

MODEL SUPPORT VECTOR MACHINE BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI PENYAKIT LIVER

Nu'man Musyaffa ¹⁾, Bakhtiar Rifai ²⁾

Sistem Informasi STMIK Nusa Mandiri Jakarta
<http://www.nusamandiri.ac.id>
numan.nmf@nusamandiri.ac.id

Teknik Informatika STMIK Nusa Mandiri Jakarta
<http://www.nusamandiri.ac.id>
bakhtiar.bri@nusamandiri.ac.id

Abstract— The liver is the largest and most important organ for our bodies. We can not live without the heart. Liver disease is an inflammation caused by viral infections, bacteria or toxic substances so that the liver can not perform its function properly and is not easily found in the early stages of diagnosing liver disease. Handling patients with liver disease at an early stage will prolong the patient's life. Many studies use the Support Vector Machine model to predict liver disease but the resulting accuracy is less accurate. In this research, the algorithm model of support vector machine and model vector machine support algorithm based on particle swarm optimization to get the rules to predict liver disease and provide more accurate value of accuracy. After testing with two models of support vector machine and support vector machine based on particle swarm optimization, get the result of algorithm model of support vector machine with accuracy value of 71,36% and accuracy value from AUC equal to 0,500, while test using support vector machine algorithm based on particle swarm optimization get the result of its accuracy value of 77.36% and AUC 0.661 for good diagnosis classification. Both of these methods have varying degrees of accuracy of 6.00% and the difference in the value of AUC 0.161.

Keywords: Liver, Support Vector Machine, Particle Swarm Optimization, Feature Selection.

Intisari— Hati adalah organ yang paling besar dan penting bagi tubuh kita. Kita tidak bisa hidup tanpa hati. Penyakit hati merupakan peradangan yang disebabkan oleh infeksi virus, bakteri atau bahan-bahan beracun sehingga hati tidak dapat melakukan fungsinya dengan baik serta tidak mudah ditemukan dalam tahap awal dalam mendiagnosis penyakit hati. Penanganan pasien dengan penyakit hati pada tahap awal akan

memperpanjang hidup pasien. Banyak penelitian menggunakan model *Support Vector Machine* untuk memprediksi penyakit hati tetapi nilai akurasi yang dihasilkan kurang akurat. Dalam penelitian yang dilakukan ini model algoritma *support vector machine* dan model algoritma *support vector machine* berbasis *particle swarm optimization* untuk mendapatkan aturan untuk memprediksi penyakit liver dan memberikan nilai yang lebih akurat dari akurasi. Setelah pengujian dengan dua model *support vector machine* dan *support vector machine* berbasis *particle swarm optimization*, mendapatkan hasil dari model algoritma *support vector machine* nilai akurasinya 71,36% dan nilai akurasi dari AUC sebesar 0.500, sedangkan pengujian menggunakan algoritma *support vector machine* berbasis *particle swarm optimization* mendapatkan hasil nilai akurasinya sebesar 77,36% dan nilai AUC 0.661 untuk tingkat diagnosa klasifikasi yang baik. Kedua metode ini memiliki berbagai tingkat akurasi sebanyak 6,00% dan selisih nilai AUC 0.161 .

Kata Kunci: Liver, *Support Vector Machine*, *Particle Swarm Optimization*, seleksi atribut.

PENDAHULUAN

Hati adalah organ yang paling besar dan penting bagi tubuh kita. Kita tidak bisa hidup tanpa hati. Penyakit hati merupakan peradangan yang disebabkan oleh infeksi virus, bakteri atau bahan-bahan beracun sehingga hati tidak dapat melakukan fungsinya dengan baik serta tidak mudah ditemukan dalam tahap awal dalam mendiagnosis penyakit hati. Penanganan pasien dengan penyakit hati pada tahap awal akan memperpanjang hidup pasien, namun sayangnya semua ahli kesehatan tidak memiliki keahlian

khusus dalam mendiagnosis medis, sehingga sistem diagnosis otomatis secara medis mungkin akan sangat bermanfaat dan membantu dengan menggunakan machine learning telah banyak digunakan dalam bidang medis. (Widodo, 2014), Hasil penelitian menggunakan support vector machine yang dilakukan (Fridayanthie, 2015) namun nilai akurasi yang dihasilkan kurang akurat, sehingga membutuhkan optimasi berbasis particle swarm optimization (PSO), (PSO) melakukan pencarian menggunakan populasi (swarm) dari individu (partikel) yang diperbaharui dari iterasi. (Handayanna, 2016). Sehingga dalam penelitian yang dilakukan bukan hanya menggunakan metode Support Vector Machine namun juga membutuhkan penggunaan optimasi menggunakan PSO untuk menseleksi atribut agar mendapatkan hasil yang akurat dan presisi

Hasil dari penelitian yang dilakukan diharapkan dapat menjadi rekomendasi serta pertimbangan dan masukan bagi ahli kesehatan dalam membuat prediksi penyakit hepatitis.

BAHAN DAN METODE

Jenis penelitian

Jenis penelitian yang digunakan dalam penelitian ini menggunakan metode *experiment*, yaitu penelitian yang melibatkan penyelidikan kepada beberapa variable menggunakan tes tertentu yang dikendalikan sendiri oleh peneliti. Metode yang digunakan untuk memprediksi adalah Algoritma *Support Vector Machine* dengan *particle swarm optimization* yang akan digunakan untuk melakukan optimasi seleksi atribut.

Algoritma Support Vector Machine

Support Vector Machine (SVM) adalah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi (Maimon, 2012). Data yang tersedia dinotasikan sebagai $x \in R^d$ sedangkan label masing-masing dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \dots, l$ yang mana l adalah banyaknya data. Diasumsikan kedua class -1 dan $+1$ dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan:

Diasumsikan kedua class -1 dan $+1$ dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan:

$$w \cdot x + b = 0 \dots\dots\dots(1)$$

Sebuah pattern x_i yang termasuk class -1 (sampel negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan:

$$w \cdot x + b = -1 \dots\dots\dots(2)$$

sedangkan pattern yang termasuk class $+1$ (sampel positif):

$$w \cdot x + b = +1 \dots\dots\dots(3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $1/||w||$. Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) *problem*, yaitu mencari titik minimal persamaan 4, dengan memperhatikan *constraint* persamaan 5.

$$\min \tau(w) = \frac{1}{2} ||w||^2 \dots\dots\dots(4)$$

$$y_i(x_i \cdot w + b) - 1 \geq 0, \quad \forall i \dots\dots\dots(5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya *Lagrange Multiplier* sebagaimana ditunjukkan pada persamaan 6:

$$L(w, b, \alpha) = \frac{1}{2} ||w||^2 - \sum_{i=1}^l \alpha_i (y_i (x_i \cdot w + b) - 1) \dots\dots\dots(6)$$

α_i adalah *Lagrange multipliers*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan L terhadap w dan b , dan memaksimalkan L terhadap α_i . Dengan memperhatikan sifat bahwa pada titik optimal gradient $L=0$, persamaan langkah 6 dapat dimodifikasi sebagai maksimalisasi *problem* yang hanya mengandung α_i saja, sebagaimana persamaan 7.

$$\text{Maximize: } \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j x_i x_j \dots\dots\dots(7)$$

$$\text{Subject to: } \alpha_i \geq 0 \quad (i=1,2,\dots,l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \dots\dots\dots(8)$$

Dari hasil dari perhitungan ini diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif inilah yang disebut sebagai *support vector*.

1. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) adalah metode pencarian penduduk, yang berasal dari penelitian untuk pergerakan organisme dari kelompok burung atau ikan, seperti algoritma genetika, *Particle Swarm Optimization* (PSO) melakukan pencarian menggunakan populasi (swarm) dari individu (partikel) yang diperbaharui dari iterasi untuk iterasi (Fei, Miao, & Liu, 2009).

Untuk menemukan solusi yang optimal, masing-masing partikel bergerak ke arah posisi sebelumnya terbaik (pbest) dan terbaik posisi global (gbest). Kecepatan dan posisi partikel dapat diperbarui sebagai berikut persamaan:

$$v_{i,m} = w \cdot v_{i,m} + c_1 * R * (pbest_{i,m} - x_{i,m}) + c_2 * R * (gbest_m - x_{i,m}) \quad x_{id} = x_{i,m} + v_{i,m} \dots\dots\dots(9)$$

Dimana:

t = menunjukkan counter iterasi

V_{ij} = kecepatan partikel i pada dimensi ke-j (nilainya terbatas antara $[-v_{max}, v_{max}]$,
 p_{ij} = posisi partikel i pada j dimensi (nilainya terbatas $[-p_{max}, p_{max}]$
 $pbest_{ij}$ = posisi pbest partikel i pada dimensi ke-j
 $gbest_{ij}$ = posisi gbest dari dimensi ke-j
 w = berat inersia (menyeimbangkan eksplorasi global dan lokal eksploitasi)
 $rand_1$ dan $rand_2$ = fungsi acak di rentang $[0, 1]$
 β = faktor kendala untuk mengontrol kecepatan berat (nilainya ke 1)
 c_1 dan c_2 adalah faktor pembelajaran pribadi dan sosial (nilainya ke 2).

Pengumpulan Data

Metode pengumpulan data dibagi menjadi dua sumber data yaitu data primer dan data sekunder. Data primer yaitu data yang dikumpulkan dari sumbernya langsung, sedangkan data sekunder yaitu data yang dikumpulkan dari peneliti sebelumnya. Data yang digunakan pada penelitian ini menggunakan data sekunder. karena diperoleh dari *Indian Liver Patient* dataset dalam UCI (singkatan dari ILPD) url : [http://archive.ics.uci.edu/ml/datasets/ILPD+\(Indian+Liver+Patient+Dataset\)](http://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset)). Masalah yang harus dipecahkan di sini adalah prediksi terjadinya penyakit Liver berisi 583 orang yang diperiksa, dan sebanyak 167 pasien tidak terdeteksi terkena penyakit liver, sehingga 416 pasien terdeteksi penyakit liver. Dengan atribut dari penyakit liver adalah usia, Jenis Kelamin, Total Bilirubin (TB) , Direct Bilirubin (DB), Alkaline phosphotase (Alkphos), Sgpt Alamine (sgpt), Sgot Proteins (Sgot), Total Proteins (TP), Albumin (Alb), rasio albumin dan Globulin (A/G) dan hasil sebagai label yang terdiri atas sakit dan sehat. Data pasien penyakit liver bisa di lihat pada tabel 1. Berikut

Tabel 1. Atribut dan data penyakit liver (UCI Repository)

Usia	Jenis Kelamin	TB	D B	Alkha pos	Sg pt	Sg ot	T P	AL B	Rasio (A/G)	Hasil
65	Wanita	0.7	0.1	187	16	18	6.8	3.3	0.9	Sakit
62	Pria	10.9	5.5	699	64	100	7.5	3.2	0.74	sakit
62	Pria	7.3	4.1	490	60	68	7	3.3	0.89	Sakit
58	Pria	1	0.4	182	14	20	6.8	3.4	1	Sakit
72	Pria	3.9	2	195	27	59	7.3	2.4	0.4	Sakit
46	Pria	1.8	0.7	208	19	14	7.6	4.4	1.3	Sakit
26	Wanita	0.9	0.2	154	16	12	7	3.5	1	Sakit
29	Wanita	0.9	0.3	202	14	11	6.7	3.6	1.1	sakit
17	Pria	0.9	0.3	202	22	19	7.4	4.1	1.2	Sehat
55	Pria	0.7	0.2	290	53	58	6.8	3.4	1	Sakit

57	Pria	0.6	0.1	210	51	59	5.9	2.7	0.8	sakit
72	Pria	2.4	1.3	260	31	56	7.4	3	0.6	Sakit
64	Pria	0.9	0.3	310	61	58	7	3.4	0.9	Sehat
74	Wanita	1.1	0.4	214	22	30	8.1	4.1	1	Sakit
61	Pria	0.7	0.2	145	53	41	5.8	2.7	0.87	Sakit
25	Pria	0.6	0.1	183	91	53	5.5	2.3	0.7	Sehat
38	Pria	1.8	0.8	342	168	441	7.6	4.4	1.3	Sakit
33	10pria	1.6	0.5	165	15	23	7.3	3.5	0.92	sehat
40	Wanita	0.9	0.3	293	232	245	6.8	3.1	0.8	Sakit
40	wanita	0.9	0.3	293	232	245	6.8	3.1	0.8	Sakit

Sumber: Musyaffa & Rifai (2017)

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 583 data, namun tidak semua data dapat digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*).

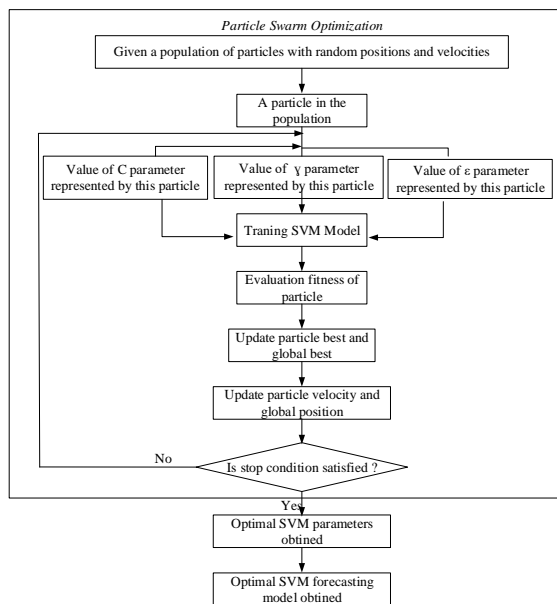
Tabel 2. Tabel Atribut yang digunakan (UCI Repository)

No	Atribut	Fungsi
1	Usia	Menunjukkan Usia (Tahun)
2	Jenis Kelamin	Menunjukkan Jenis kelamin pada pasien (female/male)
3	Total Bilirubin (TB)	Menunjukkan Sel darah merah dihati (mg / dl)
4	Direct Bilirubin (DB)	Bilirubin bebas (mg / dl)
5	Alkaline Phosphotase (alkphos)	Menunjukkan Enzim dari usus
6	Sgpt Alamine (Sgpt)	Menunjukkan Enzim didalam hati
7	Sgot Protiens (Sgot)	Menunjukkan Enzim Protein didalam hati
8	Total Protiens (TP)	Serum protein yang terdapat didalam hati (g / dl)
9	Albumin	Sintesa protein didalam hati
10	Rasio Albumin dan Globulin (Rasio A/G)	Menunjukkan Perbandingan rasio albumin dan globulin didalam hati

Sumber: Musyaffa & Rifai (2017)

Metode yang dilaksanakan

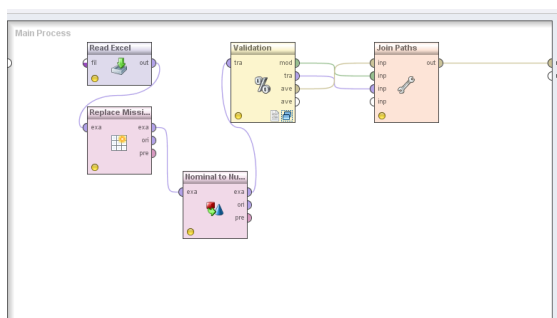
Pada tahap modeling ini dilakukan pemrosesan data traning sehingga akan membahas metode algoritma yang diuji dengan memasukan data penyakit liver kemudian di analisa dan dikomparasi. Berikut ini bentuk gambaran metode algoritma yang akan diuji.



Sumber: Musyaffa & Rifai (2017)
Gambar1. Metode yang dilaksanakan

HASIL DAN PEMBAHASAN

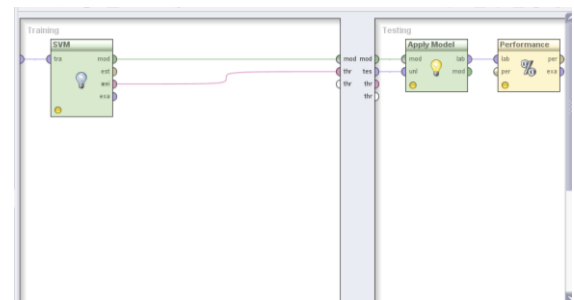
Hasil dari pengujian model yang dilakukan adalah memprediksi penyakit liver dengan *support vector machine* dan *support vector machine* berbasis *particle swarm optimization* untuk menentukan nilai *accuracy* dan *AUC*. Dalam menentukan nilai tingkat keakurasian dalam model *support vector machine* dan algoritma *support vector machine* berbasis *particle swarm optimization*. Metode pengujiannya menggunakan *cross validation* dengan desain modelnya sebagai berikut.



Sumber: Musyaffa & Rifai (2017)
Gambar2. Desain Model Validasi

Eksperimen pengujian Model Support Vector Machine

Pada penelitian penentuan hasil penyakit liver menggunakan algoritma *Support Vector Machine* berbasis pada framework RapidMiner sebagai berikut



Sumber: Musyaffa, Rifai (2018)
Gambar3. Model Pengujian validasi Support vector machine

Nilai *accuracy*, *precision*, dan *recall* dari *data training* dapat dihitung dengan menggunakan Rapid Miner. Hasil pengujian dengan menggunakan model *Support Vector Machine* didapatkan hasil pada tabel 3.

Tabel 3 diketahui dari 583 data, 416 diklasifikasikan *sakit* sesuai dengan prediksi yang dilakukan dengan metode *svm*, lalu 167 data diprediksi *sakit* tetapi ternyata hasilnya prediksi sehat, 0 data *class sehat* diprediksi sesuai, dan 0 data diprediksi tidak ternyata hasil prediksinya.

Tabel 3 Model *Confusion Matrix* untuk Metode *Support Vector Machine*

No	Prediksi		Kelas Presisi	Nilai Akurasi
	Sehat	Sakit		
1	0	416	71.36%	71.36%
2	0	167	71.36%	71.36%

Sumber: Musyaffa & Rifai (2017)

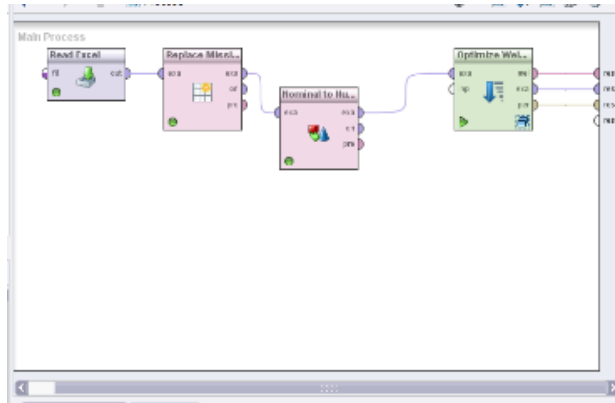
Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bias dilihat pada gambar 4 yang merupakan kurva ROC untuk algoritma *Support Vector Machine*. Kurva ROC pada gambar 4 menekspresikan *confusion matrix* dari table 3. Garis horizontal adalah *false positives* dan garis *vertical true positives*.



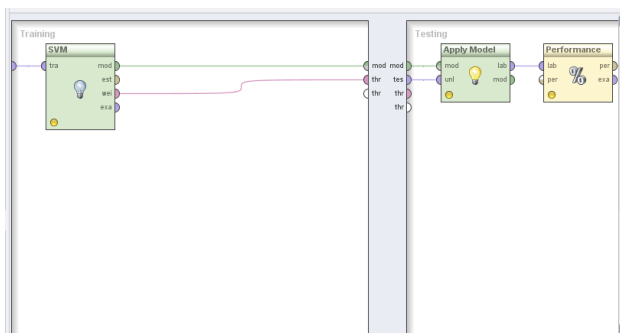
Sumber: Musyaffa & Rifai (2017)
Gambar 4. Kurva ROC algoritma SVM

Eksperimen pengujian Model Support Vector Machine berbasis Algoritma Particle Swarm Optimization (PSO).

Pada penelitian penentuan hasil penyakit Liver menggunakan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO) pada framework RapidMiner sebagai berikut



Sumber: Musyaffa & Rifai (2017)
Gambar 5. Model pengujian validasi *support vector machine* berbasis *Particle Swarm Optimization* (PSO)



Sumber: Musyaffa & Rifai (2017)
Gambar 6. Lanjutan Model pengujian validasi *support vector machine* berbasis *Particle Swarm Optimization* (PSO)

Nilai *accuracy*, *precision*, dan *recall* dari *data training* dapat dihitung dengan menggunakan RapidMiner. Hasil pengujian dengan menggunakan model *Support Vector Machine* didapatkan hasil pada table 4.

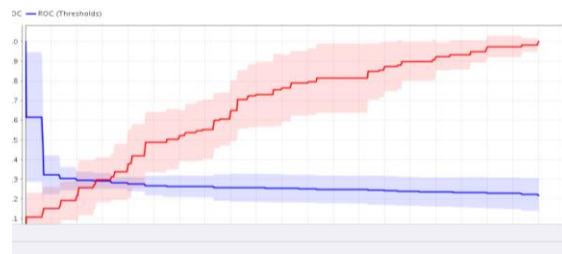
Tabel 4 diketahui dari 583 data, 295 diklasifikasikan *sakit* sesuai dengan prediksi yang dilakukan dengan metode SVM berbasis *Particle Swarm Optimization* (PSO), lalu 112 data diprediksi *sakit* tetapi ternyata hasilnya prediksi sehat, 8 data *class sehat* diprediksi sesuai, dan 1 data diprediksi sehat ternyata hasil prediksinya sakit.

Tabel 4. Model Confusion Matrix untuk metode *Support Vector Machine* berbasis *Particle Swarm Optimization*

No	Prediksi		Kelas Presisi	Nilai Akurasi
	Sehat	Sakit		
1	8	1	88.89%	77.34%
2	112	295	72.48%	77.36%

Sumber: Musyaffa & Rifai (2017)

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bias dilihat pada gambar 4 yang merupakan kurva ROC untuk algoritma *Support Vector Machine*. Kurva ROC pada gambar 4 menekspresikan *confusion matrix* dari table 3. Garis horizontal adalah *false positives* dan garis *vertical true positives*.



Sumber: Musyaffa & Rifai (2017)
Gambar 7. Kurva ROC algoritma SVM berbasis PSO

Dari hasil pengujian diatas, baik evaluasi menggunakan *confusion matrix* maupun *ROC curve* terbukti bahwa hasil pengujian algoritma SVM berbasis PSO memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma SVM. Nilai akurasi untuk model algoritma SVM sebesar 71.36% dan nilai akurasi untuk model algoritma SVM berbasis PSO sebesar 77.36 % dengan selisih akurasi 1.47%, dapat dilihat pada Tabel 5 dibawah ini:

Tabel 5. Pengujian algoritma SVM dan SVM berbasis PSO

No	Algoritma	Hasil	
		Accuracy	AUC
1	SVM	71.36%	0.500
2	SVM berbasis PSO	77.36 %	0.661

Sumber: Musyaffa & Rifai (2017)

Berdasarkan hasil *eksperimen* yang dilakukan untuk memecahkan masalah prediksi hasil prediksi penyakit liver, dapat disimpulkan bahwa hasil *eksperimen* menggunakan metode *support*

vector machine mempunyai tingkat akurasi sebesar 71.36% dan mempunyai nilai AUC sebesar 0.500. Sedangkan *eksperimen* kedua dengan penyesuaian pada parameter C dan epsilon dan *population* didapat nilai akurasi terbaik untuk algoritma yang dilakukan dengan menggunakan metode *support vector machine* berbasis *particle swarm optimization* mempunyai nilai akurasi sebesar 77.36 % dan nilai AUC sebesar 0.661.

KESIMPULAN

Penerapan kebijakan dalam penentuan prediksi penyakit liver di rumah sakit akan membawa pengaruh pada sistem, di mana melalui data-data yang ada rumah sakit akan mempermudah dalam menentukan penyakit Liver. Dengan data set yang ada serta atribut yang telah ditentukan diharapkan penentuan penyakit liver dapat lebih akurat sehingga penyakit liver yang semakin meningkat dari tahun ketahun dapat berkurang jumlah penderitanya. Dan dengan atribut yang sudah ada dan faktor penyebab yang telah diinformasikan diharapkan penderita penyakit liver maupun yang belum terkena serta mempunyai gejala penyakit liver tersebut dapat lebih meperhatikannya sehingga tidak meningkat lagi jumlah penderita penyakit liver.

REFERENSI

Fei, S. W., Miao, Y. B., & Liu, C. L. (2009). Chinese Grain Production Forecasting Method Based on Particle Swarm Optimization-based Support Vector Machine. *Recent Patents on Engineering*, 3(1), 8-12.

Fridayanthie, E. W. (2015). ANALISA DATA MINING UNTUK PREDIKSI PENYAKIT HEPATITIS DENGAN MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE. *Jurnal Khatulistiwa Informatika*, 3(1).
<http://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/1652>

Handayanna, F. (2016). PREDIKSI PENYAKIT DIA BETES MELLITUS DENGAN METODE SUPPORT VECTOR MACHINE BERBASIS PARTICLE SWARM OPTIMIZATION. *Jurnal Teknik Informatika*, 2(1), 30-37.
<http://www.ejournalab.com/index.php/jti/.00..article/view/5>

Maimon, O. (2012). *Data Mining And Knowledge Discovery Handbook*. New York Dordrecht Heidelberg London: Springer.

Musyaffa, N., & Rifai, B. (2017). Laporan Akhir Penelitian Mandiri. Jakarta: STMIK Nusa Mandiri Jakarta

Ramana, B. V., Prabu, S. P., & Venkateswarlu, N. B. 2012, Mei 21. *ILPD (Indian Liver Patient Dataset)*. Retrieved Juni 17, 2013, from MachineLearningRepository:<https://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29>

Widodo, P. (2014). Rule-Based Classifier Untuk Mendeteksi Penyakit Liver. *Bianglala Informatika*, 2(1).
<http://ejournal.bsi.ac.id/ejurnal/index.php/Bianglala/article/view/563>

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.