

Journal of Practical Computer Science

Vol.1 No.2 November 2021

ISSN Media Elektronik: 2809-8137

Diterima: 5 Desember 2021 | Direvisi: 10 Desember 2021 | Dipublikasikan: 30 Desember 2021

Optimalisasi Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Mendiagnosa Penyakit Kanker Payudara

Ari Maulana^{1,*}, Agung Nugroho², Ikhsan Romli³

^{1,2,3}Teknik Informatika, Universitas Pelita Bangsa <u>¹arimaulana221096@gmail.com</u>, <u>²agung@pelitabangsa.ac.id</u>, <u>³ikhsan.romli@pelitabangsa.ac.id</u>

ABSTRAK

Kanker payudara adalah kanker yang paling umum pada wanita dan penyebab utama kematian kanker di seluruh dunia. Klasifikasi dalam data mining merupakan dua bentuk proses analisis data yang digunakan untuk mengekstraksi model yang menggambarkan kelas data atau untuk memprediksi tren data di masa depan. Support Vector Machine (SVM) dikenal juga dengan support vector network yang merupakan metode supervised terkait dengan learning algorithm untuk analisa pola data yang digunakan untuk klasifikasi dan regresi. Seleksi fitur banyak digunakan untuk mengatasi fitur yang tidak relevan dan berlebihan. Seleksi fitur menyederhanakan sekumpulan data dengan mengurangi dimensi dan mengidentifikasi fitur yang relevan tanpa mengurangi akurasi prediksi. Penelitian ini mengguanakan algoritma Support Vector Machine dengan Particle Sarm Optimization untuk mendiagnosa penyakit kanker payudara. Hasil dari penelitian ini adalah accuracy sebesar 97.61%, precision sebesar 99.21% dan recall 96.94%. Penggunaan Particle Swarm Optimization bekerja secara efektif dalam meningkatkan nilai akurasi.

Kata kunci: Kanker payudara, klasifikasi, support vector machine, particle swarm optimizatiom.

ABSTRACT

Breast cancer is the most common cancer in women and the leading cause of cancer death worldwide. Classification in data mining are two forms of data analysis processes used to extract models that describe data classes or to predict future data trends. Support Vector Machine (SVM) is also known as support vector network which is a supervised method related to learning algorithms for analyzing data patterns used for classification and regression. Feature selection is widely used to address irrelevant and redundant features. Feature selection simplifies data sets by reducing dimensions and identifying relevant features without compromising prediction accuracy. This study uses the Support Vector Machine algorithm with Particle Swarm Optimization to diagnose breast cancer. The results of this study are accuracy of 97.61%, precision of 99.21% and recall of 96.94%. The use of Particle Swarm Optimization works effectively in increasing the accuracy value.

Keyword: Breast cancer, classification, support vector machine, particle swarm optimization.

PENDAHULUAN

Kanker payudara adalah kanker yang paling umum pada wanita dan penyebab utama kematian kanker di seluruh dunia. Meskipun etiologi kanker payudara tidak diketahui, faktor risiko berbagai kemungkinan mempengaruhi perkembangan penyakit ini termasuk faktor genetik, hormonal, lingkungan, sosiobiologis dan fisiologis. Selama beberapa dekade terakhir, risiko kanker payudara meningkat di negara-negara industri dan berkembang sebesar 1%

- 2% per tahun, tingkat kematian akibat kanker payudara menurun sedikit. Maka dari itu, penyakit kanker payudara perlu didiagnosis (Hermawanti 2012).

Klasifikasi dalam data mining merupakan dua bentuk proses analisis data yang digunakan untuk mengekstraksi model yang menggambarkan kelas data atau untuk memprediksi tren data di masa depan. Dalam proses klasifikasi terdapat 2 fase; fase pertama adalah training data dimana dalam fase ini data dipelajari dan dianalisis dengan algoritma klasifikasi. Model atau pengklasifikasi yang dipelajari disajikan dalam bentuk pola atau aturan klasifikasi; fase kedua adalah penggunaan model untuk klasifikasi, dan testing data digunakan untuk memperkirakan akurasi yang dihasilkan berdasarkan aturan klasifikasi (Kurniawan and Ivandari 2017).

Support Vector Machine (SVM) dikenal juga dengan support vector network yang merupakan metode supervised terkait dengan learning algorithm untuk analisa pola data yang digunakan untuk klasifikasi dan regresi (Nilawati and Achyani 2019).

Support Vector Machine (SVM) dikenal juga dengan support vector network yang merupakan metode supervised terkait dengan learning algorithm untuk analisa pola data yang digunakan untuk klasifikasi dan regresi. SVM merupakan usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah class pada input space (Nilawati and Achyani 2019).

SVM telah berhasil diaplikasikan dalam problema dunia nyata (*real-world problems*), dan secara umum memberikan solusi yang lebih baik dibandingkan metode konvensional seperti misalnya artificial neural network. SVM dapat mengatasi masalah klasifikasi dan regresi dengan linier ataupun nonlinier kernel yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi. SVM juga memiliki akurasi tinggi dan tingkat kesalahan yang relative kecil, kemampuan untuk mengatasi overfitting tidak membutuhkan data yang terlalu besar dan dapat digunakan untuk melakukan prediksi (Fatmawati and Windarto 2018).

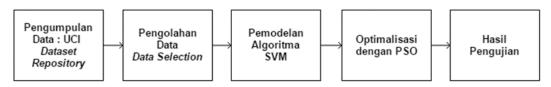
Particle Swarm Optimization merupakan salah satu algoritma optimasi yang dapat digunakan untuk pengambilan keputusan. Algoritma ini bekerja dengan cara menghitung secara terus menerus calon solusi dengan menggunakan suatu acuan kualitas. Algoritma ini mengoptimasi permasalahan dengan cara menggerakkan partikel didalam ruangan permasalahan menggunakan fungsi tertentu (Saputra and Prasetyo 2020).

Particle Swarm Optimization (PSO) merupakan salah satu optimasi metaheuristik untuk seleksi fitur karena telah terbukti kompetitif dibandingkan dengan algoritma genetika dalam beberapa kasus, terutama dalam bidang optimasi. Optimasi metaheuristik telah terbukti sebagai metodelogi yang unggul untuk mendaparkan solusi yang baik dalam waktu (Seal et al. 2015).

Particle Swarm Optimization merupakan salah satu algoritma optimasi yang dapat digunakan untuk pengambilan keputusan. Algoritma ini bekerja dengan cara menghitung secara terus menerus calon solusi dengan menggunakan suatu acuan kualitas. Algoritma ini mengoptimasi permasalahan dengan cara menggerakkan partikel didalam ruangan permasalahan menggunakan fungsi tertentu (Achyani, 2018)

METODE PENELITIAN

Dalam melakukan analisis dan klasifikasi untuk memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuatlah langkah-langkah dalam tahapan penelitian yang akan dilakukan sebagai berikut **Gambar 1.**



Gambar 1. Tahapan Penelitian

Pada penelitian ini kalsifikasi menggunakan algoritma SVM dan dioptimalisasi kinerjanya meggunakan PSO dalam pengolahan data dan yang akan dijadikan data sample dalam penelitian ini.

1. Pengumpulan data

Pengumpulan data dalam penelitian ini menggunakan metode pengumpulan data sekunder. Data sekunder adalah data yang diperoleh tidak secara langsung pada sumbernya. Sumber data yang digunakan penelitian ini berasal dari data publik yaitu *Breast Cancer Wusconsin (Diagnostic) Data Set* dari UCI Repository.

2. Pengolahan data

Berdasarkan pada data set diatas, untuk mengetahui informasi dari data yang didaptkan sebelumnya, diperlukan suatu cara atau teknik untuk mendapatkan pengetahuan tersebut dibutuhkan suatu usaha ekstraksi yang disebut data mining.

a. Data Cleaning

Data *Cleaning* (pembersihan data) merupakan penyaringan data atau pembersihan data yang tidak dibutuhkan karena adanya kesalahan data untuk menghasilkan data yang berkualitas. Pembersihan data dilakukan jika ada kesalahan berupa:

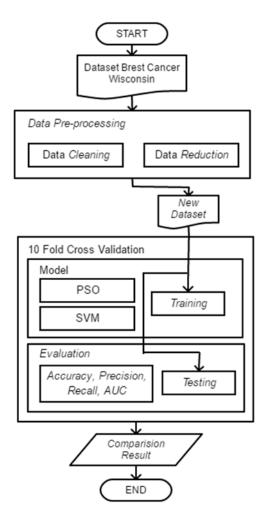
Missing data, yaitu data yang hilang, penanganan data yang hilang dapat dilakukan dengan mengabaikan tuple atau baris data yang memiliki nilai hilang, mengisi nilai secara manual, menggunakan mean/median, mengganti dengan nilai terbanyak atau dengan menerapkan nilai prediksi. Pada dataset Breast Cancer Wisconsin terdapat 1 atribut yang memiliki missing values.

b. Data Reduction

Data Reduction merupakan tahap data preprocessing dengan mengurangi jumlah data yang dihasilkan dari data asli, pengurangan tersebut dapat dilakukan dengan cara pengurangan atribut, namun tetap memilih atribut yang penting yang akan digunakan dalam teknik mining. Berkurangnya atribut tidak akan mempengauhi hasil representasi akhir karena akan menunjukan hasil yang sama.

3. Alur Pemodelan Penelitian

Model yang disulkan menggunakan algoritma SVM dengan menambahkan PSO. Berikut Gambar 2



Gambar 2. Alur Pemodelan Penelitian

Penjelasan:

- a. Start: yaitu proses memulai penelitian.
- b. Dataset Brest Cancer Wisconsin: yaitu dataset yang digunakan dalam penelitian.
- c. Data *Pre-processing*, yaitu proses yang dilakukan untuk mengolah data, dalam penelitian ini menggunakan 2 tahap yaitu:
 - Data Cleaning, yaitu proses melakukan penyaringan pada data. Data yang dicleaning dalam penelitian ini adalah ID Number karena tidak menjadi faktor dalam mendiagnosa penyakit kanker payudara.
 - 2) Data Reduction, yaitu proses pemilihan atribut yang akan digunakan setelah melakukan cleaning.
- d. New *Dataset*, yaitu data baru yang telah dilakukan pengolahan yang akan dijadikan sebagai data uji atau data tes dalam penelitian.
- e. 10 Fold Cross Validation, yaitu pengujian data uji dilakukan sebanyak 10 kali untuk mendapatkan hasil yang terbaik. Dalam penelitian menggunakan model SVM dan PSO. Pada tahap ini terdapat 2 model, yaitu:
 - 1) SVM dipilih karena memiliki akurasi tinggi dan tingkat kesalahan yang relative kecil.
 - 2) PSO dipilih sebagai optimalisasi, seleksi fitur banyak digunakan untuk mengatasi fitur yang tidak relevan dan berlebihan.

f. Evaluation, yaitu proses penelitian yang dilakukan dengan data testing menghasilkan nilai accuracy, precision, recall dan AUC.

ISSN Online: 2809-8137

- g. Comparisiom Result, yaitu perbandingan nilai akurasi dengan model SVM dan model SVM dan PSO.
- h. End, yaitu penelitian selesai.

HASIL DAN PEMBAHASAN

Analisa Data

Pada tahap ini peneliti menentukan teknik data mining yang digunakan untuk mengolah data yang sudah disiapkan sebelumnya. Dan yang sudah disiapkan untuk klasifikasi dibagi menjadi dua data, untuk data *training* (70%) dan untuk data *testing* (30%). Pembagian data training dan data testing menggunakan teknik sampling random sistematik.

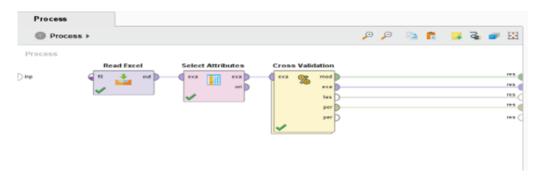
Cara penggunaan teknik ini yaitu dengan cara perandoman atau pengundian hanya dilakukan satu kali, yakni ketika menentukan unsur pertama dari sampling yang akan diambil. Penentuan unsur sampling selanjutnya ditempuh dengan cara memanfaatkan interval sampel. Dari hasil diperoleh data *testin* sebanyak 210 maka sisanya dijadikan data *training* sebanyak 699-210 = 489

Pengujian

Pada penelitian ini pengujian dilakukan sebanyak 2 kali untuk perbandingan. Berikut adalah hasil dari pengujian yang dilakukan.

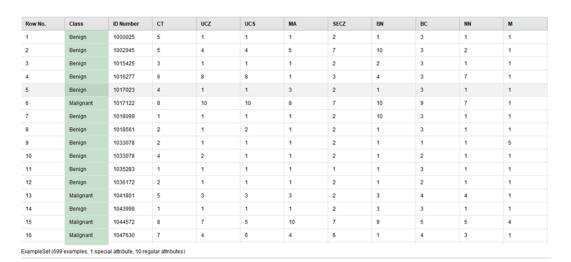
Pengujian Pertama

Pengujian pertama dilakukan dengan menggunakan select attributes pada RapidMiner Studio 9.8 untuk melakukan proses *pre-processing*. Pemodelan dbuat menggunakan algoritma klasifikasi yaitu *Support Vector Machine*. Berikut **Gambar 3**:



Gambar 3. Model Proses 1

Pada model pengujian dibuat dengan fitur *read* excel yang digunakan untuk import dataset kanker payudara. Berikut adalah hasil import *dataset* dengan fitur read excel pada RapidMiner.



Gambar 4. Import Dataset

Gambar 4 menunjukan hasil import dataset dengan 10 attribut dan 1 attribut spesial sebagai label, dan 10 atribut lainnya sebagai *class*.

Proses selanjutnya adalah proses seleksi atribut yang digunakan untuk pengujian. Penggunaan select attributes untuk memilih atribut yang digunakan.

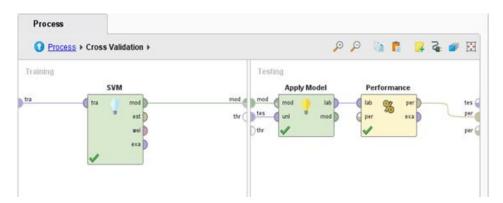
Row No.	Class	СТ	UCZ	UCS	MA	SECZ	BN	BC	NN	м
1	Benign	5	1	1	1	2	1	3	1	1
2	Benign	5	4	4	5	7	10	3	2	1
3	Benign	3	1	1	1	2	2	3	1	1
4	Benign	6	8	8	1	3	4	3	7	1
5	Benign	4	1	1	3	2	1	3	1	1
6	Malignant	8	10	10	8	7	10	9	7	1
7	Benign	1	1	1	1	2	10	3	1	1
8	Benign	2	1	2	1	2	1	3	1	1
9	Benign	2	1	1	1	2	1	1	1	5
10	Benign	4	2	1	1	2	1	2	1	1
11	Benign	1	1	1	1	1	1	3	1	1
12	Benign	2	1	1	1	2	1	2	1	1
13	Malignant	5	3	3	3	2	3	4	4	1
14	Benign	1	1	1	1	2	3	3	1	1
15	Malignant	8	7	5	10	7	9	5	5	4
16	Malignant	7	4	6	4	6	1	4	3	1

ExampleSet (699 examples, 1 special attribute, 9 regular attributes)

Gambar 5. Hasil Seleksi Atribut

Gambar 5 menujukkan dengan menggunakan fitur *select attributes* mengurangi jumlah atribut pada awal 11 menjadi 10 atribut, yang terdiri dari 9 atribut class dan 1 atribut label.

Pada penelitian dilakukan proses validasi menggunakan *cross validation* dengan nilai k= 10 fold dengan berbagai macam jenis kernel. Berikut **Gambar 6** pemodelan yang terdapat pada *cross validation*



Gambar 6. Model Cross Validation dengan SVM

Pada pemodekan Cross Validatuon didalamnya terdapat dua bagian, yaitu bagian taining dan testing. Bagian training digunakan untuk algortima klasifikasi Support Vector Machine (SVM) dan bagian testing menggunakan Apply Model untuk menampilkan confusion table, yang digunakan untuk menampilkan hasil dari accuracy, precision, recall, dan AUC. Berikut Gambar 7 yang menujukakn kernel model Support Vector Machine

```
Kernel Model

Total number of Support Vectors: 210
Bias (offset): 0.023

w[CT] = 0.523
w[UCZ] = 0.194
w[UCS] = 0.302
w[MA] = 0.330
w[SECZ] = 0.281
w[BN] = 0.437
w[BC] = 0.341
w[NN] = 0.152
w[M] = 0.270
```

Gambar 7. Kernel Model Support Vector Machine

Hasil confusion table dari pengujian pertama menggunakan algoritma Support Vector Machine menghasilkan nilai accuracy, precision, recall dan AUC. Berikut hasil dari confusion table.

Hasil accuracy menggunakan Algoritma Supoort Vector Machine dalam pengujian pertama dengan Cross Validation yaitu sebesar 95.71%. Dapat dilihat pada **Gambar 8** dibawah ini

accuracy, 35.7 fm 7-4.17m (micro average, 35.7 fm)				
	true Benign	true Malignant	class precision	
pred. Benign	126	4	96.92%	
pred. Malignant	5	75	93.75%	
class recall	96.18%	94.94%		

Gambar 8. Hasil Accuracy Support Vector Machine

accuracy: 05 71% +/ A 17% (micro avorago: 05 71%)

Hasil *precision* menggunakan Algoritma Supoort Vector Machine dalam pengujian pertama dengan Cross Validation yaitu sebesar 96.92%. Dapat dilihat pada Gambar 9

precision: 94.39% +/- 7.61% (micro average: 93.75%) (positive class: Malignant)

	true Benign	true Malignant	class precision
pred. Benign	126	4	96.92%
pred. Malignant	5	75	93.75%
class recall	96.18%	94.94%	

Gambar 9. Hasil Precision SVM

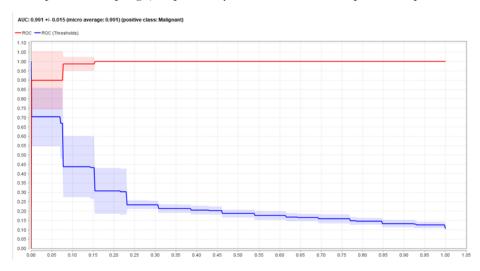
Hasil *recall* menggunakan Algoritma *Supoort Vector Machine* dalam pengujian pertama dengan *Cross Validation* yaitu sebesar 96.18%. Dapat dilihat pada **Gambar 10.**

recall: 95.00% +/- 6.45% (micro average: 94.94%) (positive class: Malignant)

	true Benign	true Malignant	class precision
pred. Benign	126	4	96.92%
pred. Malignant	5	75	93.75%
class recall	96.18%	94.94%	

Gambar 10 Hasil recall SVM

AUC yang dihasilkan pada model pengujian pertama yaitu sebesar 0.991. Dapat dilihat pada Gambar 11



Gambar 11. AUC SVM

Pengujian kedua.

Pengujian kedua dilakukan dengan menggunakan select attributes pada RapidMiner Studio 9.8 untuk melakukan proses pre-processing. Pemodelan dbuat menggunakan penambahan *model Particle Support Optimization* (PSO) dan *Support Vector Machine* (SVM) Berikut **Gambar 12**

Gambar 12. Model Proses 1 PSO dan SVM

Pada **Gambar 12** pemodelan dibuat menggunakan tambahan PSO dan menggunakan klasifikasi yang sama yaitu SVM.

Pada pemodelan Cross Validatuon didalamnya terdapat dua bagian, yaitu bagian taining dan testing. Bagian training digunakan untuk algortima klasifikasi Support Vector Machine (SVM) dan bagian testing menggunakan Apply Model untuk menampilkan confusion table, yang digunakan untuk menampilkan hasil dari accuracy, precision, recall, dan AUC.

Hasil confusion table dari pengujian kedua menggunakan algoritma Support Vector Machine dengan Particle Support Optimization menghasilkan nilai accuracy, precision, recall dan AUC. Berikut hasil dari confusion table.

Hasil accuracy menggunakan Algoritma Supoort Vector Machine dengan Particle Swarm Optimization dalam pengujian pertama dengan Cross Validation yaitu sebesar 97.61%. Dapat dilihat pada **Gambar 13**

 accuracy: 97.62% +/- 4.05% (micro average: 97.62%)

 true Benign
 true Malignant
 class precision

 pred. Benign
 127
 1
 99.22%

 pred. Malignant
 4
 78
 95.12%

Gambar 13. Hasil Accuracy SVM dengan PSO

98.73%

96.95%

Hasil precision menggunakan Algoritma Supoort Vector Machine dalam pengujian pertama dengan Cross Validation yaitu sebesar 99.21%. Dapat dilihat pada **Gambar 14.**

precision: 95.64% +/- 1.37% (micro average: 95.12%) (postuve class: mailgnant)				
	true Benign	true Malignant	class precision	
pred. Benign	127	1	99.22%	
pred. Malignant	4	78	95.12%	
class recall	96.95%	98.73%		

Gambar 14. Hasil Precisiom SVM dengan PSO

Hasil *recall* menggunakan Algoritma *Supoort Vector Machine* dalam pengujian pertama dengan *Cross Validation* yaitu sebesar 96,94%.Dapat dilihat pada **Gambar 15**

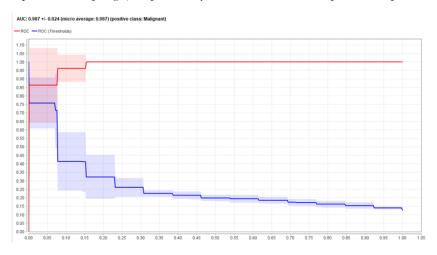
recall: 98.75% +/- 3.95% (micro average: 98.75%) (positive class: mailgnant)				
	true Benign	true Malignant	class precision	
pred. Benign	127	1	99.22%	
pred. Malignant	4	78	95.12%	
class recall	96.95%	98.73%		

Gambar 15. Hasil Recall SVM dengan PSO

class recall

ISSN Online: 2809-8137

AUC yang dihasilkan pada model pengujian pertama yaitu sebesar 0.987. Dapat dilihat pada Gambar 16



Gambar 16. AUC SVM dengan PSO

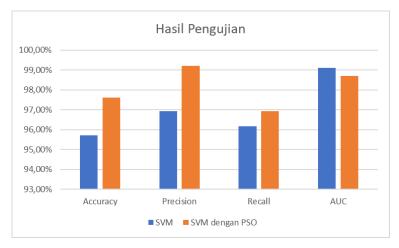
Pembahasan

Penelitian ini dilaukan sebanyak dua kali, yaitu pengujian pertama menggunakan algoritma klasifikasi Support Vector Machine dan pengujian kedua menggunakan algoritma klasifikasi Support Vector Machine dengan Particle Swarm Optimization menghasilkan accuracy, precision, recall dan AUC.

Tabel 1 Hasil dari 2 penelitian yang dilakukan

	Pengujian Pertama	Pengujian Kedua
	SVM	SVM dengan PSO
Accuracy	95,71%	97,61%
Precision	96,92%	99,21%
Recall	96,18%	96,94%
AUC	0,991	0,987

Berdasarkan hasil yang diperoleh menunjukkan bahwa adanya peningkatan nilai accuracy, precision dan recall. Pada algoritma Support Vector Machine dengan Particle Swarm Optimization meningkat jika dibandingkan hanya dengan menggunakan satu teknik pembelajaran. Hasil kenaikan nilai adalah sebesar 1,9% pada accuracy, 2,29% pada precision dan 0,76% pada recall



Gambar 17. Hasil Pengujian

KESIMPULAN

Berdasarkan pada hasil penelitian yang telah dilakukan maka dapat disimpulkan bahwa penerapan *Particle Swarm Optimization* dapat meningkatkan akurasi pada klasifikasi *Support Vector Machine* dalam mendiagnosa penyakit kanker payudara. Hasil nilai adalah *accuracy* sebesar 97.28% (naik 0.57%), *precision* sebesar 95.55% (naik 0,93%) dan *recall* 97.06% (naik 0,78%). Penggunaan Particle Swarm Optimization bekerja secara efektivitas dalam meningkatkan nilai akurasi.

DAFTAR PUSTAKA

- A. Seal, S. Ganguly, D. Bhattacharjee, M. Nasipuri, and C. Gonzalo-Martin, "Feature selection using particle swarm optimization for thermal face recognition," *Adv. Intell. Syst. Comput.*, vol. 304 AISC, no. VOLUME 1, pp. 25–35, 2015, doi: 10.1007/978-81-322-1985-9_2.
- F. Kurniawan and Ivandari, "Komparasi Algoritma Data Mining Untuk Klasifikasi Penyakit Kanker Payudara," *J. Stmik*, vol. XII, no. 1, pp. 1–8, 2017.
- K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
- L. Hermawanti, "Penerapan algoritma klasifikasi c4.5 untuk diagnosis penyakit kanker payudara," *J. Tek. Unisfat*, vol. 7, no. 1, pp. 57–64, 2012.
- L. Nilawati and Y. E. Achyani, "Optimasi Metode Particle Swarm Optimization (PSO) Pada Prediksi Penilaian Apartemen," *Paradig. J. Komput. dan Inform.*, vol. 21, no. 2, pp. 227–234, 2019, doi: 10.31294/p.v21i2.6159.
- M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.
- M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.
- R. H. Saputra and B. Prasetyo, "Improve the Accuracy of C4 . 5 Algorithm Using Particle Swarm Optimization (PSO) Feature Selection and Bagging Technique in Breast Cancer Diagnosis," pp. 47–55, 2020.
- Y. E. Achyani, "Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi Pemasaran Langsung," *J. Inform.*, vol. 5, no. 1, pp. 1–11, 2018, doi: 10.31311/ji.v5i1.2736.