

ANALISA ALGORITMA SUPPORT VECTOR MACHINE PADA DATA BUNGA IRIS

Oleh :

Muhammad Ridho Pradana¹⁾

Rio Septian Hardinata²⁾

Zulfahmi Syahputra³⁾

Fakultas Sains dan Teknologi

Universitas Pembangunan Pancabudi Medan^{1,2,3)}

E-mail:

mhdridho417@gmail.com¹⁾

rioseptianhardinata@gmail.com²⁾

zulfahmisyahputra@gmail.com³⁾

ABSTRACT

The SVM concept can be easily explained as an attempt to find the optimal hyperplane that acts as a delimiter for the two classes in the input space. The use of SVM is very suitable for binary classification problems. The SVM method performs the initial process, namely by inputting predetermined data, then the data is processed by performing a classification process using the SVM method to obtain accuracy results using that method. The initial step taken is to determine the attributes of the data that are used as parameters in classifying and getting data accuracy results. The training and test set (training set and test set) is Iris. This study found that the highest level of accuracy is good from data testing carried out using a data set that has been tested with 150 data and 3 classes of data can lead to the acquisition of analysis. Equalization is a decision-making technique in the process of summarizing the different elements by their weighting. The weighting can be objective by using statistical calculations or subjectively by giving according to certain considerations. This study shows that SVM has good results for solving problems even on small datasets. The results of the analysis using the support vector machine method showed the correct answer rate was 95.00%. These results indicate that SVM is successful and has an effect on the application of attributes to numeric values. The more attributes we use, the slower the time and speed, and the less accurate and complex it becomes.

Keywords: SVM, Iris, Classification, Accuracy

ABSTRAK

Konsep SVM dapat dengan mudah dijelaskan sebagai upaya untuk menemukan hyperplane optimal yang bertindak sebagai pembatas untuk dua kelas di ruang input. Penggunaan SVM sangat cocok untuk masalah klasifikasi biner. Metode SVM melakukan proses awal yaitu dengan menginputkan data yang telah ditentukan, kemudian data tersebut diolah dengan melakukan proses klasifikasi menggunakan metode SVM untuk mendapatkan hasil akurasi menggunakan metode tersebut. Langkah awal yang dilakukan adalah menentukan atribut-atribut data yang dijadikan parameter dalam mengklasifikasikan dan mendapatkan hasil akurasi data. Set pelatihan dan tes (set pelatihan dan set tes) adalah Iris. Penelitian ini menemukan bahwa tingkat akurasi tertinggi baik dari pengujian data yang dilakukan dengan menggunakan kumpulan data yang telah diuji dengan 150 data dan 3 kelas data dapat mengarah pada perolehan analisis. Ekualisasi adalah teknik pengambilan keputusan dalam proses meringkas elemen yang berbeda dengan pembobotannya. Pembobotan dapat bersifat obyektif dengan menggunakan perhitungan statistik atau secara subyektif dengan memberikan menurut pertimbangan tertentu. Studi ini menunjukkan bahwa SVM memiliki hasil yang baik untuk memecahkan masalah bahkan pada dataset kecil. Hasil analisis menggunakan metode support vector machine menunjukkan tingkat jawaban benar sebesar 95,00%. Hasil ini menunjukkan bahwa SVM berhasil dan berpengaruh pada penerapan atribut ke nilai numerik. Semakin banyak atribut yang kita gunakan, semakin lambat waktu dan kecepatannya, dan semakin tidak akurat dan kompleks jadinya.

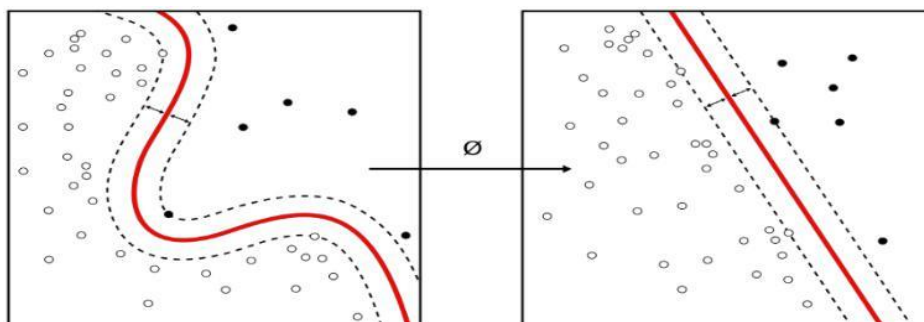
Kata kunci: SVM, Iris, Klasifikasi, Akurasi

1. PENDAHULUAN

Support Vector Machines (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian untuk menyelaraskan konsep yang baik di bidang pengenalan pola. Sebagai teknologi pengenalan pola, SVM masih tergolong baru. Namun, telah dikategorikan sebagai pengenalan pola tingkat lanjut karena kemampuannya dalam berbagai aplikasi dan merupakan salah satu topik yang paling cepat berkembang saat ini. Penelitian (Rachman, 2011) akurasi pada metode SVM lebih unggul daripada metode Naive Bayes. SVM sangat menjanjikan karena merupakan metode berbasis pembelajaran mesin, kuat, dan diterapkan dalam mengklasifikasi dan mengestimasi data. Pola data latih biasanya digunakan untuk menentukan kelas data uji, sehingga menentukan data latih yang dapat mempengaruhi hasil pengujian. Hal ini juga mempengaruhi nilai tertinggi pada precision, recall, dan accuracy dari data latih yang digunakan. (Ridwan, 2013)

2. TINJAUAN PUSTAKA

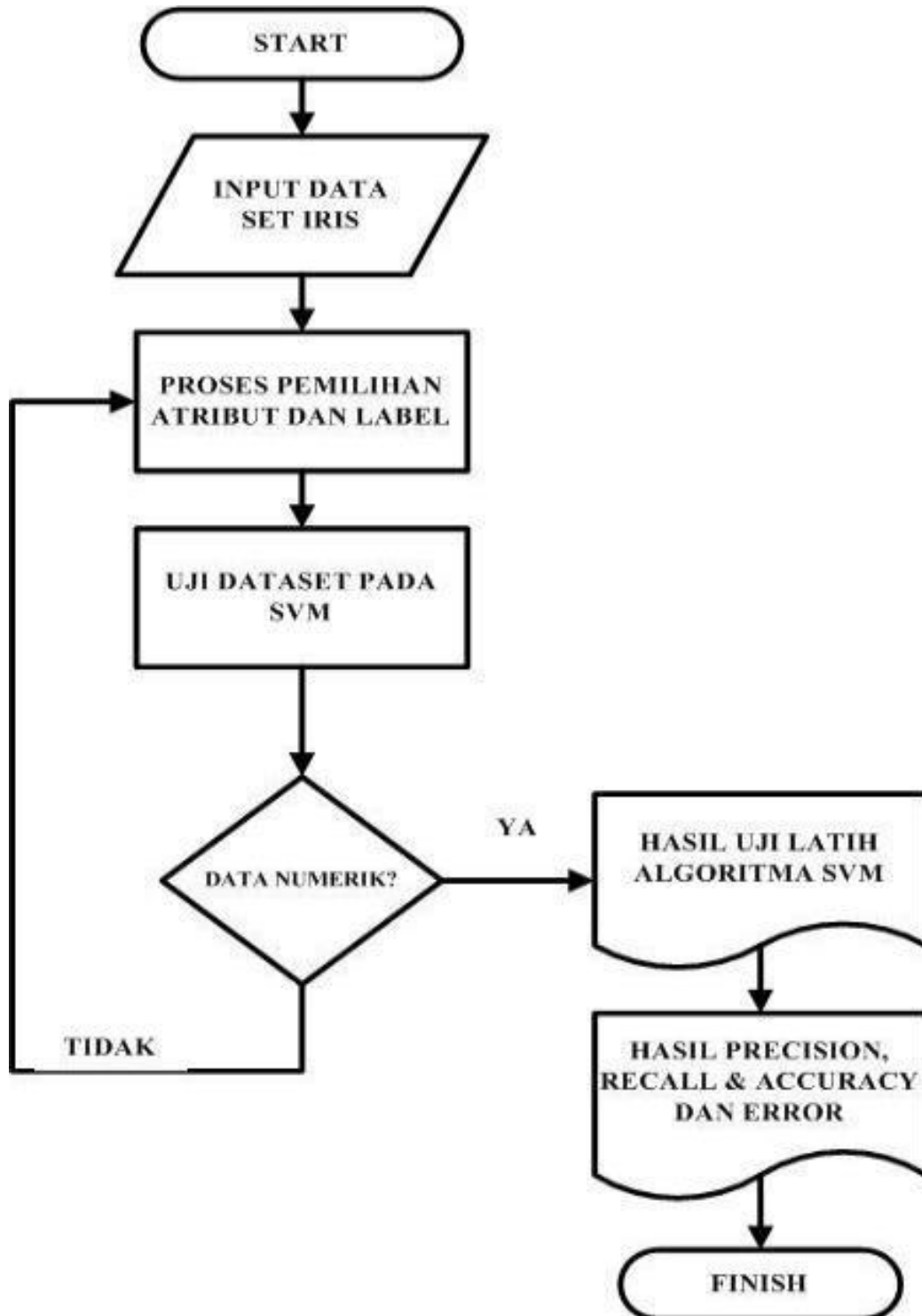
SVM digunakan sebagai salah satu metode pattern recognition, usia SVM terbilang masih relatif muda. Walaupun demikian, evaluasi kemampuannya dalam berbagai aplikasinya menempatkannya sebagai state of the art dalam pattern recognition. SVM merupakan teknik machine learning yang bekerja berdasarkan prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane optimal yang memisahkan dua kelas dalam ruang input. Penelitian ini menjelaskan teori dasar SVM dan aplikasinya dalam ilmu komputer, khususnya dalam analisis ekspresi gen yang diperoleh dari analisis microarray. Konsep SVM dapat dengan mudah dijelaskan sebagai upaya untuk menemukan hyperplane optimal yang bertindak sebagai pembatas untuk dua kelas di ruang input. Penggunaan SVM sangat cocok untuk masalah klasifikasi biner. Adapun contohnya dapat dilihat sebagai berikut:



Gambar 1. *Support Vector Machine*

3. METODE PENELITIAN

Pada bagian ini menjelaskan tentang alur penelitian dengan melakukan metode SVM. Adapun diagram yang digunakan adalah sebagai berikut:



Gambar 2 Rancangan Penelitian

Pada rancangan tersebut dengan menggunakan metode SVM dengan melakukan proses awal yaitu dengan input data yang telah ditetapkan, kemudian data tersebut diproses dengan melakukan proses klasifikasi menggunakan metode SVM untuk mendapatkan hasil akurasi dengan menggunakan metode tersebut. Langkah awal yang dilakukan adalah dengan

melakukan penentuan atribut pada data yang digunakan sebagai parameter dalam melakukan klasifikasi dan mendapatkan hasil akurasi data. Dataset pelatihan dan pengujian (training set and test set) adalah Iris. Rincian setiap catatan dapat ditampilkan pada tabel dibawah:

Tabel 1. Dataset

Data Sets	Instances	Attributes	Classes
Iris	150	4	3

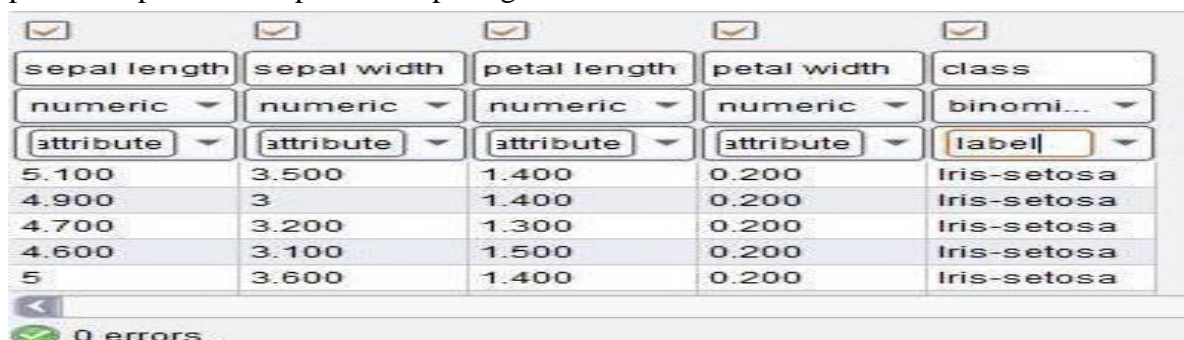
4. HASIL DAN PEMBAHASAN

Data Iris mempunyai data dengan banyak data sebesar 150 data, memiliki atribut dengan banyak 4 atribut dan memiliki kelas dengan banyaknya kelas sejumlah 3 kelas.

Table 2. Data Iris

	Atribut				kelas
	<i>sepal length</i>	<i>sepal width</i>	<i>petal length</i>	<i>petal width</i>	
Jumlah	150	150	150	150	Setosa
Data					Versicolor
					Virginica

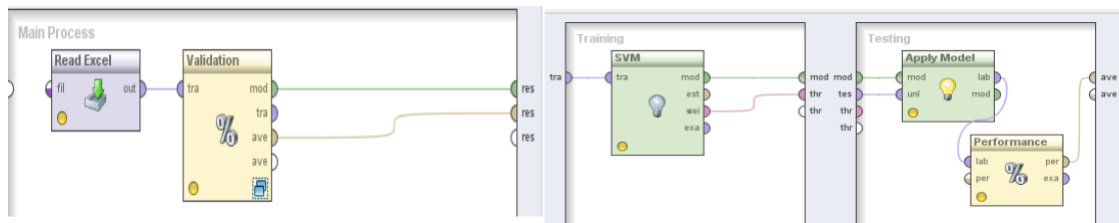
Pada SVM menerapkan sebuah model dan menggunakan dataset yang telah ditentukan dari aplikasi RapidMiner dapat dilihat pada gambar dibawah ini



Gambar 3. Penentuan Atribut Support Vector Machine

Gambar di atas adalah proses menentukan atribut dan menentukan label dengan kelas yang ditentukan dari algoritma SVM yang pada tersebut menggunakan data numerik pada atributnya dan binomial pada kelasnya. Proses tersebut digunakan untuk dapat membentk sebuah model dengan jumlah keberhasilan dengan jumlah data yang ada. Proses ini adalah proses pengolahan data dengan melakukan

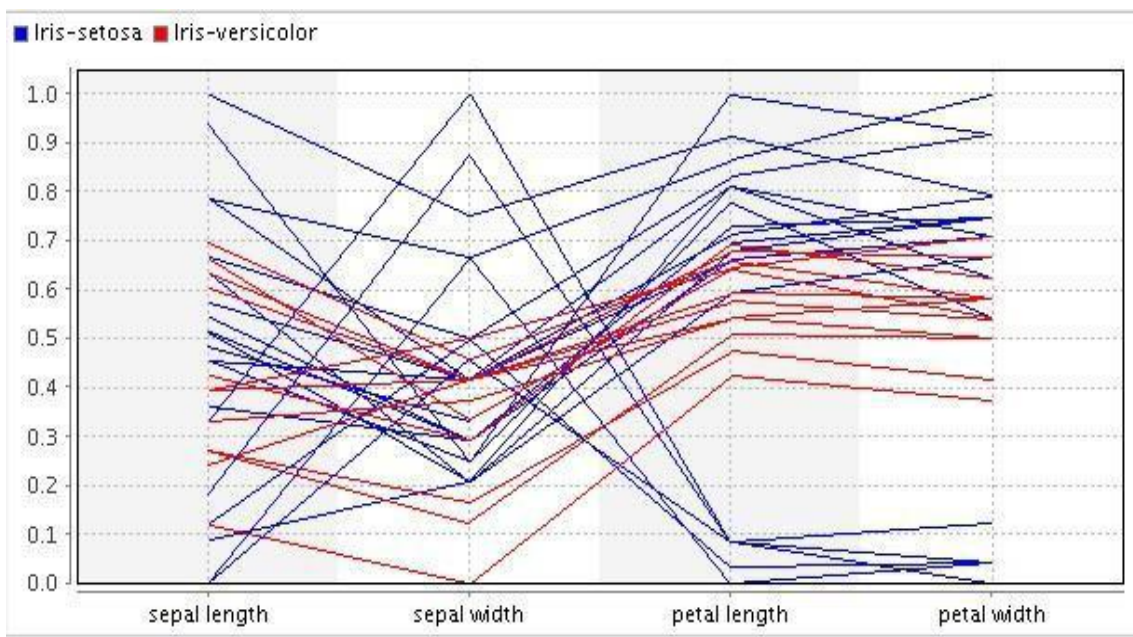
perubahan tipe data integer ke binomial. Proses ini bertujuan agar memperjelas kerja metode dengan output biner negative dan positif tanpa adanya regresi. Hal ini merupakan Analisa statistic pada referensi data yang diuji dengan tujuan dapat melakukan penentuan pengaruh dari 2 variabel atau lebih.



Gambar 4 Perancangan Model

Rancangan model SVM dengan menggunakan Rapidminer dengan menggunakan validasi serta performance classification agar dapat mengetahui

performa dari algoritma yang di uji. Rancangan model ini dilakukan agar data yang di uji dapat diproses dan mendapatkan hasil klasifikasi dan akurasinya.



Gambar 5. Grafik SVM

Pada gambar diatas adalah hasil dari grafik dari SVM dimana iris setosa berwarna biru, dan iris versi color berwarna merah. Terlihat pada gambar tersebut bahwa pengelompokan terjadi pada masing-masing atribut dan classnya. Maka dapat dilihat hasil klasifikasi tersebut dalam bentuk plot diatas. Hasil klasifikasi tersebut merupakan hasil dari klasifikasi terhadap dataset yang telah tersedia dengan banyak jumlah data yang terdapat pada masing-masing atribut adalah sepal length, sepal width, petal length, dan petal width.

Penelitian ini menemukan tingkat akurasi tertinggi yang baik dari pengujian data yang dilakukan menggunakan

kumpulan data yang telah diuji dengan 150 data dan 3 kelas data dapat mengarah pada perolehan analisis. Ekualisasi adalah teknik pengambilan keputusan dalam proses meringkas elemen yang berbeda dengan pembobotannya. Pembobotan dapat bersifat obyektif dengan menggunakan perhitungan statistik atau subyektif dengan pemberian menurut pertimbangan tertentu. Algoritme mesin vektor dukungan memiliki analisis bobot dari kumpulan data yang diuji. Hasilnya pada tabel dibawah ini:

Tabel 3. Weight Metode SVM

attribute	weight
sepal length	-43.31090974585944
sepal width	-23.673558740964314
petal length	-31.83303940079134
petal width	-11.982181949171888

Pembobotan data dilakukan untuk menekankan derajat kandungan informasi dalam data dengan menerapkan metode pembobotan yang diterapkan pada SVM untuk meningkatkan hasil klasifikasi. Penggunaan metode pembobotan dua kali yang menggunakan metode pembobotan untuk membuat antarmuka yang bagus untuk interaksi pengguna dengan mudah.

Tabel berikut menunjukkan hasil pengujian menggunakan model support

vector machine. Tabel diketahui dari 100 tanggal, 49 diklasifikasikan sebagai iris setosa menurut prediksi yang dibuat oleh metode SVM berbasis particle swarm optimization (PSO), dan versi warna memprediksi 4 tanggal, tetapi hasilnya adalah prediksi iris. Nilai 92,45%, satu iris setosa data, dan 46 versi merupakan warna prediksi dengan kelas akurasi 97,87%.
class precision

pred. Iris setosa	49	4	92.45%
pred. Iris versicolor	1	46	97.87%
class recall	98.00%	92.00%	

Tabel 5. Tabel Hasil Analisa SVM berbasis PSO

Counte r	Label	Function Value	Alpha	Abs(AI pha)	Support Vector	Sepal L ength	Sepal Width	PetalLe ngth	Petal Width
0	Iris- setosa	- 2.665049 973	1	1	support vector	4.7	3.2	1.3	0.2
1	Iris- setosa	- 2.049856 06	1	1	support vector	5.5	4.2	1.4	0.2
2	Iris- setosa	- 3.126660 307	1	1	support vector	5	3.5	1.3	0.3
3	Iris- setosa	- 2.861724 04	1	1	support vector	5	3.5	1.6	0.6
4	Iris- versic olor	1 1.526772 02	0.227 0129	0.2270 129	support vector	7	3.2	4.7	1.4
5	Iris- versic olor	1.526772 02	1	1	support vector	5.6	3	4.5	1.5
6	Iris- versic olor	0.476154 245	1	1	support vector	6.3	2.5	4.9	1.5

7	Iris- versicolor	- 0.209301 837	1	1	support vector	6.7	3	5	1.7
8	Iris- versicolor	2.695838 41	1	1	support vector	5.5	2.4	3.7	1
9	Iris- versicolor	- 0.496413 094	1	1	support vector	6	2.7	5.1	1.6
10	Iris- versicolor	1.142046 287	1	1	support vector	6.7	3.1	4.7	1.5
11	Iris- versicolor	2.793463 233	1	1	support vector	5.6	3	4.1	1.3
12	Iris- versicolor	1.924437 27	1	1	support vector	5	2.3	3.3	1
13	Iris- versicolor	2.877977 717	1	1	support vector	5.7	3	4.2	1.2

14	Iris- setosa	- 2.289697 93	1	1	support vector	6.3	3.3	6	2.5
15	Iris- setosa	- 0.331863 728	1	1	support vector	4.9	2.5	4.5	1.7
16	Iris-	- 2.525238	1	1	support	6.7	2.5	5.8	1.8

	setosa	657			vector				
17	Iris-setosa	- 1.508265 302	1	1	support vector	5.7	2.5	5	2
18	Iris-setosa	- 2.287478 787	1	1	support vector	6.5	3	5.5	1.8
19	Iris-setosa	- 1.738040 172	1	1	support vector	7.7	3.8	6.7	2.2
20	Iris-setosa	- 1.670875 551	1	1	support vector	7.7	2.6	6.9	2.3
21	Iris-setosa	- 1.323168 328	1	1	support vector	5.6	2.8	4.9	2
22	Iris-setosa	- 3.145619 324	1	1	support vector	7.2	3.2	6	1.8
23	Iris-setosa	- 0.095675 531	1	1	support vector	6.2	2.8	4.8	1.8
24	Iris-setosa	- 0.411693 436	1	1	support vector	6.1	3	4.9	1.8
25	Iris-setosa	- 2.743399 559	1	1	support vector	7.2	3	5.8	1.6

Hasil pada tabel di atas adalah hasil pengujian RapidMiner. Penelitian ini menunjukkan bahwa SVM memiliki hasil yang baik untuk menyelesaikan masalah bahkan pada dataset dengan jumlah yang kecil. Hasil analisis dengan metode support vector machine menunjukkan tingkat jawaban benar sebesar 95,00%. Hasil ini menunjukkan bahwa SVM berhasil dan berpengaruh terhadap penerapan atribut ke nilai numerik. Semakin bertambah banyak atribut yang kita gunakan, semakin lambat waktu dan kecepatannya, serta semakin tidak akurat dan rumitnya.

5. SIMPULAN

Penulis mencapai kesimpulan sebagai berikut.

1. Hasil penelitian dengan menggunakan dataset Iris mendapatkan hasil prediksi dari metode SVM dengan mendapatkan hasil nilai akurasi yang tinggi.
2. SVM dapat meningkatkan akurasi ketika rata-rata peningkatan akurasi terbaik diperoleh pada dataset iris, dengan hasil akurasi sebesar 95,00%.

6. DAFTAR PUSTAKA

Bramer,M. 2007.
Principles of Data Mining.
 London: Springer.

- Darsyah, M.Y. 2014.
Klasifikasi Tuberkulosis Dengan Pendekatan MetodeSupports Vector Machine (SVM). Jurnal Statistika Unimus, 2(2), 37-41.
- Fiska, Ryci, Rahmatil. 2017. Penerapan Teknik Dara Mining dengan Metode Support Vector Machine untuk memprediksi siswa yang berpeluang Drop Out (Studi Kasus di SMKN 1 Sutura). Sains dan Teknologi Informasi, Volume.3. No. 1.
- Gorunescu,F. 2011. *Data Mining: Concepts, Models and Techniques*. Berlin: Springer-Verlag.
- Han, J.and Kamber, M. 2006.
 “Data Mining Concepts and Techniques Second Edition”. Morgan Kauffman, San Francisco.
- Han, J., Kamber, M. 2001
 “Data Mining Concepts and Techniques”,Morgan Kaufman Pub., USA.
- Han, J.,Kamber,M.,&Pei,J. 2011. *Data Mining: Concepts and Techniques (3rded.)*. San Francisco: Morgan Kaufmann Publishers Inc.
- Pattekari, S. A., Parveen,A.2012.*Prediction System for Heart Disease Using Naive Bayes*. International Journal of Advanced Computer and Mathematical Sciences, ISSN 2230-9624, Vol. 3, No 3, Hal 290-294.

Raikwal. J.S & Saxena, Kanak. 2012.
*Performance Evaluation of SVM
and K-Nearest Neighbor Algorithm
over Medical Data set.* International
Journal of Computer Applications
(0975-8887). Volume.50. No. 14.