

TUGAS AKHIR

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
KEBIJAKAN PENYELENGGARA SISTEM ELEKTRONIK (PSE)
KOMINFO MENGGUNAKAN METODE SUPPORT VECTOR
MACHINE (SVM)**



Oleh:

Muhammad Fauzan Marzuki

1923250068

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER DAN REKAYASA
UNIVERSITAS MULTI DATA PALEMBANG
PALEMBANG
2023**

Fakultas Ilmu Komputer dan Rekayasa
Universitas Multi Data Palembang

Program Studi Informatika
Tugas Akhir Sarjana Komputer
Semester Gasal Tahun 2022/2023

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KEBIJAKAN
PENYELENGGARA SISTEM ELEKTRONIK (PSE) KOMINFO
MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)**

Muhammad Fauzan Marzuki

1923250068

Abstrak

Kebijakan PSE (Penyelenggara Sistem Elektronik) yang diberlakukan oleh KOMINFO (Kementerian Komunikasi dan Informatika Republik Indonesia) menjadi bahan perbincangan yang ramai di Indonesia, khususnya di jejaring sosial *Twitter*. Kebijakan aturan pendaftaran PSE ini sendiri memiliki berbagai pro dan kontra semasa pemberlakuan dilaksanakan, tanggapan masyarakat terhadap PSE Kominfo ini sendiri ini sendiri cukup beragam di media sosial *Twitter*. Data yang digunakan pada penelitian ini adalah sebanyak 1000 *tweet*. Penelitian ini menggunakan ekstraksi fitur *TF-IDF* dengan pembagian data menjadi 70% data latih dan 30% data uji dan menggunakan *K-Fold Cross Validation* dengan nilai $k = 5$. Pada skenario pertama hasil akurasi yang didapat sebesar 82,3% dalam mengklasifikasi 2 kelas sentimen. Pada skenario kedua setelah dilakukan *over-sampling SMOTE* untuk menyeimbangkan data didapatkan akurasi model sebesar 96,4% dalam mengklasifikasi 2 kelas sentimen. Dari kedua skenario pengujian yang telah dilakukan, menggunakan *over-sampling SMOTE* dapat meningkatkan hasil akurasi dengan sangat baik pada model yang digunakan yaitu *Support Vector Machine*.

Kata kunci: PSE, *Support Vector Machine*, Kominfo, *Twitter*, Sentimen.



BAB 1

PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, ruang lingkup, tujuan dan manfaat, metodologi penelitian dan sistematika penulisan

1.1 Latar Belakang

Media sosial adalah layanan yang memungkinkan pertukaran informasi dan topik yang berkelanjutan dan luas (Pintoko & Lhaksana, 2018). *Twitter* adalah sebuah jejaring sosial media yang cukup populer di Indonesia.. Berdasarkan laporan *We Are Social* dari *website (dataindonesia.id)*. Indonesia adalah negara yang menggunakan *Twitter* terbesar di dunia, total seluruh pemakai sosial media *Twitter* di Indonesia sendiri mencapai hingga 18,45 juta hingga akhir tahun 2022. Pengguna *Twitter* dapat menulis tentang berbagai topik, mendiskusikan hal terkini, dan mengekspresikan pendapat mereka melalui *Tweet* (Syarifuddin & Muslimin, 2022). Setiap *Tweet* dapat berupa opini, saran, ataupun kritik tentang topik tertentu. Salah satu topik yang ramai dibahas adalah Kebijakan PSE KOMINFO.

Berdasarkan aturan Peraturan MENKOMINFO No. 5 Tahun 2020, Penyelenggara Sistem Elektronik (PSE) merupakan perorangan, termasuk pejabat pemerintah, badan usaha, dan masyarakat umum, yang menyiapkan, mengatur, dan

menyelenggarakan Sistem Elektronik kepada Pengguna Sistem Elektronik untuk kebutuhannya sendiri maupun kebutuhan pihak ketiga. PSE harus terdaftar di *Web PSE* yang dikelola Kementerian Komunikasi dan Informatika Republik Indonesia (PSE Kominfo, 2022).

Kebijakan PSE (Penyelenggara Sistem Elektronik) yang diberlakukan oleh KEMKOMINFO (Kementerian Komunikasi dan Informatika Republik Indonesia) menjadi bahan perbincangan yang ramai di Indonesia, khususnya di jejaring sosial *Twitter*. Kebijakan aturan pendaftaran PSE ini sendiri memiliki berbagai pro dan kontra semasa pemberlakuan dilaksanakan (Faiz Noorhadi D, 2022), pihak masyarakat yang kontra dengan kebijakan aturan dari KOMINFO mengenai PSE menilai bahwa dengan memblokir situs seperti *Steam*, *Stack Overflow* dan *Paypal* adalah tindakan sewenang-wenang, melanggar hukum dan merugikan banyak masyarakat terkhusus dari kalangan menengah ke bawah. Pemblokiran itu dilakukan dengan alasan bahwa aplikasi atau situs PSE tersebut tidak terdaftar di web (*pse.kominfo.go.id*), berdasarkan Peraturan MENKOMINFO Nomor 5 Tahun 2020. Sedangkan masyarakat yang pro terhadap kebijakan PSE itu sendiri juga datang dari berbagai sektor, banyak yang meyakini bahwa kebijakan PSE ini dapat menjadi langkah awal untuk menjaga kedaulatan digital serta keamanan masyarakat dari beraham kasus penipuan yang ada di Indonesia.

Pada penelitian terkait yang dilakukan oleh Primandani Arsi, Retno Waluyo mengenai Analisis Sentimen Wacana Pemandangan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine* (SVM). Dalam penelitian ini *dataset* yang

digunakan sebanyak 1.236 yang di split hingga 90% data *training* dan 10 % data *testing* (9:1). Kategori positif dan negatif digunakan untuk mengkategorikan informasi teknis. Pengujian terhadap 1.236 *tweet* (404 positif dan 832 negatif) dari platform media sosial *Twitter* tentang pemindahan ibu kota menghasilkan tingkat akurasi 96,68%, tingkat presisi 95,82%, tingkat *recall* 94,04%, dan tingkat *AUC* 94,04%. (*Area Under Curve*) 0,979.

Kemudian pada penelitian terkait Analisis Sentimen Layanan Indihome Berbasis Twitter Menggunakan Metode Klasifikasi *Support Vector Machine* (SVM) Oleh Rian Tinages, Agung Triayudi dan Ira Diana Sholihati. Dalam penelitian ini diambil 10.000 *tweet* yang dijadikan sebagai *dataset* yang dibagi hingga 80% data *training* (1120 *tweet*) dan 20 % data *testing* (280 *tweet*), kemudian dihasilkan 1400 data sebagai hasil tahap penyusunan teks (542 *tweet* positif dan 858 *tweet* negatif). Saat digunakan untuk analisis sentiment layanan Indihome, metode *Support Vector Machine* memberikan hasil dengan akurasi 87%, presisi antara hasil yang diantisipasi dan data aktual sebesar 86%, tingkat perolehan Kembali 95%, dan metrik menguntungkan lainnya. Dan tingkat kesalahan sebesar 13% untuk semua data prediksi. Dan rata-rata nilai rata-rata presisi dan perbandingan daya ingat (*f1-score*) sebesar 90%.

Dan pada penelitian terkait yang dilakukan oleh Hendry Cipta Husada dan Adi Suryaputra Paramita Seputar Analisis Sentimen Maskapai di Platform *Twitter* Menggunakan Algoritma *Support Vector Machine* (SVM). Dalam penelitian ini diambil 14.000 *tweet* yang kemudian dilakukan proses *sampling* menjadi 2000 *tweet*

yang akan di split hingga 90% data *training* dan 10 % data *testing*. *Kernel RBF* menghasilkan hasil terbaik, 84,37%, dengan hasil kompleksitas 10 dan nilai *gamma* 1. Nilai akurasi rata-rata 80,41% ditemukan saat melakukan validasi silang 10 kali lipat. *Kernel RBF* hasil kompleksitas 10 dan *gamma* 1 juga memberikan nilai *precision*, *recall*, dan *f-score* terbaik, yaitu sebesar 84,33%, 84,67%, dan 84,50%.

Berdasarkan uraian dari penelitian terkait tersebut penulis mengusulkan metode *Support Vector Machine* (SVM) untuk diterapkan pada penelitian ini. Dengan penerapan Algoritma *Support Vector Machine* (SVM) ini diharapkan mendapatkan hasil akurasi yang baik serta dapat diketahui kecenderungan sentimen masyarakat Indonesia Terhadap Kebijakan Penyelenggara Sistem Elektronik (PSE) KOMINFO di media sosial *twitter*.

1.2 Rumusan Masalah

Berdasarkan penjelasan yang telah diberikan pada latar belakang, maka rumusan masalah penelitian ini yaitu:

1. Bagaimana cara pengambilan data (*crawling data*) untuk mendapatkan data latih dan data uji?
2. Bagaimana cara menganalisis dan mengklasifikasi sentimen pengguna *twitter* terhadap kebijakan penyelenggara sistem elektronik (PSE) kominfo menggunakan metode *support vector machine* (SVM) dan *SMOTE*?

1.3 Ruang Lingkup

Ruang Lingkup yang digunakan dalam penelitian ini adalah sebagai berikut :

1. *Dataset* diambil dengan melakukan *Crawling Dataset* menggunakan *API Twitter*.
2. *Crawling dataset* dimulai dari rentang tanggal 1 Juli 2022 – 10 Agustus 2022.
3. *Hashtag* yang digunakan untuk *crawling dataset* adalah #PSEKOMINFO.
4. Analisis sentimen dikelompokkan menjadi Positif dan Negatif.
5. *Dataset* yang digunakan adalah sebanyak 1000 data lalu dibagi menjadi 70% data latih dan 30% data uji.
6. Metode yang digunakan adalah *Support Vector Machine (SVM)* dan *SMOTE*.
7. Menggunakan *Kernel Linier* pada metode *Support Vector Machine (SVM)*.
8. Bahasa pemrograman yang digunakan adalah bahasa pemrograman *Python*.
9. Menggunakan *library* yang berguna sebagai alat bantu pada penelitian ini yaitu: *NLTK* dan *Sastrawi Bahasa Indonesia*.
10. Pelabelan dibantu oleh mahasiswa semester akhir, guru atau dosen bahasa Indonesia dan KBBI.
11. Menggunakan *SMOTE (Synthetic Minority Over-sampling Technique)* untuk menyeimbangkan *dataset* sehingga membuat dataset semakin akurat dan mendapatkan hasil output yang baik.

1.4 Tujuan dan Manfaat

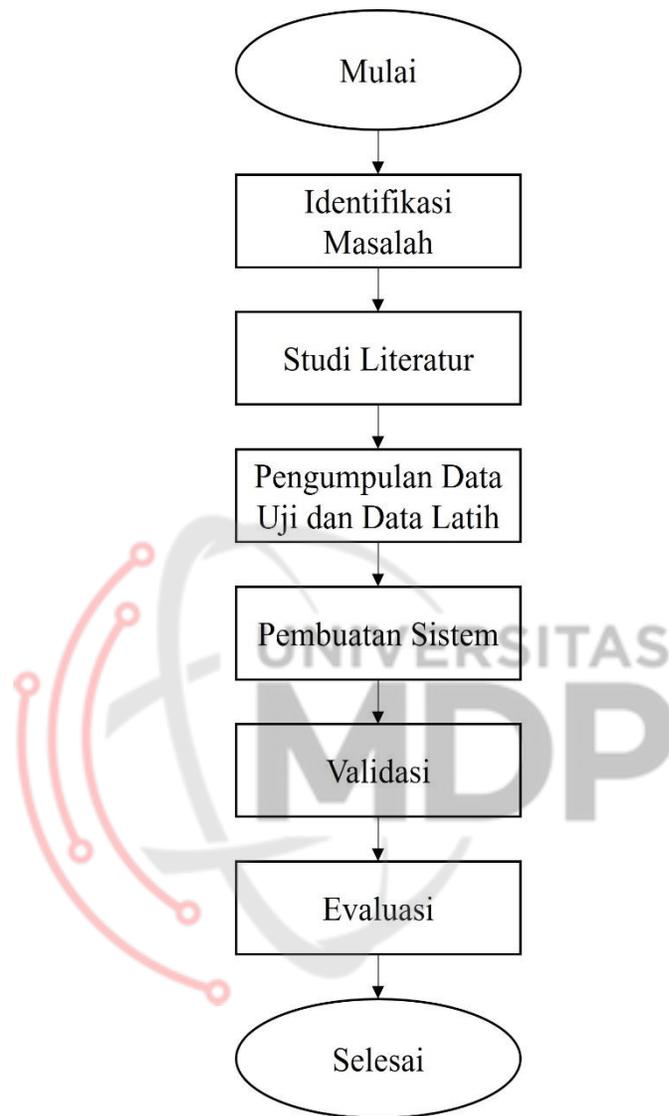
Tujuan penelitian ini adalah mengimplementasikan algoritma *Support Vector Machine* (SVM) dalam menganalisis sentimen pengguna *twitter* terhadap kebijakan penyelenggara sistem elektronik (PSE) KOMINFO.

Manfaat dari penelitian ini antara lain :

1. Memahami *Crawling Dataset* menggunakan *Google Colab*.
2. Memahami penerapan metode algoritma *Support Vector Machine* (SVM) dan *SMOTE*.
3. Mengetahui kemampuan dan tingkat akurasi dari Algoritma *Support Vector Machine* dan *SMOTE* dalam melakukan analisis sentimen.
4. Mengetahui nilai sentimen tergolong positif atau negatif terhadap PSE KOMINFO.
5. Memberikan edukasi kepada masyarakat tentang Kebijakan PSE KOMINFO.

1.5 Metodologi Penelitian

Berikut ini merupakan beberapa tahapan yang dilakukan untuk melakukan Analisis Sentimen Pengguna *Twitter* Terhadap Kebijakan Penyelenggara Sistem Elektronik (PSE) Kominfo Menggunakan Metode *Support Vector Machine* (SVM). Bagan metodologi penelitian dapat dilihat pada gambar 1.1.



Gambar 1.1 Bagan Metodologi Penelitian

1. Identifikasi Masalah

Pada tahap ini dilakukan identifikasi masalah penelitian mengenai Analisis Sentimen Pengguna Twitter Terhadap Kebijakan Penyelenggara Sistem Elektronik PSE Kominfo untuk mendapatkan opini masyarakat Indonesia mengenai Kebijakan PSE Kominfo menggunakan metode *Support Vector Machine (SVM)*.

2. Studi Literatur

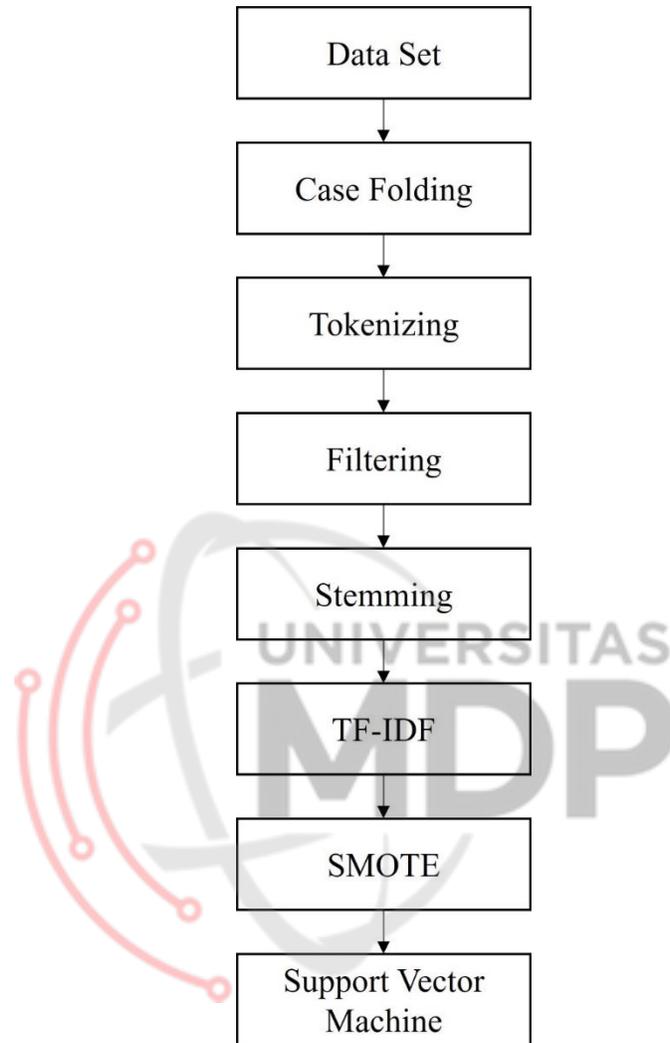
Pada tahap ini dilakukan pembelajaran literatur berupa jurnal-jurnal yang berkaitan dengan topik penelitian ini agar mendapatkan informasi dan data yang dapat digunakan untuk penelitian.

3. Pengumpulan Data Uji dan Data Latih

Pada tahap pengumpulan data dilakukan dengan mengambil data dari *Twitter* menggunakan *Google Colab*. Proses pengambilan data dilakukan dimulai dari tanggal 1 Juli 2022 – 10 Agustus 2022 dikarenakan fenomena dari PSE Kominfo itu sendiri yang masih sangat ramai dibahas dan di bicarakan oleh masyarakat Indonesia melalui *Twitter*. Data yang diambil sebanyak 1000 data. Kemudian dilakukan proses pelabelan dengan dibantu oleh mahasiswa semester akhir, guru atau dosen sastra bahasa Indonesia dan KBBI (YULIANA ROMADHONI, 2022). Setelah dilakukan pelabelan data tersebut dibagi kembali menjadi dua dengan perbandingan 70:30, yaitu data uji dan data latih. Bobot data uji sebanyak 700 data dan bobot data latih sebanyak 300 data.

4. Pembuatan Sistem

Pada tahap pembuatan sistem, setelah dataset diberi pelabelan oleh para ahli kemudian akan dilakukan tahap preprocessing. Tahap preprocessing akan dilakukan *Case Folding, Tokenizing, Filtering, Stemming, TF-IDF, SMOTE dan Support Vector Machine*. Bagan pembuatan sistem dapat pada gambar 1.2.



Gambar 1.2 Bagan Pembuatan Sistem

5. Validasi

Tahap ini untuk mengetahui apakah hasil sistem dapat diandalkan akurat dan konsisten dengan yang diinginkan, Dilakukan validasi model menggunakan metode *K-Fold Cross Validation* pada data latih yang digunakan dalam penelitian. Metode *K-Fold Cross Validation* akan diterapkan selama langkah validasi data latih. Data latih akan dipecah menjadi 5 bagian, nilai yang digunakan adalah $k = 5$. *Five-fold cross*

validation merupakan metode yang paling direkomendasikan untuk estimasi kemampuan model dalam generalisasi data (FUADAH et al., 2022).

6. Evaluasi

Pada tahap ini, sistem yang telah dirancang di evaluasi. Tujuan dari tahap ini adalah untuk menjamin keakuratan hasil yang diperoleh dan mengurangi kesalahan. Nilai *accuracy*, *precision*, *recall* dan *f1-score* dihitung pada tahap evaluasi ini dengan menggunakan pendekatan *Confusion Matrix*. Tabel *confusion matrix* ditunjukkan pada Tabel 1.1.

Tabel 1.1 Tabel Confusion Matrix

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

Jika nilai prediksi positif dan nilai aktual benar maka sentimen termasuk ke *True Positive* (TP). Jika nilai prediksi negatif dan nilai aktual benar maka sentimen termasuk ke *True Negative* (TN). Jika nilai prediksi negatif dan nilai aktual salah

maka sentimen termasuk *False Negative* (FN). Jika nilai prediksi positif dan nilai aktual salah maka sentiment termasuk *False Positive* (FP). Cara perhitungan confusion matrix dapat dilihat pada persamaan 1.1, 1.2, 1.3 dan 1.4.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (1.1)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (1.2)$$

$$Accuracy \text{ Per Class} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1.3)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (1.4)$$

Keterangan :

1. TP : volume dari data dengan sentimen positif yang berhasil diklasifikasikan oleh sistem.
2. TN : volume dari data dengan sentimen negatif yang berhasil diklasifikasikan oleh sistem.
3. FN : volume dari data dengan sentimen negatif yang salah diklasifikasikan oleh sistem.
4. FP : volume dari data dengan sentimen positif yang salah diklasifikasikan oleh sistem.

1.6 Sistematika Penulisan

Sistematika dalam penulisan laporan in adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, ruang lingkup, tujuan dan manfaat, metode penelitian dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini menjelaskan tentang teori-teori yang berkaitan tentang analisis sentimen, Metode *Support Vector Machine*, *SMOTE* dan penelitian terdahulu.

BAB 3 METODE PENELITIAN

Bab ini membahas kebutuhan hardware dan kebutuhan *software* serta metodologi penelitian yang digunakan.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang pengolahan dan pembahasan atas masalah yang telah dirumuskan serta hasil selamat penelitian.

BAB 5 SIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan dan saran yang didapat.



DAFTAR PUSTAKA

- Adam, R. (2019) Mengenal Google Colab - Structilmy, Structilmy. Dari:<https://structilmy.com/blog/2019/05/07/mengenal-google-colab/> (Diakses: 14 November 2022).
- Anwary, A.M., Hadiana, A.I. and Sabrina, P.N., 2021. Analisis Sentiment Penggunaan Vaksin Covid-19 Menggunakan Geo-Tagged Tweets Dan Algoritma Naive Bayes. *Informatics and Digital Expert (INDEX)*, 3(2), pp.75-83.
- Apa itu Twitter & Bagaimana Cara Kerjanya? (2022). Dari: <https://id.eyewated.com/apa-itu-twitter-bagaimana-cara-kerjanya/> (Diakses: 14 November 2022).
- Arsi, P. and Waluyo, R., 2021. Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), p.147.
- Atsqalani, H., Hayatin, N. and Aditya, C.S.K., 2022. Sentiment Analysis from Indonesian Twitter Data Using Support Vector Machine And Query Expansion Ranking. *Jurnal Online Informatika*, 7(1), pp.116-122
- Darwis, D., Pratiwi, E.S. and Pasaribu, A.F.O., 2020. Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, 7(1), pp.1-11.
- Fitriyah, N., Warsito, B. and Di Asih, I.M., 2020. Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM). *Jurnal Gaussian*, 9(3), pp.376-390.
- Fuadah, Yunenda Nur, et al. "Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma." *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika* 10.3 (2022): 728.
- Husada, H.C. and Paramita, A.S., 2021. Analisis Sentimen Pada Maskapai

Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), pp.18-26.

Indonesia, D. (2022) Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022, [Dataindonesia.id](https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022). Dari: <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022> (Diakses: 18 Oktober 2022).

Kominfo (2022). Dari: <https://pse.kominfo.go.id/home> (Diakses: 20 Oktober 2022).

Lukmana, D.T., Subanti, S. and Susanti, Y., 2019, June. Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine Di Twitter. In Seminar & Conference Proceedings of UMT.

Medhat, W., Hassan, A. and Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), pp.1093-1113.

Muhammadi, R.H., Laksana, T.G. and Arifa, A.B., 2022. Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, 8(1).

Pintoko, B.M. and Lhaksana, K.M., 2018. Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier. *eProceedings of Engineering*, 5(3).

Praghakusma, A.Z. and 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)*, p.33.

Pro Kontra Aturan PSE - Jurusan Informatika - Fakultas Teknologi Industri - Universitas Islam Indonesia (2022). Available at: <https://informatics.uui.ac.id/2022/08/15/pro-kontra-aturan-pse/> (Diakses: 15 Oktober 2022).

Ropikoh, Isnin Apriyatin, et al. "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19." *Journal of Applied Informatics and Computing* 5.1 (2021): 64-73.

Rachmi, H. and Surniandari, A., 2020. Classification of Text Mining Review Oil Disfusser Products Using Naive Bayes Classification: Classification of Text Mining Review Oil Disfusser Products Using Naive Bayes Classification. *Jurnal*

Mantik, 4(1), pp.187-192.

Romadhoni, Y., 2022. Klasifikasi Kalimat Perbincangan Masyarakat Terhadap Pandemi Covid-19 Pada Twitter Dengan Metode Long Short-Term Memory (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).

Sianturi, J. and Hajjah, A., 2021. Analisis Sentimen Prosesor AMD Ryzen menggunakan Metode Support Vector Machine. SATIN-Sains dan Teknologi Informasi, 7(2), pp.129-141.

Styawati, S., Hendrastuty, N. and Isnain, A.R., 2021. Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. Jurnal Informatika: Jurnal Pengembangan IT, 6(3), pp.150-155.

Sutoyo, E. and Almaarif, A., 2020. Twitter Sentiment Analysis Of The Relocation Of Indonesia's Capital City. Bulletin of Electrical Engineering and Informatics, 9(4), pp.1620-1630.

Syaifuddin, A. and Muslimin, M., 2022, September. Analisis Sentimen Pada Sosial Media Tentang Implementasi Kebijakan Pse Kominfo Menggunakan Algoritme Lexicon Based. In Seminar Nasional Fakultas Teknik (Vol. 1, No. 1, pp. 7-14).

Tineges, R., Triayudi, A. and Sholihati, I.D., 2020. Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). Jurnal Media Informatika Budidarma, 4(3), pp.650-658.