

KOMPARASI ALGORITMA BERBASIS *NEURAL NETWORK* DALAM MENDETEKSI PENYAKIT HEPATITIS

SURANTO SAPUTRA
surantosaputra@yahoo.com

Program Studi Teknik Informatika
Fakultas Teknik, Matematika dan Ilmu Pengetahuan Alam
Universitas Indraprasta PGRI

***Abstract.** Hepatitis aninfectious disease of the liver and is very dangerous in the world, several studies have been conducted to correctly diagnose patients is unknown but the most accurate method in predicting disease hepatitis in patients. In this study a comparison algorithm Multilayer Perceptron (MLP) and Support Vector Machine (SVM) algorithms to determine the most accurate in predicting disease hepatitis. This study uses secondary data in the form of hepatitis disease data obtained from the University of California Irvine Machine Learning repository of data. Testing the algorithms are carried out by using software that is known that RapidMiner algorithm Support Vector Machine (SVM) has the highest value of accuracy is 90.64%, while the algorithm Multilayer Perceptron (MLP) has an accuracy of 84.38%. Thus the algorithm Support Vector Machine (SVM) to predict the hepatitis disease better.*

KeyWords : Neural Network, Multilayer Perceptron, Support Vector Machine, rapid miner.

Abstrak. Penyakit aninfectious Hepatitis dari liver and sangat berbahaya di dunia, beberapa studi telah dilakukan untuk benar mendiagnosa pasien tidak diketahui tetapi metode yang paling akurat dalam memprediksi hepatitis penyakit pada pasien. Dalam penelitian ini algoritma perbandingan Multilayer Perceptron (MLP) dan Support Vector Machine (SVM) algoritma untuk menentukan yang paling akurat dalam memprediksi hepatitis penyakit. Penelitian ini menggunakan data sekunder berupa data penyakit hepatitis yang diperoleh dari University of California Irvine Machine Learning repository data. Pengujian algoritma dilakukan dengan menggunakan perangkat lunak yang diketahui bahwa algoritma RapidMiner Support Vector Machine (SVM) memiliki nilai tertinggi akurasi adalah 90,64%, sedangkan algoritma Multilayer Perceptron (MLP) memiliki akurasi 84,38%. Jadi algoritma Support Vector Machine (SVM) untuk predic penyakit hepatitis yang lebih baik.

Kata kunci: Neural Network, Multilayer Perceptron, Support Vector Machine, penambang cepat.

PENDAHULUAN

Penyakit hepatitis merupakan urutan pertama dari berbagai penyakit hati diseluruh dunia. Penyakit ini sangat berbahaya bagi kehidupan karena penyakit hepatitis dapat menimbulkan kematian 1-2 juta orang setiap tahunnya. Tantangan yang dihadapi oleh organisasi kesehatan adalah mendiagnosa pasien dengan benar, diagnosa atau prediksi yang buruk dapat menyebabkan konsekuensi yang mendatangkan malapetaka yang kemudian tidak dapat diterima.

Salah satu alternatif sebagai solusi dari masalah tersebut adalah membuat suatu sistem data mining yang bisa melakukan penelusuran pada data historis untuk

mengidentifikasi pola dan memprediksi trend didasarkan pada sifat-sifat yang teridentifikasi sebelumnya, kemudian memberikan alternatif pengobatan atau pencegahan bila ditemukan indikasi yang mengarah pada timbulnya penyakit hepatitis. Informasi yang dihasilkan untuk selanjutnya bisa digunakan oleh edukator hepatitis maupun dokter sebagai dasar untuk melakukan tindakan-tindakan yang diperlukan.

Untuk menjawab tantangan tersebut beberapa penelitian telah dilakukan dalam bidang kesehatan untuk mendapatkan prediksi penyakit dengan lebih akurat, namun belum diketahui metode apa yang paling akurat dalam memprediksi penyakit pasien.

METODE

Jenis penelitian yang digunakan pada penelitian ini adalah sebagai berikut:

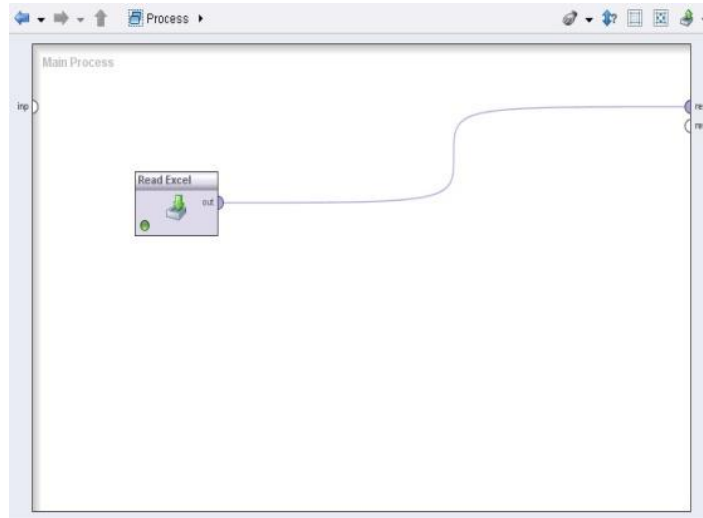
1. Penelitian Eksperimen komparatif penelitian yang dilakukan melalui penyelidikan tentang perlakuan pada parameter dan variabel yang semuanya tergantung pada peneliti itu sendiri.
2. Penelitian ini bertujuan untuk melakukan komparasi dan evaluasi model *Neural Network* menggunakan *Multilayer Perceptron* (MLP) dan *Support Vector Machine* (SVM) untuk mengetahui algoritma yang memiliki keakuratan lebih tinggi dalam memprediksi penyakit hepatitis.

Pada penelitian ini, sumber data diperoleh secara sekunder, yaitu melalui pengumpulan data yang pernah dibuat sebelumnya dan dipublikasi melalui *website*. Data yang digunakan berasal dari sumber sekunder, yaitu data diperoleh dari UCI (*University of California, Irvine*) *Machine Learning Repository* dan diunduh melalui alamat web <http://archive.ics.uci.edu/ml/datasets/Statlog+project> Sampel dari Penelitian ini adalah data profile penderita penyakit hepatitis, data tersebut yang bersifat *public* yang didapatkan dari *University of California Irvine machine learning data repository*.

Dalam melakukan eksperimen ini, penulis menggunakan model *Cross- Industry Standard Process for Data Mining* (CRISP-DM) (SVM). Tahap *modelling* dilakukan dengan menggunakan *software rapid miner*.

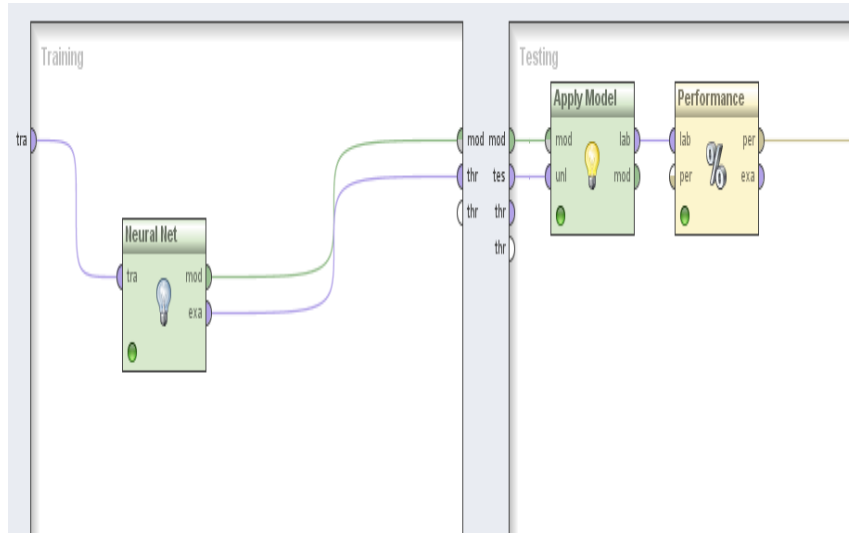
HASIL DAN PEMBAHASAN

Pembuatan model algoritma *Support Vector Machine* (SVM) diawali dengan pembacaan file data (*Read Excell*). *Data training* disimpan dalam satu file *Excell 2003*.



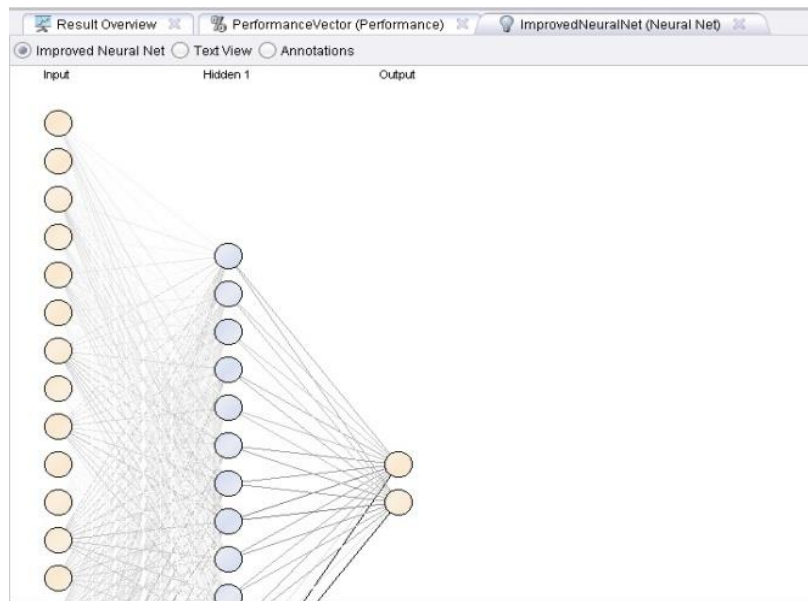
Gambar 1. Desain Model Validasi *Neural Network*

Langkah selanjutnya adalah penentuan model algoritmanya, dalam hal ini adalah *Multilayer Perceptron* (MLP).



Gambar 2. Desain Model Validasi *Multilayer Perceptron* (MLP).

neural net yang dihasilkan dari pengolahan data *training* dengan metode *neural network* adalah *Multilayer Perceptron* yang dihasilkan dari data *training*



Gambar 3. *Neural Network* Prediksi Penyakit Hepatitis

Terdiri dari tiga *layer*, yaitu *Input layer* terdiri dari 20 simpul, sama dengan 19 jumlah atribut prediktor ditambah satu simpul bias. Pada pembahasan ini digunakan satu *hidden layer* yang terdiri dari 12 simpul ditambah satu simpul bias. Di bagian *output layer* terdapat dua simpul yang mewakili atribut kelas yaitu tidak dan sakit.

Untuk setiap data pada data *training*, dihitung input untuk simpul berdasarkan nilai input dan jaringan saat itu. Bobot awal untuk *input layer*, *hidden layer*, dan bias diinisialisasi

secara acak biasanya berkisar antara -0,1 hingga 1,0. Simpul bias terdiri dari dua, yaitu pada *input layer* yang terhubung dengan simpul-simpul pada *hidden layer*, dan pada *hidden layer* yang terhubung pada *output layer*.

Setelah semua nilai awal diinisialisasi, kemudian dihitung masukan, keluaran, dan *error*. Selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid. Setelah didapat nilai dari fungsi aktivasi, hitung nilai *error* antara nilai yang diprediksi dengan nilai yang sesungguhnya. Setelah nilai *error* dihitung, selanjutnya dibalik ke *layer* sebelumnya (*backpropagated*). Nilai *Error* yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi. Hasil perhitungan akhir *backpropagation* fungsi aktivasi untuk simpul pada *hidden layer* terdapat pada lampiran Hidden layer . Kolom pertama pada lampiran *hidden layer* merupakan atribut yang dinyatakan berupa simpul pada *input layer*

Tabel 1 adalah nilai akhir fungsi aktivasi pada *output layer*. Kolom pertama pada Tabel 1 menyatakan *class*, yaitu atribut kelas yang dinyatakan dengan simpul pada *output layer* seperti pada gambar 3 nilai yang terdapat pada kolom berlabel angka satu sampai dua belas adalah nilai bias terbaru yang terdapat pada relasi antara simpul pada *hidden layer* dan simpul pada *output layer*.

Tabel 1. Nilai Bobot Akhir Untuk Output Layer

1	Output (Sigmoid)												Thres hold
	1	2	3	4	5	6	7	8	9	10	11	12	
Tidak	1.033	2.290	2.037	2.203	0.974	1.636	1.505	1.562	2.427	2.168	1.513	1.448	-5.210
Sakit	-1.036	-2.308	-2.030	-2.136	-0.984	-1.639	-1.459	-1.604	-2.457	-2.142	-1.486	-1.528	5.210

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan *learning rate*. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training cycles*:

Tabel 2. Eksperimen Penentuan Nilai *Training Cycle MultiLayer Perceptron*

<i>Learning Rate</i>	MLP	
	<i>Accuracy</i>	AUC
0.3	84,38%	0,327
0.9	84,38%	0,324

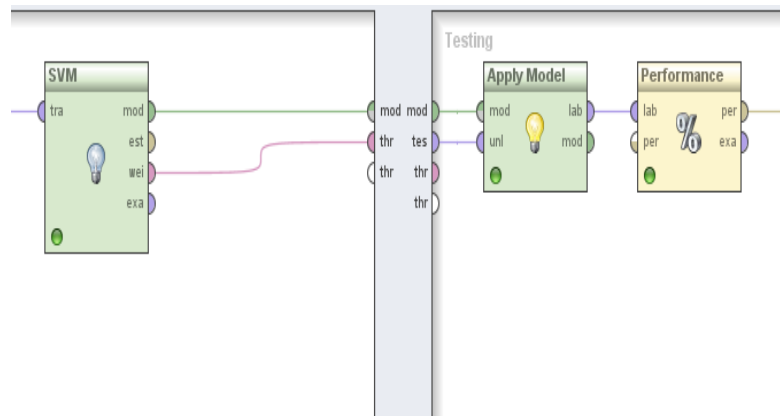
Hasil terbaik pada *eksperimen Multilayer Perceptron* (MLP) dengan menggunakan data training diatas adalah dengan *Learning Rate* =0.3 dihasilkan *accuracy* 84,38% dan AUC 0.327 seperti pada table 2 sedangkan dengan data testing Mempunyai nilai *accuracy* dan AUC yang sama yaitu *accuracy* 81,25% dan AUC 0.958 seperti pada table 3.

Tabel 3. Eksperiment Penentuan Nilai *Testing Cycle Multilayer Perceptron*

<i>Learning Rate</i>	MLP	
	<i>Accuracy</i>	AUC
0.3	81.25%	0,958
0.9	81.25%	0,958

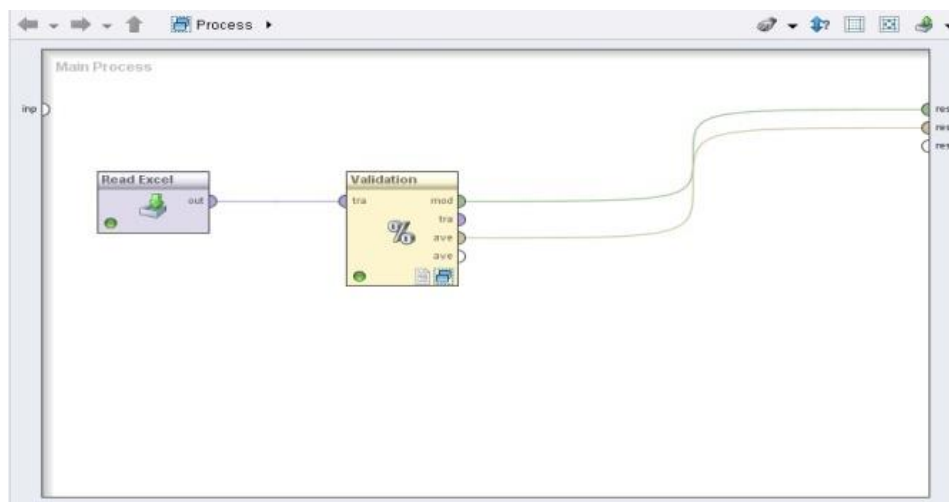
Algoritma *Support Vector Machine (SVM)*

Pembuatan model algoritma *Support Vector Machine (SVM)* diawali dengan pembacaan file data (*Read Excell*). *Data training* disimpan dalam satu file *Excell 2003*. Langkah selanjutnya adalah penentuan model algoritmanya, dalam hal ini adalah *Support Vector Machine (SVM)*.



Gambar 4. Model Algoritma *Support Vector Machine (SVM)*

Data tersebut kemudian divalidasi



Gambar 5. Desain Model *Support Vector Machine*

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C, epsilon. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training cycle*. Hasil terbaik pada *eksperiment Support*

Vector Machine (SVM) diatas adalah dengan $C = 0.0$ dan $Epsilon = 0.0$ dihasilkan *accuracy* 90,64% dan AUC 0.838 sedangkan dengan data testing adalah dengan $C=1.0$ dan $Epsilon=0.0$ menghasilkan nilai *accuracy* 87.50% dan AUC 0,958

Evaluasi dan Validasi Hasil

Dalam penulisan ini misalkan, metode yang digunakan, yaitu algoritma *Multilayer Perceptron* (MLP) dan *Support Vector Machine* (SVM), kemudian dilakukan komparasi keduanya dan mengukur metode mana yang paling akurat. Metode klasifikasi bisa dievaluasi berdasarkan beberapa kriteria seperti tingkat akurasi, kecepatan, kehandalan, skalabilitas, dan interpretabilitas^[7].

Penelitian ini bertujuan untuk mengukur tingkat akurasi dua algoritma dalam memprediksi penyakit hepatitis, kemudian menganalisa akurasi dengan membandingkan kedua metode tersebut.

Pengujian Model

Model yang telah dibentuk diuji tingkat akurasinya dengan memasukan data uji yang berasal dari data *training*. Karena data yang didapat dalam penelitian ini setelah proses *preprocessing* hanya 64 data maka digunakan metode *Xvalidation* untuk menguji tingkat akurasi. Untuk nilai akurasi model untuk metode *Multilayer Perceptron* (MLP) sebesar 84.38% dan metode *Support Vector Machine* (SVM) sebesar 90,64%.

Selain itu dalam penelitian ini diuji juga dengan menggunakan *confusion matrix* dan kurva ROC.

Confusion Matrix

Evaluasi dengan menggunakan model *confusion matrix* dilakukan pada kedua model algoritma yang telah terbentuk. Perhitungan berdasarkan data *training*, diketahui dari 64 data, 53 diklasifikasikan tidak, sesuai dengan prediksi yang dilakukan dengan metode *Multilayer Perceptron* (MLP), lalu 1 data diprediksi sakit. Perhitungan berdasarkan data *testing*, diketahui dari 16 data, 10 diklasifikasikan tidak, sesuai dengan prediksi yang dilakukan dengan metode *Multilayer Perceptron* (MLP), lalu 3 data diprediksi sakit.

Confusion matrix untuk metode *Support Vector Machine* (SVM) dengan menggunakan data *training*. Diketahui dari 64 data, 55 diklasifikasikan tidak, sesuai dengan prediksi yang dilakukan dengan metode *Support Vector Machine* (SVM), lalu 3 data diprediksi sakit. *Confusion matrix* untuk metode *Support Vector Machine* (SVM) dengan menggunakan data *testing*. Diketahui dari 16 data, 10 diklasifikasikan tidak, sesuai dengan prediksi yang dilakukan dengan metode *Support Vector Machine* (SVM), lalu 3 data diprediksi sakit.

Sebelum diterapkan pada data baru, terlebih dahulu dilakukan pengujian akurasi terhadap model yang telah terbentuk, hasil pengujian dengan menggunakan data *training* dan data *testing*. Pengujian akurasi dengan menggunakan data *training* diperoleh algoritma *Support Vector Machine* (SVM) memiliki akurasi paling tinggi, yaitu sebesar 90,64%. Sedangkan jika menggunakan data *testing*, algoritma *Multilayer Perceptron* (MLP) dan *Support Vector Machine* (SVM) memiliki akurasi nilai yang sama yaitu sebesar 81,25%.

Kurva ROC

Hasil perhitungan dari kedua buah algoritma divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.327 dimana diagnosa hasilnya *Excellent classification*. Kurva ROC mengekspresikan *confusion matrix* Garis X adalah *false positives* dan garis Y *true positives*. Hasil yang didapat dari pengolahan ROC untuk algoritma *Multilayer*

Perceptron (MLP) dengan menggunakan data *testing* sebesar 0.958 dengan tingkat diagnosa *excellent classification*

Kurva ROC untuk algoritma *Support Vector Machine* (SVM) yang terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.838 dimana diagnosa hasilnya *fair classification*.

Hasil yang didapat dari pengolahan ROC untuk algoritma Algoritma *Support Vector Machine* (SVM) sebesar 0.958 dengan tingkat diagnosa *excellent classification*

Setelah dilakukan perhitungan pada kurva ROC maka didapatkan hasil perbandingan tersebut. Perbandingan hasil perhitungan nilai AUC untuk metode *Multilayer Perceptron* (MLP) dan *Support Vector Machine* (SVM) dapat dilihat pada Tabel 12.

Tabel 4 Komparasi Nilai AUC

	<i>Multilayer Perceptron</i> (MLP)		<i>Support Vector Machine</i> (SVM)	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
AUC	0.327	0.958	0.838	0.958

Analisis Hasil Komparasi

Model yang dihasilkan dengan metode *Multilayer Perceptron* (MLP) dan *Support Vector Machine* (SVM) diuji menggunakan metode *Cross Validation*, terlihat perbandingan nilai *accuracy* pada Tabel 4, untuk metode *Multilayer Perceptron* (MLP) memiliki nilai *accuracy* yang paling tinggi dibandingkan dengan metode *Support Vector Machine* (SVM). Tabel 5 membandingkan *accuracy* dan AUC dari tiap metode. Terlihat bahwa nilai *accuracy Support Vector Machine* (SVM) paling tinggi begitu pula dengan nilai AUC-nya.

Tabel 5. Komparasi Nilai Accuracy dan AUC

	<i>Multilayer Perceptron</i> (MLP)	<i>Support Vector Machine</i> (SVM)
Accuracy	84.38 %	90.64 %
AUC	0.906	0.991

Penerapan Algoritma Terpilih

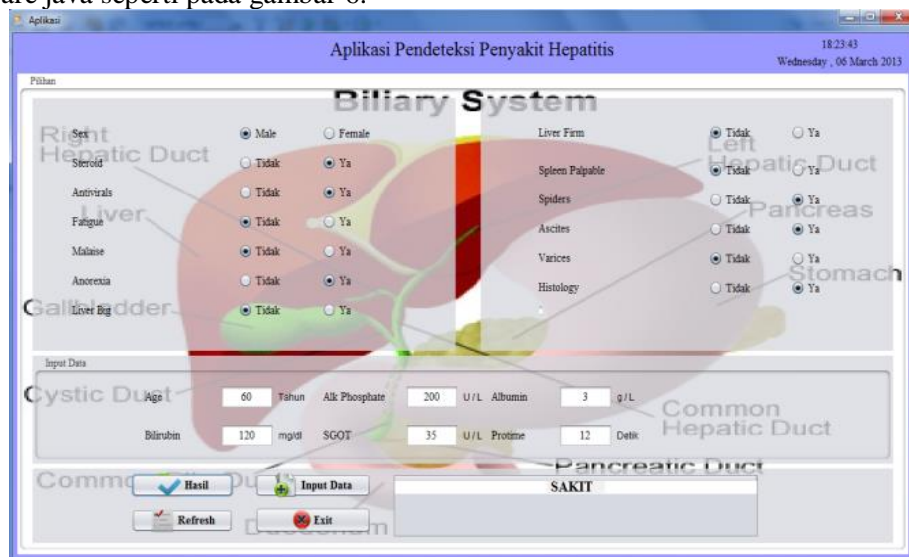
Berdasarkan hasil perbandingan akurasi pada tabel 5, algoritma terpilih sebagai algoritma terbaik dalam prediksi penyakit hepatitis yaitu algoritma *Support Vector Machine* (SVM) yang memiliki tingkat akurasi yang paling tinggi dengan persentase 90.64%. Berdasarkan algoritma yang terpilih maka selanjutnya dilakukan penerapan pada data baru.

Hasil penerapan *rule* algoritma *Support Vector Machine* (SVM) terhadap data baru sejumlah 10 record data dimana 5 data diprediksi sakit dan 5 data diprediksi tidak. Dengan tingkat akurasi sebesar 100 % seperti pada tabel 6.

Tabel 6. Model *Confusion Matrix Support Vector Machine (SVM)* Menggunakan Data Baru

accuracy: 100.00% +/- 0.00% (mikro: 100.00%)			
	true Tidak	true Sakit	class precision
pred. Tidak	5	0	100.00%
pred. Sakit	0	5	100.00%
class recall	100.00%	100.00%	

Rule hasil klasifikasi dari algoritma *Support Vector Machine (SVM)* diterapkan kedalam pembuatan aplikasi untuk mendeteksi penyakit hepatitis dengan menggunakan software java seperti pada gambar 6.



Gambar 6. Aplikasi untuk mendeteksi penyakit hepatitis

Pada aplikasi untuk mendeteksi penyakit hepatitis pada gambar 6 dihasilkan klasifikasi sakit dan tidak. Input data pasien pada program tersebut sesuai dengan atribut yang dibutuhkan, kemudian klik tombol Hasil, maka secara otomatis tampil hasil sakit atau tidak. Untuk menginput kembali data baru klik tombol Input Data, untuk keluar dari aplikasi klik tombol Exit.

PENUTUP

Simpulan

Penelitian ini memkomparasi keakuratan dua buah algoritma data mining dalam memprediksi penyakit hepatitis, algoritma yang dikomparasikan yaitu *Multilayer Perceptron (MLP)* dan *Support Vector Machine (SVM)*. Kedua algoritma tersebut dievaluasi dan divalidasi dengan menggunakan *confusion matrix*, dan kurva ROC.

Dari hasil evaluasi dan validasi diketahui bahwa algoritma *Support Vector Machine (SVM)* memiliki akurasi yang paling tinggi yaitu sebesar 90,64%., diikuti oleh algoritma *Multilayer Perceptron (MLP)* yang memiliki akurasi 84,38%. Dengan demikian,

algoritma *Support Vector Machine* (SVM) dapat memprediksi penyakit hepatitis dengan baik, sehingga bisa digunakan untuk membantu dokter atau tenaga medis didalam pengambilan keputusan untuk memprediksi penyakit hepatitis.

Saran

Agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

1. Hasil penelitian ini diharapkan bisa digunakan pada rumah sakit atau instansi lain yang terkait untuk lebih meningkatkan akurasi dalam memprediksi penyakit hepatitis.
2. Untuk mendukung pengambilan keputusan dan pengembangan sistem informasi manajemen strategik, model ini dapat diterapkan pada rumah sakit atau instansi lain yang terkait dengan menerapkan sistem yang menggunakan perangkat keras dan perangkat lunak, disertai dengan pembuatan *Standard Operational Procedure* dan pelatihan bagi *end-user*.
3. Penelitian semacam ini dapat dikembangkan pada unit bisnis serupa atau yang lainnya. Penelitian ini dapat dikembangkan dengan algoritma yang lain misalkan saja dengan metode statistik lainnya seperti *Decision Tree* (C4.5), *Naive Bayes*, *Radial Basis Function* (RBF) juga untuk metode *soft computing* lainnya seperti algoritma genetika.
4. Dalam penelitian ini dapat pula dikembangkan bahasa pemrogramnya dengan menggunakan bahasa pemrograman lainnya seperti bahasa pemrograman Delphi, Visual Basic dan Hypertext Preprocessing (PHP).

DAFTAR PUSTAKA

- Bhuvanewari dan Kalaiselvi (2012). **Naive Bayesian Classification Approach in Healthcare Applications**. *International Journal of Computer Science and Telecommunications Volume 3, Issue 1, January* , 106-112.
- Han. J., dan Kember. M. (2006). **Data Mining Concepts and Techniques**. San Fransisco: Morgan Kauffman
- Larose D. T. (2005). **Discovering Knowledge in Data**. New Jersey: John Willey & Sons, Inc
- Larose D. T. (2006). **Data Mining Methods and Models**. Canada: John Wiley dan Sons, inc.
- Myatt. G. J. (2007). **Making Sense of Data A Practical Guide to Exploratory Data Analysis and Data Mining**. New Jersey: Wiley-Interscience.
- Venkates P. & S. Anitha. (2006). **Application of a radial Basis Function Neural Network for Diagnosis of Diabetes Mellitus**, Chennai, India: Current Science, Vol. 91, No. 9.
- Vercellis C. (2009). **Business Intelligent: Data Mining and Optimizzation for Decision Making**. Southern Gate, Chichester, West Sussex, United Kingdom : John Wiley & Sons Ltd.
- Haykin, S. (1999). **Neural networks a comprehensive foundation**, Second Edition, Upper Saddle River, N.J.: Prentice-Hall International, Inc.
- Kusriani, dan Luthfi. T. E. (2009). **Algoritma Data Mining**. Yogyakarta: Penerbit Andi.
- Dong. Y., Xia. Z., Tu. M., dan Xing. G. (2007). **An Optimization Method For Selecting Parameters In Support Vector Machines**. *Sixth International Conference On Machine Learning And Applications* , 1.
- Huang. K., Yang. H., King. I., dan Lyu. M. (2008). **Machine Learning Modeling Data Locally And Globally**. Berlin Heidelberg: Zhejiang University Press, Hangzhou And Springer-Verlag GmbH.

Gorunescu F. (2011). **Data Mining Concepts, Models and Techniques**. Berlin Heidelberg: Springer Verlag.

Bramer. M. (2007). **Principles of Data Mining**. London: Springer

Ester Monica. (2002) . **Keperawatan Medikal Bedah**. Jakarta: EGC, 93

Kothari. C. R. (2004). **Research Methology Methods and Techniques**. India: New Age International Limited.

Sugiyanto. (2008). **Metode Penelitian Kuantitatid Kualitatif R dan D**. Bandung: Alfabeta.

Cohen, L., Manion, L., & Morrison, K. (2005). **Research Methods in Education 5th Edition**. London: The Taylor & Francis e-Library