

KLASIFIKASI SUPPORT VECTOR MACHINE BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK ANALISA SENTIMEN PENGGUNA APLIKASI PEDULILINDUNGI

Astrid Noviriandini¹⁾, Hermanto²⁾, Yudhistira³⁾

¹ Teknologi Komputer, Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Senen, Jakarta Pusat

² Teknologi Komputer, Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Senen, Jakarta Pusat

³ Teknologi Komputer, Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Senen, Jakarta Pusat

Co Responden Email: astrid.asv@bsi.ac.id

Article history

Received 07 January 2022

Revised 31 March 2022

Accepted 14 April 2022

Available online 22 April 2022

Keywords

Covid-19, Google Play,
PeduliLindungi, Review,
Support Vector Machine,
Particle Swarm Optimization.

Riwayat

Diterima 07 Januari 2022

Revisi 31 Maret 2022

Disetujui 14 April 2022

Terbit 22 April 2022

Kata Kunci

Covid-19, Google Play,
PeduliLindungi, Review,
Support Vector Machine,
Particle Swarm Optimization.

Abstract

Covid-19 is an infectious disease that has spread to Indonesia. Monitoring the spread of Covid-19 in Indonesia is handled by the Ministry of Communication and Information (KOMINFO) by creating the PeduliLindungi application which can be found on Google Play. Users will choose applications that have good reviews, but monitoring reviews from the public is not easy, so the author wants to know the analysis of user reviews of the PeduliLindungi application based on user comments using the Support Vector Machine algorithm based on Particle Swarm Optimization. The test results with an accuracy value = 93.0% and AUC value = 0.977. For this reason, the application of the PSO-based Support Vector Machine in this research has a higher accuracy so that it can be used to provide solutions to sentiment analysis problems in reviewing comments from the Pedulilindungi application users on google play.

Abstrak

Covid-19 adalah penyakit menular yang sudah menyebar ke Indonesia. Pemantauan penyebaran Covid-19 di Indonesia ditangani oleh Kementerian Komunikasi dan Informatika (KOMINFO) dengan membuat aplikasi PeduliLindungi yang dapat ditemukan di Google Play. Pengguna akan memilih aplikasi yang memiliki ulasan yang bagus, tetapi untuk memantau ulasan dari masyarakat tidak mudah sehingga penulis ingin mengetahui analisis review pengguna aplikasi PeduliLindungi berdasarkan komentar pengguna dengan menggunakan algoritma Support Vector Machine berbasis Particle Swarm Optimization. Hasil tes dengan nilai akurasi = 93,0% dan nilai AUC = 0,977. Untuk itu, penerapan Support Vector Machine berbasis PSO pada penelitian ini memiliki akurasi yang lebih tinggi sehingga dapat digunakan untuk memberikan solusi terhadap permasalahan analisis sentimen pada review komentar pengguna aplikasi Pedulilindungi di google play.

PENDAHULUAN

Dunia dilanda bala penyakit yang menyerang sistem pernafasan manusia, penyakit ini dikenal dengan nama Covid-19. Covid-19 ditetapkan sebagai wabah yang terjadi dengan melibatkan manusia, karena penyebarannya sudah mencapai sebagian besar Negara yang ada di dunia ini maka wabah tersebut ditetapkan sebagai pandemic dunia pada bulan maret 2020 (Santoso et al., 2021).

Pemantauan penyebaran Covid-19 di Indonesia ditangani oleh Pemerintah Indonesia menggunakan berbagai macam cara

salah satunya yaitu menggunakan aplikasi android yang dipasang Google Play. Aplikasi yang dibuat oleh pemerintah kali ini adalah aplikasi PeduliLindungi yang menjadi wadah penanganan dan partisipasi masyarakat untuk saling membagikan data lokasi saat berpergian agar penelusuran riwayat kontak dengan penderita Covid-19 di Indonesia dapat dilakukan. Hal ini tentunya membuat masyarakat menjadi sering membicarakan pendapatnya melalui aplikasi PeduliLindungi yang terdapat pada Google Play.

Aplikasi PeduliLindungi ini ditetapkan melalui Keputusan Menteri Komunikasi dan Informatika Nomor 171 Tahun 2020

sebagaimana telah diubah dengan keputusan Menteri Komunikasi dan Informatika Nomor 253 Tahun 2020 tentang Perubahan Atas Keputusan Menteri Komunikasi dan Informatika Nomor 171 Tahun 202 tentang penetapan Aplikasi PeduliLindungi Dalam Rangka Pelaksanaan Surveilans Kesehatan Penanganan Corona Virus Disease 2019 (Covid-19).

Sebagai aplikasi yang digunakan dalam pelaksanaan surveilans kesehatan dalam menangani penyebaran Covid-19, maka dari itu aplikasi PeduliLindungi ini harus terus dipertahankan karena sangat membantu masyarakat dalam upaya penelusuran, pelacakan dan pemberian peringatan Covid-19(Eka Putri & Radja Erland, 2021).

Fokus utama penelitian ini adalah melihat review aplikasi PeduliLindungi yang ada di Google Play. Review pengguna aplikasi PeduliLindungi dapat dipengaruhi oleh beberapa hal yang belum menjadi perhatian oleh pihak Kementerian Komunikasi dan Informatika (KOMINFO) sebagai pembuat aplikasi ini. Hal ini memungkinkan terjadi karena adanya beberapa faktor yang harus diperbaiki sebagai informasi pengembangan aplikasi. Opini yang diberikan oleh masyarakat mengenai pelayanan aplikasi PeduliLindungi tentang kepuasan atau berkurangnya masyarakat yang menggunakan layanan ini, maka dengan banyaknya opini-opini yang diberikan membuat masyarakat menjadi lebih selektif dalam memilih penyedia layanan yang baik sehingga dengan adanya hal tersebut pemerintah khususnya KOMINFO mengetahui kepuasan masyarakat yang diberikan sehingga dapat memperbaiki dan meningkatkan kualitas layanan yang akan diberikan kepada masyarakat. Maka dari itu untuk melihat hal tersebut dilakukanlah sebuah analisa sentiment pengguna.

Analisis sentiment merupakan proses menentukan opini seseorang yang diwujudkan dalam bentuk teks dan dikategorikan menjadi sentiment positif atau negatif(Noviriandini et al., 2021). Pengguna internet banyak menuliskan opini dan segala hal yang menjadi perhatian mereka. Opini tentang apa yang mereka rasakan ini dapat berupa perasaan positif, netral ataupun negatif yang dapat diungkapkan dengan cara yang kompleks.

Text Mining dikenal dengan text data mining atau pencarian pengetahuan di basis data textual adalah sebuah proses untuk melakukan pencarian pengetahuan yang berfokus kepada data yang berbentuk dokumen atau teks(Noviriandini et al., 2021).

Support Vector Machine (SVM) adalah metode pembelajaran supervised yang menganalisis data dan mengenali pola untuk klasifikasi dan regensi, SVM bekerja menggunakan cara mencari *hyperplane* terbaik dengan memaksimalkan jarak antar kelas, *hyperplane* merupakan sebuah fungsi yang digunakan untuk pemisah antar kelas. SVM merupakan sebuah pengklasifikasi maka diberi suatu himpunan pelatihan yang ditandai sebagai milik salah satu dari kelas kategori, algoritma SVM membangun sebuah model yang memprediksi apakah data yang baru diproses merupakan salah satu dalam kategori yang lain(Rahman et al., 2021).

Particle Swarm Optimization (PSO) dapat diartikan sebagai kelompok burung mencari makanan disuatu daerah. Burung tersebut tidak tahu dimana makanan terserbut berada, tapi mereka tahu seberapa jauh makanan itu berada, strategi terbaik untuk menemukan makanan tersebut adalah mengikuti burung yang dekat dengan makanan tersebut(Wahyudi et al., 2019).

Beberapa penelitian terdahulu yang digunakan untuk menunjang penelitian ini adalah yang pertama dengan judul “Algoritma Klasifikasi *Naïve Bayes* dan *Support Vector Machine* dalam Layanan Komplain Mahasiswa” dilakukan pengujian model dengan membandingkan dua metode data mining yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes* dapat dibuktikan dengan nilai akurasi dan nilai AUC dari masing-masing algoritma yaitu untuk SVM nilai akurasi = 84,45% dan nilai AUC = 0,922, sedangkan untuk algoritma *Naïve Bayes* nilai akurasi = 69,75% dan nilai AUC = 0,687 (Hermanto et al., 2020). Berdasarkan penelitian terdahulu yang kedua dengan judul “*Analysis of User Reviews for the PeduliLindungi Application on Google Play Using Naïve Bayes Algorithm Based on Particle Swarm Optimization*” bisa dilihat untuk algoritma NB berbasis PSO mendapatkan nilai akurasi = 69,00% dan nilai AUC = 0,659 (Mustopa et al., 2020).

Dari latar belakang yang sudah diuraikan diatas, peneliti akan melakukan analisis review pengguna aplikasi PeduliLindungi berdasarkan komentar pengguna aplikasi tersebut menggunakan metode *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (PSO). Proses pengujian data pada Rapidminer dimulai dengan proses pembentukan model dengan data pada bagian pertama pembagian data training dan data testing. Setelah melakukan pengujian hasil Akurasi dapat diukur menggunakan *Accuracy* dan *AUC* serta akan ditampilkan dalam bentuk kurva ROC, tentunya dapat diketahui aplikasi PeduliLindungi manakah yang memiliki tingkat akurasi paling tinggi sesuai dengan hasil penelitian menggunakan model algoritma *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (PSO).

METODE PENELITIAN

A. Bussines Understanding

Tahapan ini berfokus kepada pemahaman tujuan dari proyek dan kebutuhan secara perspektif bisnis kemudian mengubah pengetahuan ini menjadi definisi masalah penambangan data dan rencana awal yang dirancang untuk mencapai tujuan(Lestari & Mahdiana, 2021). Data yang akan digunakan pada penelitian ini adalah *review* komentar pengguna aplikasi PeduliLindungi pada *google play* yang memberikan komentar, untuk menentukan status *review* komentar pengguna yang ada di *google*.

Pada tahapan ini juga dilakukan pemahaman untuk mencari metode algoritma terbaik agar dapat membantu saat proses pengolahan data yang akan dilakukan dengan cara membandingkan hasil dari algoritma dan *Support Vector Machine* (SVM) pada *review* komentar pengguna aplikasi PeduliLindungi.

B. Data Understanding

Tahapan ini adalah proses memahami data yang akan digunakan sebagai bahan yang akan diteliti untuk bisa dilakukan ke tahap setelahnya yaitu *processing*. Data *review* pengguna PeduliLindungi mengambil data dari *google play*, dimana *review* komentar pengguna diambil sebanyak 1000 data. Dari *review* pengguna tersebut kemudian dikelompokkan berdasarkan kategori

komentar yang diberikan oleh pengguna PeduliLindungi.

C. Data Preparation

Tahap selanjutnya adalah melakukan persiapan data sebelum data dilakukan modelling atau disebut dengan *Data Preparation*. Untuk tahap ini mempersiapkan data untuk melakukan langkah-langkah yang disebut dengan *text preprocessing*. Dalam *Text Preprocessing* terdapat beberapa tahap, yaitu *Case Folding*, *Tokenizing*, *Stopword Removal* dan *Stemming* (Salam et al., 2018).

Pada tahap ini menggunakan dua aplikasi *preprocessing*, yang pertama menggunakan Gata Framework yang diakses melalui link <http://gataframework.com/textmining> yang dapat digunakan secara gratis juga mudah dalam penggunaan dikarenakan tidak harus membuat *account* untuk memakai servicenya dan dilanjutkan *preprocessing* dari Rapidminer, berikut adalah tahapannya :



Sumber:

<http://gataframework.com/textmining>

Gambar 1. Tampilan *tools* Gata Framework Text Mining

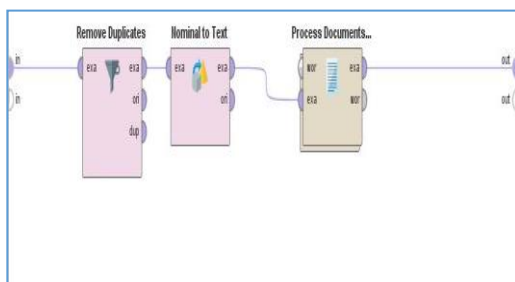
Gata Framework merupakan alternatif dalam pre-processing teks berbahasa Indonesia yang dikombinasikan dengan aplikasi RapidMiner untuk memproses kata-kata dalam bahasa Indonesia, hal ini dikarenakan dalam aplikasi RapidMiner sudah ada fasilitas kamus untuk mengubah akronim, dan stopword, tetapi masih terbatas pada bahasa Inggris, Cina, dan Arab, sedangkan untuk bahasa Indonesia masih belum tersedia. Dari hasil pre-processing dengan menggunakan Gata Framework, maka data set akan dilakukan pre-processing lagi dengan menggunakan *tools* RapidMiner untuk membersihkan data agar lebih baik lagi hasilnya

D. Remove Duplicates

Remove Duplicates merupakan tahapan data preparation selanjutnya yang digunakan pada software rapidminer. Remove duplicates digunakan untuk menghilangkan text yang sama atau duplikat. Hal ini dilakukan agar data tidak dipenuhi oleh text yang sama sehingga memperlambat proses running software untuk menganalisa model.

E. Nominal Text

Nominal to text merupakan operator yang ada dalam rapidminer yang berfungsi untuk mengubah semua angka yang ada dalam text menjadi sebuah text. Sehingga angka yang ada akan dianggap jenis data text bukan numeric atau nominal. Gambar 2. memperlihatkan bagaimana penggunaan operator ini digunakan pada proses yang ada pada rapidminer.



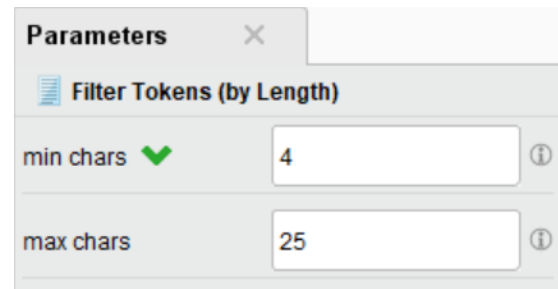
Gambar 2. Tampilan Proses Nominal to Text

F. Transform Case

Operator yang digunakan pada tahapan ini adalah untuk mengubah huruf kapital yang masih ada pada text akan diubah menjadi huruf kecil semua. Hal ini dilakukan agar ketikan dilakukan proses ke dalam model klasifikasi terdapat keseragaman huruf dan tidak terjadi kesalahan dalam proses tokenize.

G. Filter Token (By Length)

Filter Token adalah proses yang ada pada data preparation untuk menghilangkan sejumlah kata (setelah proses tokenize) dengan panjang karakter tertentu. Pada penelitian ini panjang minimum karakter yang digunakan adalah 4 karakter dan panjang maksimum 25 karakter. Artinya kata yang panjangnya kurang dari 4 karakter dan lebih dari 25 karakter akan dihilangkan. Untuk mendapatkan hasil seperti ini maka dilakukan setting pada Parameters dari operator ini.



Gambar 3. Tampilan Proses Filter Token (By Length)

H. Filter Stopword (Dictionary)

Selanjutnya adalah penggunaan operator Stopword Removal (by Dictionary) yang berfungsi untuk menghilangkan kata-kata yang tidak hubungan dengan isi text. Pada tahapan sebelumnya dengan menggunakan service text mining Gata framework telah dilakukan namun ada beberapa kata yang belum dapat bisa dihilangkan oleh service sebelumnya karena belum dimasukkan sebagai kata yang harus dihapus. Maka dengan operator Stopword Removal (by Dictionary) peneliti dapat mendaftarkan kata yang harusnya dihapus dari text.

HASIL DAN PEMBAHASAN

A. Tahapan Pemodelan

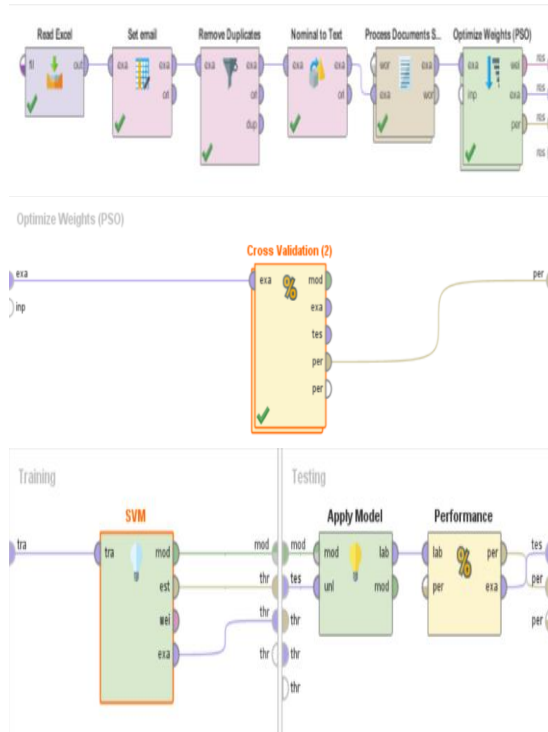
Merupakan tahap pemilihan teknik mining dengan menentukan algoritma yang akan digunakan. Tool yang digunakan adalah RapidMiner versi 9.1. Hasil pengujian model yang dilakukan adalah menganalisa review komentar pengguna Pedulilindungi pada google playstore dengan label data komentar positif dan negatif menggunakan algoritma Naive Bayes dan *Support Vector Machine (SVM)* Berbasis Particle Swarm Optimization untuk mendapatkan nilai akurasi terbaik. Berikut adalah desain model Rapidminer yang digunakan yaitu:

B. Tahapan Pengujian

Pengaturan dan penggunaan operator serta parameter dalam frameworks RapidMiner sangat berpengaruh terhadap akurasi dan model yang terbentuk, secara lebih jelas pengujian terhadap algoritma Naive Bayes dan *Support Vector Machine (SVM)* Berbasis

Particle Swarm Optimization tersebut adalah sebagai berikut:

1. Proses model ini algoritma Naive Bayes dan *Support Vector Machine (SVM)* Berbasis Particle Swarm Optimization dilakukan proses 10 fold validation dapat dilihat pada gambar 4:



Gambar 4. Model Pengujian Algoritma *Support Vector Machine (SVM)*

Gambar diatas adalah model pengujian algoritma *Support Vector Machine (SVM)* Berbasis Particle Swarm Optimization menggunakan rapidminer, diawali dari memasukan data kemudian mengatur set role yang nantinya menentukan label disana dan nominal text lalu keproses dokumen. Pada pengujian ini, data digunakan adalah data bersih yang telah melalui preprocessing. Data tersebut diambil dari operator Read Excel, hal ini dilakukan karena dataset disimpan dalam bentuk Excel (.xlsx). Process documents from files untuk mengkonversi files menjadi dokumen. Process validasi terdiri dari data training dan data testing. Kemudian masuk ke model desain algoritmanya diproses ke dalam operator cross validation *Support Vector Machine (SVM)* Berbasis Particle Swarm Optimization didalamnya ada perhitungan algoritmanya kemudian modelnya diapply

setelah itu masuk ke penilaian performancenya barulah muncul hasil nilai accuracy dan aucnya.

C. Model Evaluasi

Tahapan evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil dibuat pada langkah sebelumnya. Untuk evaluasi digunakan 10-fold cross validation. Dari hasil pengujian model dari algoritma yang dipakai adalah untuk menghasilkan sebuah nilai Accuracy (*Confusion Matrix*) dan AUC (*Area Under Curve*). Maka mendapatkan hasil grafik ROC dengan nilai AUC (*Area Under Curve*).

D. Nilai Accuracy Algoritma *Support Vector Machine (SVM)* Berbasis Particle Swarm

accuracy: 93.00% +/- 3.50% (micro average: 93.00%)

	true Positif	true Negatif	class precision
pred. Positif	181	9	95.26%
pred. Negatif	19	191	90.95%
class recall	90.51%	95.51%	

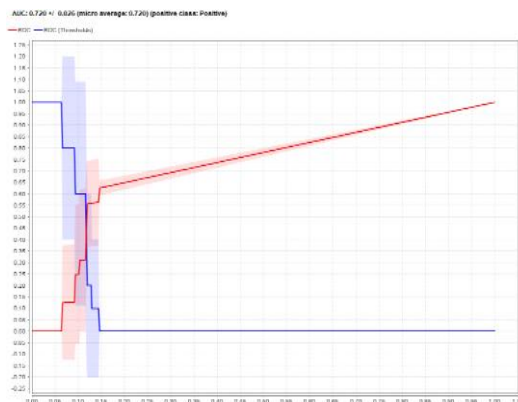
Gambar 5. Nilai Accuracy Algoritma *Support Vector Machine (SVM)* Berbasis Particle Swarm

Jumlah *True positif (TP)* adalah 181 *record* diklasifikasikan sebagai *positif* dan *False positif (FP)* adalah 19 *record* diklasifikasikan sebagai *Negatif*. Berikutnya *True negatif (TN)* 191 diklasifikasikan sebagai *Negatif* dan 9 *record False Negatif (FN)* diklasifikasikan sebagai *Negatif*. Berdasarkan tabel 1 diatas menunjukkan bahwa, tingkat akurasi review pengguna aplikasi Pedulilindungi dengan menggunakan algoritma *Support Vector Machine (SVM)* Berbasis Particle Swarm Optimization adalah sebesar 93,00%.

E. Nilai AUC dari aplikasi Pedulilindungi algoritma *Support Vector Machine (SVM)* Berbasis Particle Swarm Optimization:

1. Berikut Hasil pengujian data review pengguna Pedulilindungi dengan data 1364 menggunakan metode *Support Vector Machine (SVM)* Berbasis

Particle Swarm Optimization menghasilkan AUC = 0.977:



Gambar 6. Hasil pengujian data review pengguna

KESIMPULAN

Pada penelitian ini setelah dilakukan *preprocessing* dan pengujian model dengan menggunakan metode data mining *support vector machine* berbasis *pso*. Dapat diketahui bahwa nilai akurasi untuk menentukan bahwa analisa review komentar pengguna aplikasi Pedulilindungi, dapat dibuktikan dengan nilai akurasi dan nilai *AUC* dari algoritma SVM berbasis PSO nilai akurasi = **93,0%** dan nilai *AUC* = **0.977**. Dalam penelitian dapat diketahui bahwa tingkat akurasi yang didapatkan oleh algoritma Support Vector Machine berbasis PSO, Untuk itu, penerapan Support Vector Machine berbasis PSO pada penelitian ini memiliki akurasi yang lebih tinggi sehingga dapat digunakan untuk memberikan solusi terhadap permasalahan analisis sentimen pada review komentar pengguna aplikasi Pedulilindungi di google play.

Untuk Penelitian lanjutan dengan mengambil jumlah data yang lebih besar lagi dalam kurun waktu lebih dari 3 bulan pada google play, sehingga akan mendapatkan nilai akurasi yang lebih baik lagi, juga Melakukan analisis review pengguna aplikasi berdasarkan jumlah rate bintang 1-5 dan jumlah 'like' yang didapatkan. Selain itu Menambahkan metode klasifikasi lain untuk membandingkan dengan metode dari Supervised Learning lainnya dari model yang diusulkan, seperti penggunaan Algoritma C-45, Random Forest dan lainnya.

REFERENSI

- Eka Putri, C., & Radja Erland, H. (2021). Aplikasi Pedulilindungi Mitigasi Bencana Covid-19 Di Indonesia. *Jurnal Pustaka Komunikasi*, 4(1), 66–78. <https://journal.moestopo.ac.id/index.php/pustakom/article/view/1321%0Ahttps://doi.org/10.32509/pustakom.v4i1.1321>
- Hermanto, H., Mustopa, A., & Kuntoro, A. Y. (2020). Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 211–220. <https://doi.org/10.33480/jitk.v5i2.1181>
- Lestari, D. A., & Mahdiana, D. (2021). Penerapan Algoritma K-Nearest Neighbor pada Twitter untuk Analisis Sentimen Masyarakat Terhadap Larangan Mudik 2021. *Jurnal Informatik*, 17(2), 123–131. <https://ejournal.upnvj.ac.id/index.php/informatik/article/view/3629/1405>
- Mustopa, A., Hermanto, Anna, Pratama, E. B., Hendini, A., & Risdiansyah, D. (2020). Analysis of user reviews for the pedulilindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization. *2020 5th International Conference on Informatics and Computing, ICIC 2020*, 2. <https://doi.org/10.1109/ICIC50835.2020.9288655>
- Noviriandini, A., Bina, U., Informatika, S., & Pusat, J. (2021). ANALISA SENTIMEN TERHADAP BELAJAR ONLINE PADA MASA COVID-19 MENGGUNAKAN ALGORITMA SUPPORT VECTOR. 5(1), 129–136.
- Rahman, O. H., Abdillah, G., & Komarudin, A. (2021). Classification of Hate Speech on Social Media Twitter Using Support Vector Machine. *RESTI Journal (Systems Engineering and Information Technology)*, 5(1), 17–23.
- Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia). *Prosiding SINTAK*, 480–486.

- Santoso, A. K., Noviriandini, A., Kurniasih, A., Wicaksono, B. D., & Nuryanto, A. (2021). Klasifikasi Persepsi Pengguna Twitter Terhadap Kasus Covid-19 Menggunakan Metode Logistic Regression. *JIK (Jurnal Informatika Dan Komputer)*, 5(2), 234–241.
- Wahyudi, I., Bahri, S., & Handayani, P. (2019). *Aplikasi Pembelajaran Pengenalan Budaya Indonesia*. V(1), 135–138.
<https://doi.org/10.31294/jtk.v4i2>