

OPTIMASI SUPPORT VECTOR MACHINE MENGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK KLASIFIKASI MULTICLASS DATASET

Rissa Nurfitriana Handayani¹, Djajasukma Tjahjadi²

¹Universitas Adhirajasa Reswara Sanjaya
e-mail: rissa@ars.ac.id

²STMIK Likmi
e-mail: djaja@likmi.ac.id

Abstrak

Data mining merupakan proses pengembangan model yang bertujuan untuk memahami pola dari hasil analisis data dan prediksi. Kumpulan data atau yang biasa disebut *dataset* terdiri dari atribut, data dan kelas. Salah satu permasalahan yang terdapat pada dataset yaitu *multiclass dataset*. *Multiclass* merupakan *dataset* yang memiliki kelas klasifikasi *polynomial*. Algoritma *Support Vector Machine (SVM)* banyak digunakan oleh peneliti untuk metode klasifikasi. *SVM* dipilih karena mampu menentukan *hyperplane* terpisah untuk memaksimalkan *margin* antara 2 kelas yang berbeda. Namun, *SVM* terdapat kekurangan pada saat dilakukan pemilihan parameter, untuk itu diterapkan *Particle Swarm Optimization (PSO)* untuk meningkatkan hasil akurasi. *PSO* digunakan untuk optimasi bobot sedangkan *SVM* digunakan untuk klasifikasi. *Dataset* yang digunakan merupakan *dataset* dari *UCI Machine Learning Repository*. Penerapan *SVM* menghasilkan nilai akurasi sebesar 78.30% pada *breast tissue dataset*, 85.81% pada *vertebral column dataset*, 50.50% pada *obesity levels dataset*, 92.58% pada *hepatitis C virus (HCV) dataset*, dan sebesar 52.63% pada *exasens dataset*. Kemudian setelah dilakukan penerapan optimasi *PSO* terhadap algoritma *SVM* diperoleh hasil 83.96% pada *breast tissue dataset*, 87.42% pada *vertebral column dataset*, 59.64% pada *obesity levels dataset*, 96.59% pada *hepatitis C virus (HCV) dataset*, dan sebesar 56.40% pada *exasens dataset*. Setelah diterapkan *PSO* terhadap *SVM* diperoleh peningkatan akurasi sebesar 1% - 9%. Hasil percobaan menghasilkan bahwa metode yang digunakan mampu menghasilkan kinerja yang baik, dibandingkan dengan hasil klasifikasi lain dengan peningkatan yang signifikan dengan nilai p uji-t 0.019.

Kata Kunci: *Dataset Medis, Klasifikasi, Multiclass Dataset, Particle Swarm Optimization, Support Vector Machine*

Abstract

Data mining is a model development process aimed at understanding the patterns of data analysis and predictions. Collections of data or otherwise known as *datassets* consist of attributes, data and class. One of the many problems with *datassets* is *multiclass datassets*. *Multiclass* is a *datasset* who has *polypar classification classes*. The algorithm *support vector machine (SVM)* is used extensively by researchers for classification methods. *SVM* was selected because it was able to specify a separate *hyperplane* to maximize the *margin* between two different classes. Namum, *SVM* there is a shortage at the time of perimeter selection, for it applied *particle swarm optimization (ps)* to increase accuracy. *Pso* is used for *weightlessness optimization*, whereas *SVM* is used for classification. The *datasset* used is a *datasset* of the *uci machine learning revalued*. *SVM* application results in a 78 percent accuracy rate, 8.81% in *natural asset planets*, 50.50% in the *obesity of datasset column*, 92.58% in *datasset hepatitis c (HCV), datasset*, and by 52.63% in the *datasset exasens.63%*. Then, after a *pso optional application* of *SVM* algorithm results in 83.96% on the *breast tissue datasset*, 87.42% on *mandate of datasset column*, 59.59% on *datasset hepatitis c (HCV)*, and by 56.40% on the *datasset exasens*. After *pso* applied to *SVM* obtained a 1% - 9% increase in accuracy. The results of the experiment produced that the methods used were able to produce good performance, compared with other classifications with significant increases by a p uji-t 0019 value.

Keywords: *Medical Dataset, Classification, Multiclass Dataset, Particle Swarm Optimization, Support Vector Machine*

1. Pendahuluan

Data mining merupakan representasi proses ekstraksi informasi yang berguna, tidak diketahui sebelumnya dan tersembunyi dari data (Jamsa, 2021), dan mengembangkan model yang bertujuan untuk memahami fenomena dari hasil analisis data dan prediksi (Kantardzic, 2020). *Data mining* merupakan proses ekstraksi pengetahuan dari jumlah data yang besar yang tersimpan dalam komputer, proses pada *data mining* lebih banyak didasarkan pada teknologi informasi (Prasetio & Ripandi, 2019).

Kumpulan ekstraksi pengetahuan yang digunakan untuk menganalisis *data mining* biasa disebut *Dataset*. *Dataset* terdiri dari kumpulan objek dan sifat atau karakteristik dari suatu objek itu sendiri atau biasa disebut atribut (Listiana & Muslim, 2017). Kumpulan data atau yang biasa disebut *dataset* terdiri dari atribut, data dan kelas. Salah satu permasalahan yang terdapat pada *dataset* yaitu *multi class* (Kantardzic, 2020).

Dataset dengan karakteristik *multiclass* merupakan dataset yang memiliki kelas klasifikasi lebih dari 2 atau biasa disebut *polynomial* (Sucipto et al., 2016). *Breast tissue dataset* (Marques & Jossinet, 2010) merupakan contoh dari *multi class dataset* hal ini ditunjukkan dengan terdapat 6 kelas yaitu: *carciroma, fibro-adenoma, mastopathy, glandular, connective* dan *adipose*. *Vertebral column dataset* (Barreto et al., 2011), juga merupakan *multiclass dataset*, karena memiliki 3 kelas yaitu: *hernia, spondylolisthesis*, dan *normal*. *Estimation of Obesity Levels based on Eating Habits and Physical Condition Data Set* (Palechor & Manotas, 2019) merupakan bagian dari *multiclass dataset* karena memiliki 7 kelas, terdiri dari: *insufficient weight, normal weight, over weight level 1, over weight level 2, obesity type 1, obesity type 2, dan obesity type 3*. *Hepatitis C Virus (HCV) dataset* (Lichtinghagen et al., 2020) memiliki 5 kelas yaitu: *0 - blood donor, 0s - suspect blood donor, 1 - hepatitis, 2 - fibrosis, 3 - cirrhosis*. Dan kemudian pada *exasens dataset* (Zarrin & Roekendorf, 2020) memiliki 3 kelas yaitu: *non-smoker, ex-smoker, dan active smoker*.

Teknik klasifikasi pada *dataset* dengan karakteristik *multi-class* yang dapat

diterapkan adalah *Support Vector Machine, Naïve Bayes*, dan *k-Nearest Neighbor*. (Azis et al., 2017; Maurya et al., 2014). Pada penelitian ini, digunakan algoritma *Support Vector Machine* karena dapat memberikan hasil klasifikasi yang baik (Liu et al., 2018), hal ini disebabkan karena SVM dapat menentukan *hyperplane* atau bidang pemisah dengan memilih bidang dengan optimal *margin* maka generalisasi pada SVM dapat terjaga dengan sendirinya, tingkat generalisasi pada SVM tidak dipengaruhi oleh jumlah data latih, dengan menentukan parameter *soft margin, noise* dapat dikontrol sehingga makin besar parameter *soft margin*, makin besar pula pinalti yang dikenakan pada kesalahan pada klasifikasi sehingga proses pelatihan semakin ketat (Nugroho, 2020).

Support Vector Machine (SVM) merupakan metode *supervised learning* untuk menganalisis data dan membuat poa yang dapat digunakan untuk klasifikasi (Jayadeva et al., 2017). Metode SVM mampu menentukan *hyperplane* terpisah sehingga dapat memaksimalkan *margin* antara dua kelas yang berbeda (Murty & Raghava, 2016). Akan tetapi, (Basari et al., 2013) menyatakan bahwasannya SVM memiliki kekurangan dalam pemilihan parameter. Untuk meningkatkan hasil akurasi, pada penelitian ini diterapkan metode optimasi *Particle Swarm Optimization (PSO)*.

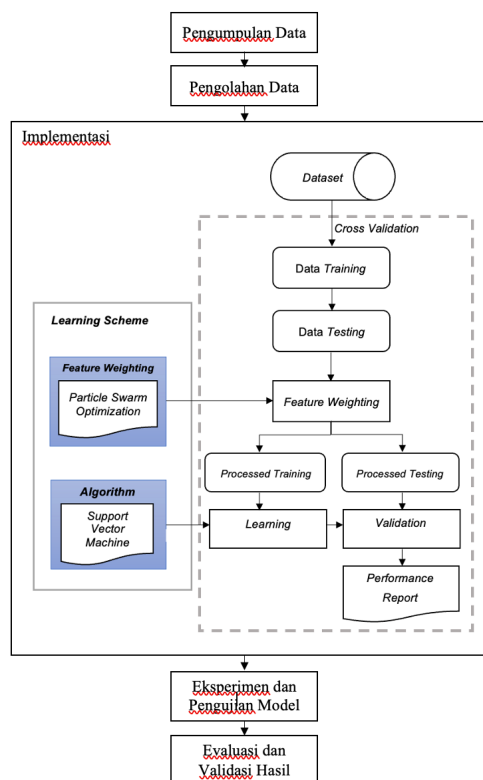
Particle Swarm Optimization (PSO) merupakan algoritma *searching*, berbasis populasi dan dapat diinisiasi dengan partikel (Couceiro & Ghamisi, 2016). Ketika diterapkan *PSO* terdapat beberapa metode untuk melakukan optimasi diantaranya meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau *variabel* yang dipakai, menseleksi atribut (*atribut selection*), optimasi bobot (*feature weighting*) dan *feature selection* (Basari et al., 2013). Penerapan *PSO* merupakan penentu parameter yang akan mempengaruhi nilai *accuracy* pada klasifikasi (Indrayuni, 2016).

Selain itu penelitian terdahulu yang dilakukan oleh (Prasetio, 2020) melakukan penelitian klasifikasi terhadap 5 (lima) *dataset* sebagai tolak ukur, namun menggunakan algoritma *k-Nearest Neighbor* dengan optimasi algoritma genetika. Dengan

demikian, penulis terdorong untuk melakukan penelitian serupa menerapkan klasifikasi terhadap 5 *dataset* medis dengan algoritma *support vector machine (SVM)* dengan metode optimasi *particle swarm optimization (PSO)*.

2. Metode Penelitian

Berdasarkan penjelasan yang dipaparkan pada pendahuluan, penelitian ini menggunakan metode penelitian eksperimen, dengan *dataset* yang digunakan adalah *dataset* yang bersifat publik yang dapat diakses melalui UCI Repository (<https://archive.ics.uci.edu/ml/index.php>), digunakan 5 *dataset* medis yang memiliki sifat *multiclass dataset*. Berikut merupakan tahapan penelitian yang akan diimplementasikan digambarkan melalui Gambar 1.



Gambar 1. Metode Penelitian

Adapun beberapa tahapan dalam penelitian ini yaitu sebagai berikut:

1. Pengumpulan data
Pada tahap ini dilakukan pengumpulan *dataset* yang akan diujicoba, terdapat 5 (lima) *dataset* medis dengan karakteristik *multiclass*. Berikut merupakan *dataset* yang digunakan:

a. Breast Tissue Dataset

Breast tissue dataset merupakan *dataset* jaringan payudara dipublikasikan pada tahun 2010. *Dataset* ini terdiri dari 106 data, terdiri dari 9 (sembilan) *atribut* yang merupakan karakteristik dari *breast tissue dataset*, yaitu: I0 (*impedivity (ohm) at zero frequency*), PA500 (*phase angle at 500 KHz*), HFS (*high-frequency slope of phase angle*), DA (*jarak impedansi antara ujung spektral*), AREA (*area under spectrum*), A/DA (*area normalized by DA*), MAX IP (*maximum of the spectrum*), DR (*jarak antara I0 dan titik maksimum frekuensi*), P (*panjang spectral curve*), dan satu label *breast tissue* dengan 6 label (Marques & Jossinet, 2010).

Dari 106 data yang terdapat pada *breast tissue dataset*, terdiri dari 21 data dengan label Car (*Carcinoma*), 15 data dengan label Fad (*Fibro-adenoma*), 18 data dengan label Mas (*Mastopathy*), 16 data dengan label Gla (*Glandular*), 14 data dengan label Con (*Connective*), dan 22 data dengan label Adi (*Adipose*).

b. Vertebral Column Dataset

Vertebral column dataset merupakan *dataset* mengenai *Orthopaedics* yang dipublikasikan pada tahun 2011. *Dataset* ini terdiri dari 310 data pasien, terdiri dari 6 (enam) *atribut* yang merupakan karakteristik dari *vertebral column dataset*, yaitu: *pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, degree spondylolisthesis*, dan satu label *vertebral column* dengan 3 label (Barreto et al., 2011).

Dari 310 data pasien yang terdapat pada *vertebral column dataset*, terdiri dari 60 data pasien dengan label *Hernia*, 150 data pasien dengan label *Spondylolisthesis*, dan 100 data pasien dengan label *Normal*.

c. Obesity Levels

Estimation of obesity levels based on eating habits and physical condition dataset merupakan *dataset* mengenai tingkat obesitas berdasarkan pada kebiasaan makan dan kondisi fisik yang dipublikasikan pada tahun 2019. *Dataset* ini terdiri dari 2.111 data, terdiri dari 16 (enam belas) *atribut* yang merupakan karakteristik dari *obesity dataset*, yaitu: *Gender, Age, Height, Weight, Family History with Overweight, FAVC, FCVC*,

NCP, CAEC, Smoke, CH20, SCC, FAF, TUE, CALC, Mtrans dan satu label *obesity dataset* dengan 7 kelas (Palechor & Manotas, 2019).

Dari 2.111 data pasien yang terdapat pada *obesity dataset*, terdiri dari 272 data pasien dengan label *Insufficient Weight*, 287 data pasien dengan label *Normal Weight*, 351 data pasien dengan label *Obesity Type I*, 297 data pasien dengan label *Obesity Type II*, 324 data pasien dengan label *Obesity Type III*, 290 data pasien dengan label *Overweight Level I*, dan 290 data pasien dengan label *Overweight Level II*.

d. *Hepatitis C Virus (HCV)*

Hepatitis C Virus (HCV) dataset merupakan dataset mengenai hepatitis C yang dipublikasikan pada tahun 2020. *Dataset* ini terdiri dari 615 data pasien, terdiri dari 12 (dua belas) atribut yang merupakan karakteristik dari HCV *dataset*, yaitu: *Age, Sex, ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, PROT*, dan satu label diagnosis dengan 5 label (Lichtinghagen et al., 2020).

Dari 615 data pasien yang terdapat pada HCV *dataset*, terdiri dari 533 data dengan label 0 = *Blood Donor*, 7 data pasien dengan label 0s = *Suspect Blood Donor*, 24 data pasien dengan label 1 = *Hepatitis*, 21 data pasien dengan label 2 = *Fibrosis*, 30 data pasien dengan label 3 = *Cirrhosis*.

e. *Exasens Dataset*

Exasens dataset merupakan data sampel air liur dari pengidap *asthma, COPD, infeksi, dan HC*, yang dipublikasikan pada tahun 2020. *Dataset* ini terdiri dari 399 data pasien, terdiri dari 8 (delapan) atribut yang merupakan karakteristik dari *exasens dataset*, yaitu: *ID, Imaginary Part (min Average), Imaginary Part (max Average), Real Part (min. Average), Real Part (max. Average), Gender, Age, Smoking* dan satu label diagnosis dengan 4 label (Zarrin & Roeckendorf, 2020).

Dari 399 data pasien yang terdapat pada HCV *dataset*, terdiri dari 80 data pasien dengan label *Asthma*, 79 data pasien dengan label *COPD*, 160 data pasien dengan label *HC*, dan 80 data pasien dengan label *Infected*.

2. Pengolahan data awal

Pada penelitian ini penulis menggunakan *medical dataset* yang terdiri dari 5 *dataset* yang digunakan sebagai *benchmark*. Untuk pengembangan dan pengujian model yang digunakan, data akan dibagi menjadi *data training* dan *data testing* menggunakan *Cross Validation*. *Data training* digunakan untuk pengembangan model dan *data testing* untuk pengujian model. *Dataset* yang sudah dibagi menjadi *data training* dan *data testing* dapat langsung diolah dengan menggunakan model yang sudah ditentukan akan menghasilkan nilai yang dapat menjadi indikator seberapa baik model yang digunakan implementasi.

3. Eksperimen dan pengujian model

Pada tahap ini eksperimen dan pengujian model dilakukan dengan tahapan sebagai berikut:

- a. Menyiapkan *dataset* untuk kebutuhan eksperimen.
- b. Membagi *dataset* menjadi *data training* dan *data testing*.
- c. Melakukan pembobotan fitur atau atribut menggunakan *particle swarm optimization*.
- d. Melakukan *training* terhadap *dataset* menggunakan model *support vector machine* dengan optimasi *particle swarm optimization* untuk pengembangan model.
- e. Melakukan *testing* terhadap *dataset* menggunakan model *support vector machine* dengan optimasi *particle swarm optimization* untuk pengujian model.

4. Evaluasi dan validasi model

Pada tahap ini dilakukan evaluasi terhadap eksperimen dan pengujian pada model yang diusulkan sehingga dapat diketahui hasil kinerja model yang digunakan dalam penelitian ini.

3. Hasil dan Pembahasan

Perbandingan hasil akurasi pada penerapan *Support Vector Machine* sebelum dan sesudah optimasi untuk mengevaluasi hasil keseluruhan dari penelitian ini. Penelitian ini menggunakan data yang sama yaitu, 5 *medical dataset* sebagai *benchmarked*. Perbandingan hasil penelitian ini dapat dilihat pada tabel 1.

Tabel 1. Perbandingan Hasil Penelitian

Datasets	SVM	Metode yang diusulkan
<i>Breast Tissue Dataset</i>	78.30%	83.96%
<i>Vertebral Column Dataset</i>	85.81%	87.42%
<i>Obesity Levels Dataset</i>	50.50%	59.64%
<i>Hepatitis C Virus (HCV) Dataset</i>	92.85%	96.59%
<i>Exasens Dataset</i>	52.63%	56.39%

Hasil eksperimen yang dituangkan pada Tabel 1 menyatakan bahwa metode yang diusulkan dapat meningkatkan akurasi dari kelima *dataset benchmark* dengan peningkatan sebesar 1% - 9% dibandingkan dengan algoritma *Support Vector Machine* tanpa optimasi. Peningkatan kinerja tertinggi diperoleh oleh klasifikasi *dataset obesity levels* dengan peningkatan 9.14% dengan hasil setelah dioptimasi 59.64%. Sedangkan peningkatan kinerja terendah diperoleh dari klasifikasi *dataset vertebral column* sebesar 1.61% meningkat menjadi 87.42%. Peningkatan performansi pada *dataset breast tissue* sebesar 6,18% menjadi 84.36%. *Dataset hepatitis C virus (HCV)* meningkat sebesar 3,74% menjadi 96.59%, dan *dataset exasens* meningkat sebesar 3.75% menjadi 56.40%.

Tabel 2. Hasil Eksperimen dari Optimasi Bobot

Nama Dataset	SVM + Particle Swarm Optimization	SVM + Forward Selection	
<i>Breast Tissue Data Set</i>	84.36%	74.64 %	75.73%
<i>Vertebral Column Data Set</i>	87.42%	87.10 %	86.77%
<i>Obesity Levels Datas</i>	59.64%	61.39 %	57.65%
<i>Hepatitis C</i>	96.59%	93.50 %	93.66%

Nama Dataset	SVM + Particle Swarm Optimization	SVM + Forward Selection	
<i>Virus (HCV) Data Set</i>			
<i>Exasens Data Set</i>	56.40%	56.65 %	55.64%

Berdasarkan hasil percobaan yang digambarkan melalui Tabel 2 menyatakan bahwa metode yang diusulkan yaitu optimasi bobot pada *Support Vector Machine* menggunakan *Particle Swarm Optimization* terhadap lima *dataset* medis yang digunakan, memiliki hasil akurasi yang lebih unggul dibandingkan dengan metode optimasi lain seperti *Forward Selection* dan *Backward Elimination*. Hasil dari penerapan *forward selection* dan *backward elimination* mengalami penurunan sebesar 0.035% - 2.45% namun pada penerapan optimasi tersebut juga mengalami peningkatan akurasi sebesar 0.007% - 7.15% dibandingkan dengan penerapan *Particle Swarm Optimization*, dan hasil terendah diperoleh dengan penerapan *forward selection*.

Berdasarkan hasil percobaan pada penelitian ini, untuk mengetahui apakah metode optimasi yang diusulkan dapat meningkatkan kinerja dalam klasifikasi *dataset* medis secara signifikan. Pengujian menggunakan uji signifikansi dilakukan, *t-Test Paired Two Sample for Means* digunakan dalam hasil antara sebelum dan sesudah menggunakan metode optimasi yang diusulkan.

Hasil pengujian uji-t menunjukkan bahwa metode yang diusulkan dapat meningkatkan kinerja *support vector machine* dalam hal akurasi secara signifikan pada semua *dataset* yang ditandai dengan p uji-t < 0.05 . Hasil Uji-t dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengujian t-Test

	SVM + PSO	
<i>Mean</i>	0.71998	0.76882
<i>Variance</i>	0.0375075	0.0318043
<i>Observations</i>	9	5
<i>Pearson Correlation</i>	5	5
	0.9914206	
	8	

	SVM + PSO	
<i>Hypothesized Mean Difference</i>	0	
<i>df</i>	4	
<i>t Stat</i>	-3.7960442	
<i>P(T<=t) one-tail</i>	0.00958441	
<i>t Critical one-tail</i>	2.13184679	
<i>P(T<=t) two-tail</i>	0.01916881	
<i>t Critical two-tail</i>	2.77644511	

Tabel 4. Hasil Pengujian t-Test SVM dengan Optimasi Lain

Algoritma	p Value of t-Test	Hasil
<i>Particle Swarm Optimization</i>	0.019	Sig. ($p < 0.05$)
<i>Forward</i>	0.328	Not Sig. ($p > 0.05$)
<i>Backward</i>	0.296	Not Sig. ($p > 0.05$)

Hasil pengujian uji-t menunjukkan bahwa pemilihan optimasi *Particle Swarm Optimization* dapat meningkatkan performansi *Support Vector Machine* dalam hal akurasi secara signifikan pada semua *dataset*, hal ini ditandai nilai p dari *t-Test* sebesar 0.019 yang artinya jika nilai p pada *t-Test* < 0.05 memiliki arti signifikan. Dari hasil penelitian ini dapat disimpulkan bahwa *Support Vector Machine* dapat dioptimasi menggunakan *Particle Swarm Optimization* pada lima *dataset* medis yang dijadikan tolak ukur.

4. Kesimpulan

Berdasarkan pada seluruh tahapan yang sudah dilakukan dalam penelitian penanganan 5 *dataset* medis menggunakan metode *Support Vector Machine* berbasis *Particle Swarm Optimization* ini, dapat disimpulkan hasil akurasi menggunakan *Support Vector Machine* diterapkan pada 5 *dataset* yang dijadikan *benchmark* menghasilkan akurasi yang kurang baik. Akan tetapi, setelah dilakukan hasil *testing* menggunakan *Support Vector Machine* dan *Particle Swarm Optimization* menghasilkan peningkatan akurasi yang baik sebesar 1% - 9%. Penerapan *Particle Swarm Optimization* terhadap *Support Vector Machine* dinilai secara signifikan dapat meningkatkan

performansi, hal ini dibuktikan dengan nilai p dari *t-Test* sebesar 0.019.

Berdasarkan pada seluruh tahapan penelitian dan kesimpulan yang telah dilakukan, maka terdapat beberapa usulan pada penelitian berikutnya dapat diterapkan metode yang sama namun pada *dataset* yang berbeda. Kemudian, dapat diterapkan pula metode optimasi lain seperti *Genetic Algorithm*

Referensi

- Azis, A. I. S., Suhartono, V., & Himawan, H. (2017). *Model Multi-Class SVM Menggunakan Strategi 1V1 Untuk Klasifikasi Wall-Following Robot Navigation Data*. 13, 18.
- Barreto, G. de A., Neto, A. R. da R., & Filho, H. A. F. da M. (2011). *Vertebral Column Data Set*. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets/vertebral+column>
- Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). *Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization*. *Procedia Engineering*, 53, 453–462.
- Couceiro, M., & Ghamisi, P. (2016). *Fractional Order Darwinian Particle Swarm Optimization: Applications and Evaluation of an Evolutionary Algorithm*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-19635-0>
- Indrayuni, E. (2016). *Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization*. *EVOLUSI: Jurnal Sains Dan Manajemen*, 4(2).
- Jamsa, K. (2021). *Introduction to Data Mining and Analytics*. Jones & Bartlett Learning LLC.
- Jayadeva, Khemchandani, R., & Chandra, S. (2017). *Twin Support Vector Machines* (Vol. 659). Springer International Publishing. <https://doi.org/10.1007/978-3-319-46186-1>
- Kantardzic, M. (2020). *Data Mining Concept, Model, Method, and Algorithm* (3rd ed.). Wiley-IEEE Press.
- Lichtinghagen, R., Klawonn, H., & Hoffmann, G. (2020). *HCV Data*. UCI Machine Learning Repository.

- <https://archive.ics.uci.edu/ml/dataset/s/HCV+data>
- Listiana, E., & Muslim, M. A. (2017). *PENERAPAN ADABOOST UNTUK KLASIFIKASI SUPPORT VECTOR MACHINE GUNA MENINGKATKAN AKURASI PADA DIAGNOSA CHRONIC KIDNEY DISEASE*.
- Liu, Y., Wen, K., Gao, Q., Gao, X., & Nie, F. (2018). *SVM based multi-label learning with missing labels for image annotation*. 78, 307–317. <https://doi.org/10.1016/j.patcog.2018.01.022>
- Marques, J., & Jossinet, J. (2010). *Breast Tissue Data Set*. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets/breast+tissue#>
- Maurya, R., Singh, S. K., Maurya, A. K., & Kumar, A. (2014). GLCM and Multi Class Support Vector Machine based Automated Skin Cancer Classification. *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*. <https://doi.org/10.1109/IndiaCom.2014.6828177>
- Murty, M. N., & Raghava, R. (2016). *Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-41063-0>
- Nugroho, I. (2020). *Perbandingan Prediksi Fasies Melalui Beberapa Machine Learning*. Pertamina University.
- Palechor, F. M., & Manotas, A. de la H. (2019). *Estimation of Obesity Levels based on Eating Habits and Physical Condition Data Set*. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition+>
- Prasetio, R. T. (2020). Genetic Algorithm to Optimize k-Nearest Neighbor Parameter for Benchmarked Medical Datasets Classification. *Jurnal Online Informatika*, 5(2), 153. <https://doi.org/10.15575/join.v5i2.656>
- Prasetio, R. T., & Ripandi, E. (2019). Optimasi Klasifikasi Jenis Hutan Menggunakan Deep Learning Berbasis Optimize Selection. *Jurnal Informatika*, 6(1), 100–106. <https://doi.org/10.31311/ji.v6i1.5176>
- Sucipto, Kusriani, & Taufiq, E. L. (2016). *Classification Method of Multi-class on C4.5 Algorithm for Fish Diseases*. 5.
- Zarrin, P. S., & Roeckendorf, N. (2020). *Exasens Data Set*. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets/breast+tissue#>