

Optimasi Algoritma Support Vector Machine untuk Analisis Klasifikasi Teks Pemintaan Informasi di Platform Online Shop

Imannudin Akbar¹, Marwondo M. Iqbal², Nugraha³

^{1,2,3}Informatika, Universitas Informatika dan Bisnis, Indonesia

iman@unibi.ac.id

Info Artikel

Sejarah artikel :

Diterima September 2023

Direvisi September 2023

Disetujui September 2023

Diterbitkan September 2023

ABSTRACT

The use of technology in the field of trade and sales is increasingly growing. Product information is an important role in building consumer trust when making purchasing decisions. Therefore, classification analysis is needed to help potential consumers in drawing conclusions. Classification analysis aims to conclude and identify data and classify its polarity. The Support Vector Machine (SVM) algorithm is widely used by many researchers for use in classification analysis. This algorithm was chosen because it can identify separate hyperplane to maximize the margin between 2 different classes. However, the Support Vector Machine (SVM) has deficiencies in parameter selection, so the selection of the Particle Swarm Optimization (PSO) feature is applied to improve accuracy. The results showed that implementation of the Support Vector Machine (SVM) has an accuracy value of 81.48% and an AUC value of 0.825, while optimization using Particle Swarm Optimization (PSO) has an accuracy value of 89.78% and an AUC of 0.902. The application of Particle Swarm Optimization (PSO) has been proven to improve the performance of the Support Vector Machine (SVM) algorithm.

Keywords : Classification Analysis; Information Request Text; Online Store Platform; Particle Swarm Optimization; Support Vector Machine.

ABSTRAK

Pemanfaatan teknologi dalam bidang perdagangan dan penjualan semakin berkembang. Informasi produk berperan penting untuk membangun kepercayaan konsumen ketika menentukan keputusan dalam pembelian produk. Oleh karena itu, diperlukan analisis klasifikasi untuk membantu calon konsumen untuk menarik kesimpulan. Analisis klasifikasi bertujuan untuk menyimpulkan dan mengidentifikasi pada data dan mengklasifikasikan polaritas. Algoritma *Support Vector Machine* (SVM) banyak digunakan oleh banyak peneliti untuk digunakan dalam analisis klasifikasi. Algoritma ini dipilih karena dapat mengidentifikasi *hyperplane* terpisah untuk memaksimalkan margin antara 2 kelas yang berbeda. Akan tetapi, SVM memiliki kekurangan pada pemilihan parameter, maka diterapkan seleksi fitur *Particle Swarm Optimization* (PSO) untuk meningkatkan akurasi. Hasil penelitian menunjukkan bahwa penerapan SVM memiliki nilai akurasi 81.48% dan nilai AUC 0,825, sedangkan setelah dioptimasi menggunakan *Particle Swarm Optimization* (PSO) memiliki nilai akurasi 89.78% dan AUC 0,902. Terbukti bahwa penerapan *Particle Swarm Optimization* (PSO) dapat meningkatkan kinerja algoritma *Support Vector Machine* (SVM).

Kata Kunci : Analisis Klasifikasi; *Particle Swarm Optimization*; Platform Online Shop; *Support Vector Machine*; Teks Pemintaan Informasi.

PENDAHULUAN

Dinamika perekonomian tidak terlepas dari perkembangan serta kemajuan teknologi, infrastruktur, produktifitas dan inovasi. Ilmu pengetahuan yang terus berkembang memberikan dampak terhadap kemajuan teknologi yang semakin canggih. Inovasi yang dimunculkan dari teknologi memberikan kemudahan dan manfaat untuk manusia dalam beraktivitas [1]. Dengan berkembangnya *E-commerce* secara cepat di seluruh dunia, segmentasi *online customer* menarik perhatian pebisnis dan akademisi [2].

Penggunaan *E-commerce* pada saat ini merupakan syarat bagi sebuah organisasi atau perusahaan, agar perusahaan itu dapat bersaing secara global [3]. Dengan pesatnya perkembangan *E-commerce* tersebut membuat banyak orang yang mulai menggunakan *E-commerce* sebagai transaksi jual beli untuk memungkinkan perusahaan mengenali target pasar yang lebih tepat dan mengoptimalkan kampanye pemasaran mereka [2].

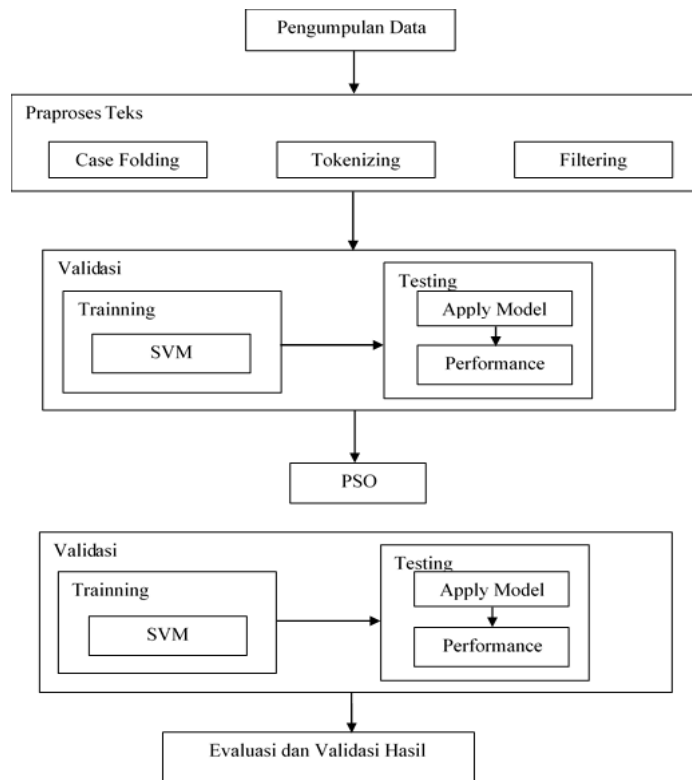
Pada penelitian sebelumnya, sistem yang dikembangkan memiliki nilai yang sama, serta memenuhi nilai *minimum support* dan *minimum confidence* yang ditentukan. Dalam pengaplikasian metode *grid search* untuk mencari akurasi model masih dilakukan secara manual maka pembakuan kata pada corpus tidak memberikan dampak yang signifikan terhadap akurasi dan cenderung menurunkan nilai akurasi [4].

Metode *Support Vector Machine (SVM)* merupakan suatu metode yang menggunakan *machine learning* yang memiliki hasil lebih baik dalam hal klasifikasi [5]. Sedangkan *Particle swarm optimization (PSO)* adalah algoritma bionik populer di mana sekumpulan solusi yang dihasilkan secara acak bergerak di ruang pencarian untuk mendapatkan solusi optimal [6]. Pemilihan *Support Vector Machine (SVM)* dikarenakan algoritma ini mampu mengidentifikasi *hyperplane* terpisah sehingga dapat memaksimalkan margin antara 2 kelas yang berbeda [7]. SVM memiliki kekurangan pada pemilihan parameter atau fitur untuk mempengaruhi hasil akurasi secara signifikan, pada penelitian ini akan diterapkan metode *Particle Swarm Optimization (PSO)* [8]. *Particle Swarm Optimization (PSO)* adalah suatu teknik optimasi yang biasa digunakan untuk menerapkan dan memodifikasi beberapa parameter [9].

Penelitian ini bertujuan untuk mengoptimisasi algoritma *Support Vector Machine (SVM)* menggunakan *Particle Swarm Optimization (PSO)* untuk analisis klasifikasi teks dalam konteks permintaan informasi pada platform *online shop* agar memiliki tingkat akurasi yang dapat dicapai dan juga menganalisis klasifikasi teks permintaan informasi.

METODE

Sesuai dengan judul yang dikemukakan, pada dasarnya penelitian ini menggabungkan langkah-langkah SVM dan PSO. Pengolahan data dilakukan menggunakan perangkat lunak *RapidMiner Studio*. Secara lengkap Langkah-langkah penelitian dijelaskan pada gambar 1 sebagai berikut :



Gambar 1. Metode Penelitian

Berdasarkan gambar di atas dapat dijelaskan sebagai berikut :

1. Pengumpulan Data

Data yang digunakan adalah data pertanyaan informasi pada toko *online* juncoshop. Data yang digunakan merupakan data *query* dan *non query* untuk permintaan informasi produk. Data *query* merupakan data informasi produk dan data *non query* merupakan data bukan informasi produk. Data ini kemudian diolah sehingga dapat digunakan untuk proses klasifikasi teks.

positif,Hai, barang ini ready?,mudah,12,88,0,136363636363636,			
positif,kak Jersey ini ada Pasangan Celananya ga yah?,mudah,23,77,0,298701298701299,			
positif,Hai, barang ini ready?,mudah,12,88,0,136363636363636,			
negatif,gimana ka?,susah,84,16,5,25,			
positif,sudah 1 paket kaos dan celana + no punggung,mudah,23,77,0,298701298701299,			
positif,ada yg polos buat wanita ?,mudah,23,77,0,298701298701299,			

Gambar 2. Contoh Data

2. Praproses Teks

Dalam penelitian ini pengolahan bahasa dilakukan dengan praproses teks untuk menyeragamkan bentuk kata dan mengurangi volume kosakata. Tahap ini terdiri dari:

- a. *Case Folding* yaitu proses dalam mengubah semua huruf dalam teks menjadi huruf kecil. Karakter selain huruf akan dihilangkan.

Tabel 1. Case Folding

Teks Input	Teks Output
Kak Jersey ini ada Pasangan Celananya ga yah?	kak jersey ini ada pasangan celananya ga yah

- b. *Tokenizing* yaitu sebuah proses untuk memisahkan setiap kata dalam suatu kalimat sehingga menghasilkan kumpulan kata-kata yang berdiri sendiri.

Tabel 2. *Tokenizing*

Teks Input	Teks Output
kak jersey ini ada pasangan celananya ga yah	kak jersey ini ada pasangan celananya ga yah

- c. *Filtering* yaitu proses untuk mengambil kata penting dari hasil *tokenizing*. Apabila terdapat kata yang sama, maka kata itu akan dibuang dari koleksi.

Tabel 3. *Filtering*

Teks Input	Teks Output
kak jersey ini ada pasangan celana nya ga yah	kak jersey ada pasangan celana ga yah

3. **Klasifikasi**

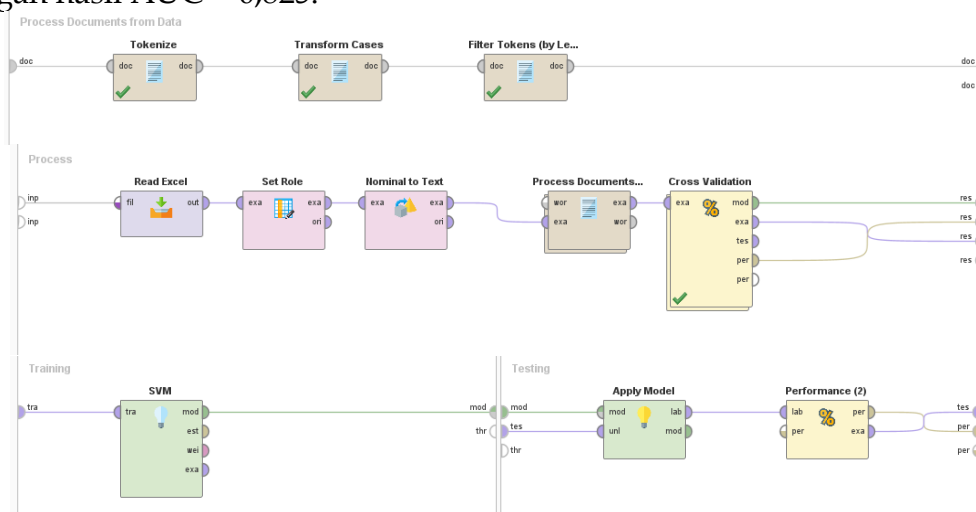
Tahap ini dilakukan klasifikasi pada permintaan informasi produk menggunakan metode *Support Vector Machine (SVM)* menggunakan *Particle Swarm Optimization (PSO)* dengan menggunakan perangkat lunak RapidMiner Studio. Klasifikasi teks yang dilakukan akan dibagi menjadi 2 kategori yaitu kategori *query* dan *non query* dengan hasil berupa model klasifikasi dan akan dibagi menjadi 2 yaitu *data training* dan *data testing*.

HASIL DAN PEMBAHASAN

Berdasarkan metodologi yang telah dipaparkan sebelumnya, berikut hasil penelitian yang dilakukan sesuai urutan langkahnya.

1. *Support Vector Machine (SVM)*

Pada penelitian ini, *dataset* yang digunakan sudah dipraproses sebelumnya menggunakan *RapidMiner Studio*, sehingga *dataset* sudah siap untuk pengolahan data. Bentuk penerapan SVM dalam RapidMiner Studio digambarkan seperti pada gambar 3. Pada pengujian menunjukkan bahwa penerapan *Support Vector Machine (SVM)* dihasilkan *accuracy* sebesar 81.48% dengan hasil $AUC = 0,825$.



Gambar 3. Penerapan Model *Support Vector Machine (SVM)*

a. Confusion Matrix

Berikut merupakan hasil dari penerapan model pada RapidMiner, menunjukkan hasil dari *confusion matrix* pada algoritma *Support Vector Machine* dengan hasil terbaik.

Tabel 4. Confusion Matrix Support Vector Machine

	true negatif	true positif	class precision
pred. negatif	109	23	82.58%
pred. positif	2	1	33.33%
class recall	98.20%	4.17%	

Dengan menggunakan persamaan-persamaan berikut:

Persamaan untuk menghitung nilai accuracy:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

Persamaan untuk menghitung nilai recall:

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

Persamaan untuk menghitung nilai precision:

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

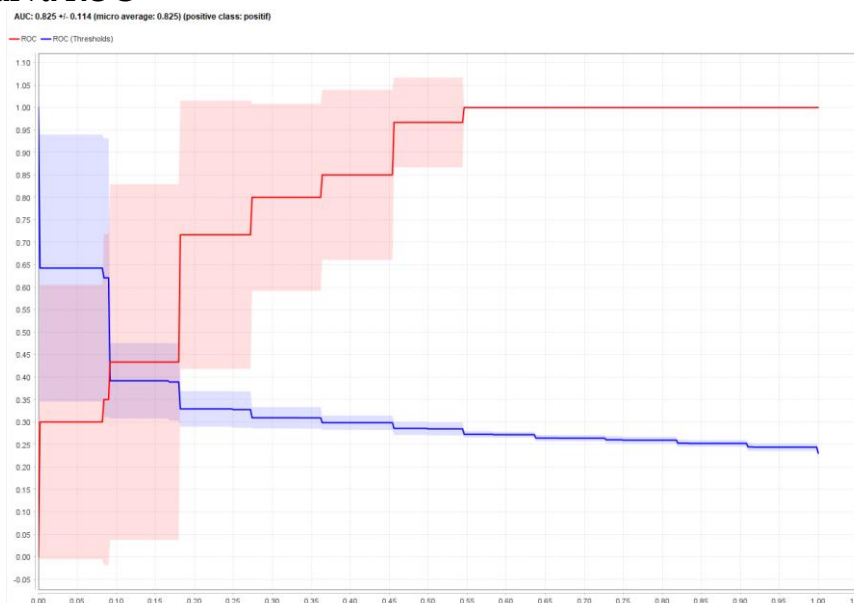
Didapatkan hasil sebagai berikut:

$$Accuracy = \frac{1 + 109}{23 + 109 + 1 + 2} = 81,48\%$$

$$Recall = \frac{1}{1 + 109} = 9,10\%$$

$$Precision = \frac{1}{1 + 2} = 33,33\%$$

b. Kurva ROC



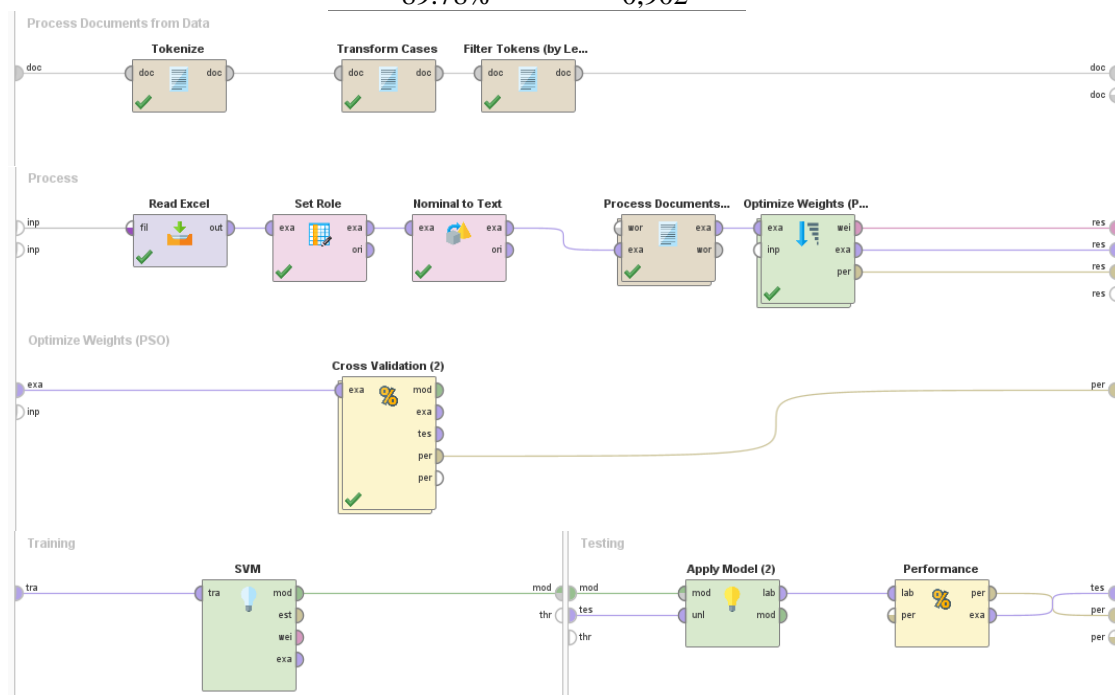
Gambar 4. Kurva ROC Support Vector Machine

Berikut merupakan hasil dari penerapan model pada RapidMiner, mengevaluasi akurasi dan klasifikasi secara visual menggunakan *Support Vector Machine (SVM)* mendapatkan hasil Accuracy 81,48%, Recall 9,10%, Precision 33,33%, dan AUC 0,825.

2. Gabungan *Support Vector Machine* dengan *Particle Swarm Optimization*
 Pada langkah ini, SVM digabungkan dengan PSO sebagaimana digambarkan pada gambar 5. Dengan menggunakan *dataset* yang sama, hasil pengujian menunjukkan bahwa *accuracy* yang didapat sebesar 89.78% dengan hasil AUC = 0,902.

Tabel 5. Hasil Penentu pada SVM

Accuracy	AUC
89.78%	0,902



Gambar 5. Penerapan Model *Support Vector Machine* dengan *Particle Swarm Optimization*

a. Confusion Matrix

Berikut merupakan hasil dari penerapan model pada *RapidMiner*, menunjukkan hasil dari *confusion matrix* pada algoritma *Support Vector Machine* dengan *Particle Swarm Optimization*.

Tabel 6. Confusion Matrix *Support Vector Machine* dengan *Particle Swarm Optimization*

	true negatif	true positif	class precision
pred. negatif	110	13	89.43%
pred. positif	1	11	91.67%
class recall	99.10%	45.83%	

$$\begin{aligned}
 \text{Accuracy} &= \frac{11 + 110}{13 + 110 + 11 + 1} \\
 &= 89,78\%
 \end{aligned}$$

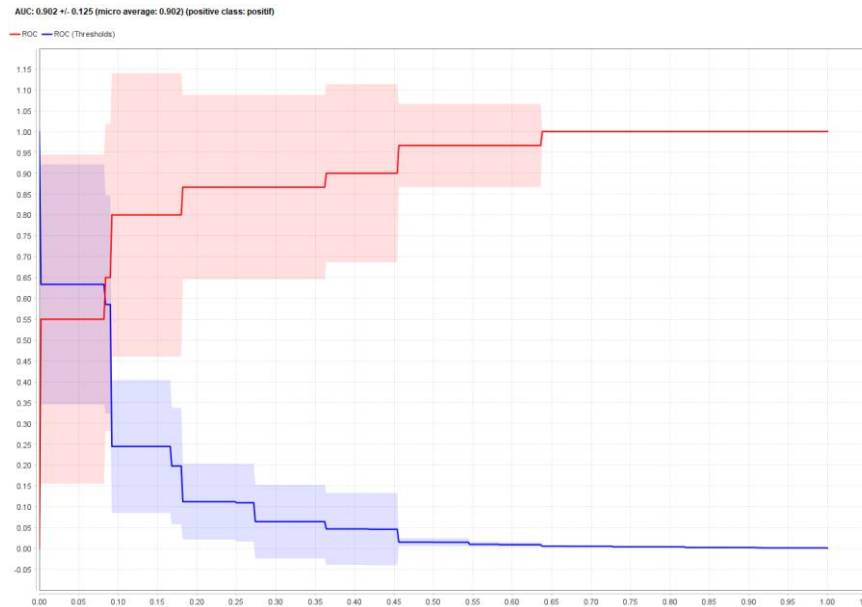
$$\text{Recall} = \frac{11}{11 + 110}$$

$$= 50,00\%$$

$$\text{Precision} = \frac{11}{11 + 1}$$

$$= 91,67\%$$

b. Kurva ROC



Gambar 6. Kurva ROC Support Vector Machine dengan Particle Swarm Optimization

Hasil dari penerapan model pada *RapidMiner*, penggabungan SVM dan PSO mendapatkan hasil *Accuracy* 89,78%, *Recall* 50,00%, *Precision* 91,67%, dan AUC 0,902.

3. Evaluasi dan Validasi Hasil SVM dan SVM + PSO

Sebelum dilakukan optimasi dengan PSO, SVM menghasilkan nilai *accuracy* 81.48% dan nilai AUC 0,825 yang berarti klasifikasi terbaik, kemudian hasil penerapan optimasi SVM menggunakan PSO memiliki nilai *accuracy* 89.78% dan AUC 0,902 yang berarti ada peningkatan nilai *accuracy* tersebut.

Tabel 7. Evaluasi dan Validasi Hasil Penelitian

Support Vector Machine			
accuracy 81.48%		AUC 0,825	
	true negatif	true positif	class precision
pred. negatif	109	23	82.58%
pred. positif	2	1	33.33%
class recall	98.20%	4.17%	
Support Vector Machine + Particle Swarm Optimization			
accuracy 89.78%		AUC 0,902	
	true negatif	true positif	class precision
pred. negatif	110	13	89.43%
pred. positif	1	11	91.67%
class recall	99.10%	45.83%	

PENUTUP

Berdasarkan hasil pembahasan, dapat disimpulkan bahwa SVM dapat dioptimalkan menggunakan PSO terutama pada dalam konteks permintaan informasi pada *platform online shop*. *Accuracy* dan AUC meningkat hal ini dapat dilihat dari sebelumnya nilai *accuracy* sebesar 81.48% dan nilai AUC 0,825. Setelah dilakukan penggabungan SVM dan PSO menghasilkan nilai *accuracy* sebesar 89.78% dan AUC 0,902.

DAFTAR PUSTAKA

- [1] P. H. L. S. H. W. P. Eri Yanti Nasution, "Perkembangan Transaksi Bisnis E-Commerce terhadap Pertumbuhan Ekonomi di Indonesia," *Jurnal Ekonomi & Ekonomi Syariah Vol 3*, pp. 506-519, 2020.
- [2] M. S. M. Niki Ratama, "Sosialisasi Penggunaan Ecommerce Dalam Perkembangan Bisnis Di Era Digital," *Abdi Jurnal Publikasi Vol 1*, pp. 6-12, 2022.
- [3] Alwendi, "Penerapan E-Commerce Dalam Meningkatkan Daya Saing Usaha," *Jurnal Manajemen Bisnis vol 17*, pp. 317-325, 2020.
- [4] D. A. RACHMAWATI, "Klasifikasi Teks Permintaan Informasi Untuk Aplikasi Online Shop Menggunakan Algoritma Support Vector Machine (Studi Kasus: Bento Shop)," Institut Teknologi Sepuluh Nopember Surabaya, Surabaya, 2016.
- [5] F. A. M. K. S. H. D. G. Muhammad Iqbal Ahmadi, "Sentiment Analysis Online Shop On The Play Store Using Method Support Vector Machine (SVM)," *Seminar Nasional Informatika 2020 (SEMNASIF 2020)*, pp. 196-203, 2020.
- [6] S. S. Rizqi Agung Permana, "Review Analisis Produk Marketplace Online Pada Algoritma Support Vector Machine," *Jurnal Ilmiah Informatika*, pp. 50-58, 2021.
- [7] M. K. Jakub Nalepa, "Selecting training sets for support vector machines:," *Artif Intell Rev*, p. 857-900, 2018.
- [8] R. N. Handayani, "Optimasi Algoritma Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Produk Tokopedia Menggunakan PSO," *Media Informatika Vol.20 No.2*, pp. 97-108, 2021.
- [9] A. Taufik, "Optimasi Particle Swarm Optimization Sebagai Seleksi Fitur Pada Analisis Sentimen Review Hotel Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes," *Jurnal Teknik Komputer*, pp. 40-47, 2017.
- [10] R. D. H. a. Eliyani, "Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi," *J. Edukasi dan Penelit. Inform vol. 7, no. 1*, pp. 58-63, 2021.
- [11] I. C. P. P. A. Arsyia Monica Pravina, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, pp. 2789-2797, 2019.