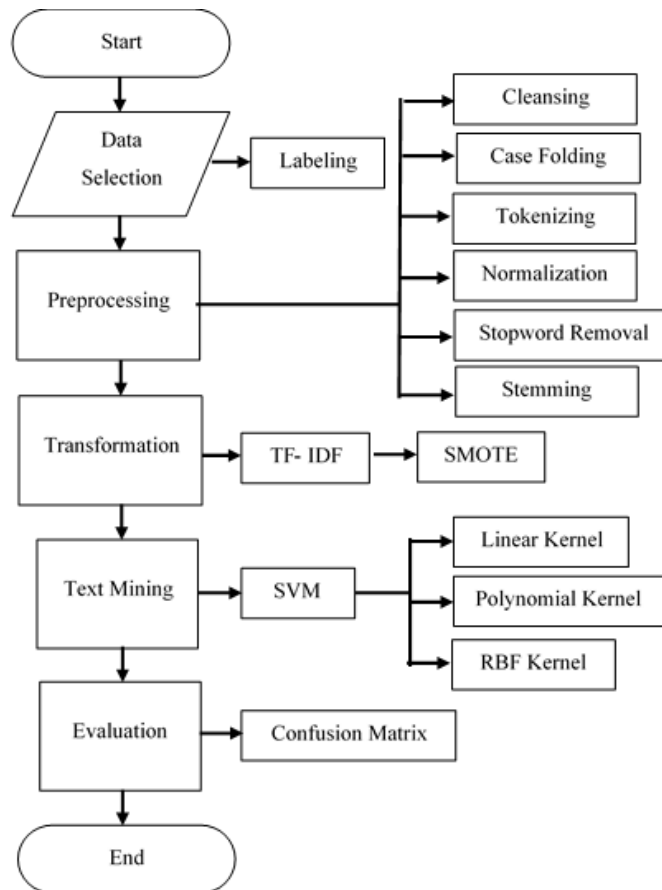
Proses *transformation* ini menggunakan Algoritme *Term Frequency – Inverse Document Frequency* (TF-IDF). Algoritme *Term Frequency -- Invers Document Frequency* (TF-IDF) adalah proses di mana setiap kata diberi bobot untuk menilai seberapa dekat hubungannya dengan dokumen (Meylita, P., S., & Dito, P., U., 2019).

Text Mining

Text mining adalah praktik menganalisis teks untuk mendapatkan informasi yang relevan untuk tujuan tertentu dengan menerapkan konsep dan teknik data mining untuk menemukan pola dalam teks (Meylita, P., S., & Dito, P., U., 2019). Tujuan utama *text mining* adalah memproses data teks sehingga memudahkan manusia untuk mengetahui informasi tidak terstruktur pada teks (Thalita., 2021).

Evaluation

Evaluation merupakan proses pengukuran performa akhir dari model yang telah dibuat. Pada tahap ini, peneliti menggunakan *Confusion Matrix*. Rancangan penelitian yang akan dilakukan pada penelitian ini dapat dilihat pada gambar 2



Gambar 2. Rancangan Penelitian

HASIL DAN PEMBAHASAN

Data Selection

Pada tahap ini, dilakukan pengambilan data tweet dari media sosial Twitter. Data yang diambil menggunakan teknik *scrapping* dengan bantuan *package* Twint dengan *tools* JupyterLab. Data yang diambil merupakan tweet dengan kata kunci “perubahan rute krl” dengan rentang waktu dari tanggal 26 Mei 2022 sampai 28 Februari 2023. Selanjutnya, peneliti menghapus kolom *username* commuterline yang mana kolom tersebut merupakan tweet dari admin commuterline. Sedangkan pada penelitian ini, peneliti fokus pada tweet para pengguna KRL Commuter Jabodetabek. Data yang tersisa selanjutnya masuk ke tahap labeling. Untuk pelabelan data dilakukan secara manual dan dibagi menjadi 3 kelas yaitu positif, netral, dan negatif. Contoh hasil pelabelan data dapat dilihat Tabel 1.

Tabel 1 Contoh Hasil Pelabelan Data

Data Tweet	Label
Ya allah trims banget dengan perubahan rute krl gue gak perlu transit manggarai buat ke bekasi hshshshs jadi gak capek-capek banget kalau PP :)	Positif
Untuk pekerja yang merasakan dampak perubahan rute KRL ini, semoga diberi kekuatan lebihðŸ™† cc: @hrdbacot https://t.co/8nkWvkw8tS	Netral
#KRL #CommuterLine Setelah berjalan beberapa bulan selayaknya @CommuterLine @KAI121 perlu meninjau kembali perubahan rute	Negatif

#KRL Bogor - Kampung Bandan dan sebaliknya, rute Cikarang - Stasiun Kota dan sebaliknya. Transit pindah #KRL di stasiun manggarai menyusahkan bagi lansia	
---	--

Perbandingan hasil pelabelan dapat dilihat pada Tabel 2.

Tabel 2 Perbandingan Hasil Pelabelan

Label	Jumlah
Positif	17
Netral	184
Negatif	140

Berdasarkan Tabel 2, terdapat masalah yaitu *imbalance data*. Penanganan masalah tersebut dijelaskan pada tahap *Transformation*.

Preprocessing

Tahap preprocessing bertujuan untuk memperbaiki data mentah hasil data selection menjadi data yang dapat digunakan. Pada tahap ini terdapat 6 proses yaitu:

1. Cleansing

Contoh proses *cleansing* adalah sebagai berikut:

<i>Before</i>	<i>After</i>
Kusedih karna perubahan rute krl ini. Bikin makin banyak transit :(transit tuh capek @InfoKRL @CommuterLine	Kusedih karna perubahan rute krl ini Bikin makin banyak transit transit tuh capek InfoKRL CommuterLine

2. Case Folding

Contoh proses *case folding* adalah sebagai berikut:

<i>Before</i>	<i>After</i>
Kusedih karna perubahan rute krl ini Bikin makin banyak transit transit tuh capek InfoKRL CommuterLine	kusedih karna perubahan rute krl ini bikin makin banyak transit transit tuh capek infokrl commuteline

3. Tokenizing

Contoh proses *tokenizing* adalah sebagai berikut:

<i>Before</i>	<i>After</i>
kusedih karna perubahan rute krl ini bikin makin banyak transit transit tuh capek infokrl commuteline	“kusedih” “karna” “perubahan” “rute” “krl” “ini” “bikin” “makin” “banyak” “transit” “transit” “tuh” “capek” “infokrl” “commuteline”

4. Normalization

Contoh proses *normalization* adalah sebagai berikut:

<i>Before</i>	<i>After</i>
“kusedih” “karna” “perubahan” “rute” “krl” “ini” “bikin” “makin” “banyak” “transit” “transit” “tuh” “capek” “infokrl” “commuteline”	“kusedih” “karena” “perubahan” “rute” “krl” “ini” “buat” “semakin” “banyak” “transit” “transit” “itu” “capek” “infokrl” “commuteline”

5. Stopword Removal

Contoh proses *stopword removal* adalah sebagai berikut:

<i>Before</i>	<i>After</i>
“kusedih” “karena” “perubahan” “rute” “krl” “ini” “buat” “semakin” “banyak” “transit” “transit” “itu” “capek” “infokrl” “commuteline”	“kusedih” “karena” “perubahan” “rute” “krl” “banyak” “transit” “transit” “capek” “infokrl” “commuteline”

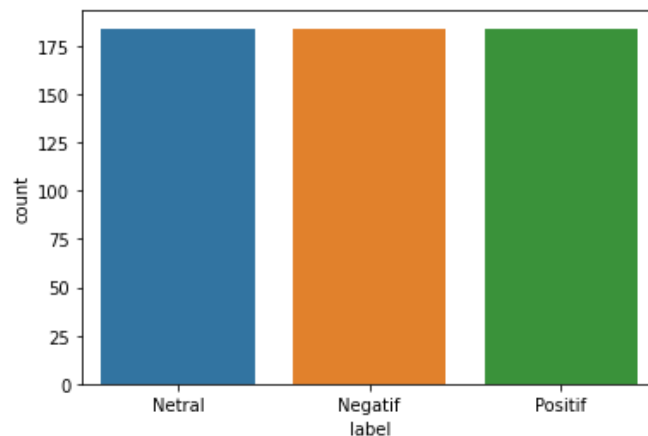
6. Stemming

Contoh proses *stemming* adalah sebagai berikut:

<i>Before</i>			<i>After</i>		
“kusedih”	“karena”	“perubahan”	“kusedih”	“karena”	“ubah”
“rute”	“krl”	“banyak	“krl”	“banyak	“transit”
“transit”	“capek”	“infokrl”	“capek”	“infokrl”	“commuterline”
“commuterline”					

Transformation

Pada tahap ini data diubah menjadi bentuk vector menggunakan algoritme *Term Frequency -- Inverse Document Frequency* (TF-IDF) agar data dapat diolah untuk tahapan selanjutnya. Selanjutnya, peneliti melakukan penanganan masalah imbalance data menggunakan algoritme *Synthetic Minority Oversampling Technique* (SMOTE).



Hasil penggunaan algoritme *Synthetic Minority Oversampling Technique* (SMOTE) dapat dilihat pada gambar 3.

Gambar 3. Hasil penggunaan algoritme SMOTE

Berdasarkan Gambar 3, penggunaan algoritme *Synthetic Minority Oversampling Technique* (SMOTE) dapat merubah hasil pelabelan yang sebelumnya label positif berjumlah 17, label netral berjumlah 184, dan label negatif berjumlah 140 menjadi label positif, label netral dan label negatif berjumlah sama besar yaitu 184.

Text Mining

Pada tahap ini data diklasifikasikan berdasarkan sentimen negatif, netral, dan positif dengan menggunakan 3 skenario pembagian dataset yaitu 90:10, 80:20, 70:30 dan 3 kernel *Support Vector Machine* yaitu *Kernel Linear*, *Kernel RBF*, dan *Kernel Polynomial*. Tahap ini juga membandingkan pengaruh penggunaan algoritme *Synthetic Minority Oversampling Technique* (SMOTE).

Evaluation

Tahap *evaluation* bertujuan untuk mengukur performa model yang diimplementasikan dengan *confusion matrix* dengan parameter *accuracy*, *precision*, *recall*, *f-measure*. Hasil perbandingan akurasi pada penelitian ini dapat dilihat pada Tabel 3.

Tabel 3 Hasil Perbandingan Akurasi

Skenario	Kernel SVM	Akurasi	
		Sebelum SMOTE	Setelah SMOTE
Skenario 1 (90:10)	Kernel RBF	60%	85,7%
	Kernel Linear	68%	87,5%
	Kernel Polynomial	60%	85,7%

Skenario 2 (80:20)	Kernel RBF	57,9%	88,2%
	Kernel Linear	71%	85,5%
	Kernel Polynomial	57,9%	88,2%
Skenario 3 (70:30)	Kernel RBF	57%	86,7%
	Kernel Linear	72,8%	86,7%
	Kernel Polynomial	57%	86,7%

Berdasarkan Tabel 3, hasil akurasi terbaik didapati oleh RBF Kernel dan Polynomial Kernel pada skenario 2 (80:20) dengan nilai yang sama besar yaitu 88,2%. Hasil *confusion matrix RBF kernel* dengan skenario 2 dapat dilihat pada Tabel 4.

Tabel 4 *Confusion Matrix RBF Kernel* dengan Skenario 2

	Sebelum SMOTE			Setelah SMOTE		
	Negatif	Netral	Positif	Negatif	Netral	Positif
Negatif	6	26	0	32	10	0
Netral	0	34	0	3	35	0
Positif	0	3	0	0	0	31

Hasil *confusion matrix polynomial kernel* dengan skenario 2 dapat dilihat pada Tabel 5.

Tabel 5 *Confusion Matrix Polynomial Kernel* dengan Skenario 2

	Sebelum SMOTE			Setelah SMOTE		
	Negatif	Netral	Positif	Negatif	Netral	Positif
Negatif	6	26	0	32	10	0
Netral	0	34	0	3	35	0
Positif	0	3	0	0	0	31

KESIMPULAN

Cara memperoleh sentimen publik adalah dengan proses scrapping menggunakan *library* Twint pada media sosial Twitter dan menghasilkan 355 data tweet. Setelah data terkumpul, kolom dan baris yang tidak relevan dihapus dan menghasilkan 341 data. Selanjutnya 341 data tersebut memasuki tahap labeling yang dilakukan dengan bantuan seorang ahli Bahasa Indonesia untuk menjadi validator. Kemudian Dari 18 kali percobaan klasifikasi yang telah dilakukan, dengan menggunakan 3 skenario pembagian dataset dengan 3 kernel SVM dan dengan membandingkan pengguna SMOTE didapati hasil akurasi terbaik oleh *RBF kernel* dan *Polynomial Kernel* pada skenario 2 (80:20) dengan nilai yang sama besar yaitu 88,2%. Hasil sentimen publik terhadap perubahan rute KRL Commuter Jabodetabek menggunakan data dari sosial media Twitter menghasilkan 17 label positif, 184 label netral dan 140 label negatif dari total 341 data.

REFERENCES

- Arifiyanti, A. A., & Wahyuni, E. D. (2020). SMOTE: Metode penyeimbang kelas pada klasifikasi data mining. *Scan: Jurnal Teknologi Informasi dan Komunikasi*, 15(1), 34-39
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147.
- Aulia, P. M. T. (2021) Summarization Teks Dengan Teknik Ekstraksi Pada Artikel Berbahasa Indonesia Menggunakan Support Vector Machine.

- BRAHIMI, B., TOUAHRIA, M. & TARI, A., 2019. Improving sentiment analysis in Arabic: A combined approach. *Journal of King Saud University - Computer and Information Sciences*. [online] Available at: <<https://doi.org/10.1016/j.jksuci.2019.07.011>>
- Carventes, j., Lamont, F. G., Mazahua, L. R., & Lopez, A. (2019). A comprehensive survey on support vector machine classification: application challenges and trends. *Neurocomputing*, 408, 189-215
- Deolika, A., Kusriani, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. (*JurTI*) *Jurnal Teknologi Informasi*, 3(2), 179-184.
- Eshardiansyah, R. P., Sulistiyowati, N., & Jajuli, M. (2021). Algoritme C4. 5 Untuk Klasifikasi Jenis Kekerasan pada Anak (Kasus DP3A Kabupaten Karawang). *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 5(2), 687-696.
- Fitriyyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 5(3), 279-285.
- Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer genomics & proteomics*, 15(1), 41-51.
- Muhajir, I. A., Yusuf, D., & Hannie, H. (2022). Analisis Hubungan Popularitas Studio Animasi Dengan Anime Menggunakan Metode Pengambilan Data Web Scraping Pada Situs Myanimelist. *Net. Jurnal Ilmiah Wahana Pendidikan*, 8(16), 258-275.
- Praghakusma, A. Z., & Charibaldi, N. Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus: Komisi Pemberantasan Korupsi). *Jurnal Sarjana Teknik Informatika ISSN*, 2338(5197), 33.
- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Riandari, F., & Simangunsong, A. (2019). Penerapan algoritme c4. 5 untuk mengukur tingkat kepuasan mahasiswa. *Terakreditasi Dikti*, 3(2), 1-7.
- Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *ikraith-informatika*, 3(2), 1-9.
- Simatupang, M. P., & Utomo, D. P. (2019). Analisa Testimonial Dengan Menggunakan Algoritme Text Mining Dan Term Frequency-Inverse Document Frequency (Tf-Idf) Pada Toko Allmeeart. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1).
- Siringoringo, R. (2018). Klasifikasi data tidak seimbang menggunakan algoritme SMOTE dan k-nearest neighbor. *Journal Information System Development (ISD)*, 3(1).
- WANG, Q., LIU, K. & MA, K., 2019. Emotional Analysis of Public Opinions in Colleges and Universities: Based on Naive Bayesian Classification Method. *Journal of Physics*
- Zhao, Bo, 2017, *Web scraping*, Springer International Publishing AG (outside the USA) 2017.