

PREDIKSI PENYAKIT LIVER DENGAN MENGGUNAKAN METODE DECISION TREE DAN NEURAL NETWORK

Popon Handayani¹, Elah Nurlelah², Mugi Raharjo³, Panji Madya Ramdani⁴

^{1,2,3}STMIK Nusa Mandiri Jakarta

Jln. Damai No 8, Warung Jati, Marga Satwa, Pasar Minggu, Jakarta Selatan, Indonesia

⁴Universitas BSI

Jln. Kamal Raya No.18 Cengkareng, Jakarta Barat, Indonesia

¹popon.pph@nusamandiri.ac.id, ²elah.enl@bsi.ac.id, ³mugi.mou@bsi.ac.id, ⁴panji.pmr@bsi.ac.id

Page | 55

Abstrak— Penyakit Liver merupakan penyakit dimana disebabkan oleh berbagai faktor yang merusak hati, seperti virus, penggunaan alkohol dan lainnya. Dalam hal penanganan pasien pada tahap awal sangatlah penting untuk kelangsungan pasien dan penyakit hati tidak mudah ditemukan pada stadium awal. Untuk itu kami melakukan sebuah penelitian dengan menggunakan dua metode yaitu metode Decision Tree dan Metode Neural Network untuk mengetahui nilai akurasi. Berdasarkan hasil perbandingan, diperoleh neural network terbaik dalam mendeteksi penyakit hati.

Keywords— Liver, Decision Tree, Neural Network..

I. PENDAHULUAN

Kesehatan merupakan salah satu hal terpenting dalam hidup manusia, Hal ini menjadi dasar bahwasanya banyak ditemukannya temuan-temuan ilmiah baik berupa temuan obat-obatan, alat kesehatan atau penemuan-penemuan baru di bidang kesehatan. Berbicara mengenai kesehatan tentunya berhubungan dengan penyakit, sejak dahulu banyak sekali penyakit-penyakit bermunculan entah itu datangnya dari virus, bakteri, parasite, sel kanker dan lain-lain. Salah satunya adalah organ tubuh hati, Hati penting untuk mencerna makanan, menyingkirkan tubuh kita dari zat beracun dan menyimpan energi tubuh untuk digunakan bila diperlukan. Dalam industri kesehatan dan medis keakuratan prediksi sebuah penyakit sangatlah penting dan memerlukan keputusan yang efektif dalam mengambil suatu analisa dan keakuratan prediksi suatu penyakit yang diderita pasien (Rifai, 2013).

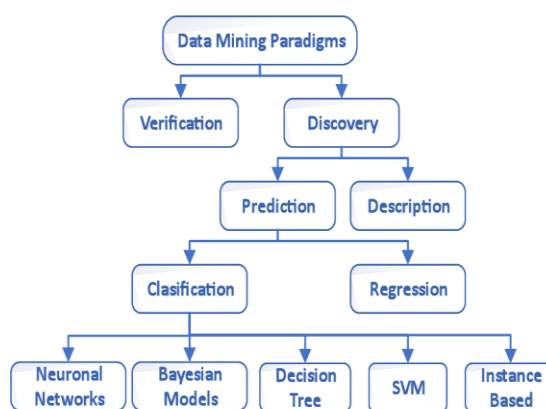
Dalam mendiagnosis ada atau tidak penyakit liver dapat digunakan acuan dari hasil tes fungsi hati yang dilaksanakan di laboratorium. Tes tersebut antara lain yaitu *transaminase serum*, fosfatase alkali, total bilirubin, bilirubin terkonjugasi, total protein, albumin, serta rasio albumin dan globulin. Dari hasil tes tersebut dapat dilihat hasil tes yang signifikan sebagai ciri adanya gangguan fungsi hati dengan menggunakan algoritma pohon klasifikasi karena dapat memperoleh informasi mengenai data klasifikasi pasien penyakit liver (Abdurrahman et al 2014).

Salah satu metode yang dapat digunakan Untuk memprediksi penyakit hipertensi adalah dengan menggunakan data mining. Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Muzakir et al, 2016).

Teknik dalam penelitian ini menggunakan dalam *data mining* adalah Algoritma *decision tree* (pohon keputusan). Dengan menggunakan teknik *decision tree* menggunakan algoritma C4.5, akan menghasilkan pola, yaitu berupa *rules* yang dapat memprediksi penyakit liver. Pada studi ini juga akan menggunakan algoritma *neural network* dengan optimalis diharapkan hasil dari prediksi yang dilakukan bisa lebih akurat. Dan bisa digunakan untuk prediksi penyakit liver lebih baik

II. TINJAUAN PUSTAKA

Banyak metode Data Mining yang dipergunakan dengan berbagai tujuan dan hasil yang dicapai. *Predictive analytics* merupakan proses penggunaan data, algoritma statistic dan *machine learning* untuk mengidentifikasi kemungkinan hasil berdasarkan *historical data*. Dimana *prediction* adalah salah satu cabang dari *discovery data mining* [1].



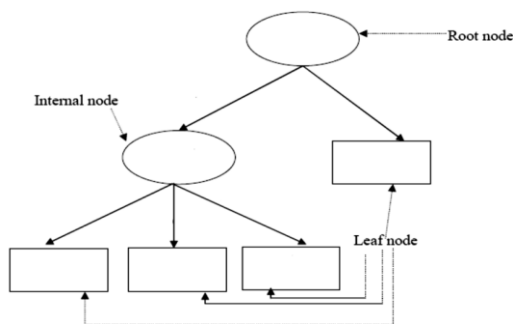
Gbr 1. Data Mining Taxonomy [1]

Pada Gambar 1 menunjukkan data mining terdiri dari dua jenis yaitu *data mining verification oriented* berorientasi melakukan verifikasi data dan *data*

mining discovery oriented digunakan untuk menemukan aturan baru atau pola mandiri dari sebuah set data. Metode *prediction* merupakan turunan dari *data mining discovery oriented* yang bertujuan membangun model perilaku untuk memperoleh sampel baru yang mampu memprediksi nilai-nilai dari satu atau lebih variabel yang terkait dengan sample. Selain itu prediksi dikembangkan untuk memperoleh pola yang membentuk pengetahuan dengan cara yang dapat dimengerti dan mudah dioperasikan. *Data mining prediction* terdiri dari teknik *classification* dan *regression*, algoritma yang dapat digunakan untuk teknik *classification* antara lain *Neural Networks*, *Bayesian Models*, *Decision Tree*, *Support Vector Machine* (SVM), dan *Instance Based*.

Classification adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep/kelas data sehingga dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dimana model ini berupa aturan "jika-maka", berupa *decision tree*, formula matematis atau *neural network*. Proses *classification* dibagi menjadi 2 (dua) fase yaitu fase *learning* dan *test*, pada fase *learning* sebagian data yang telah diketahui kelas datanya dipergunakan untuk membentuk model perkiraan, yang selanjutnya masuk ke dalam fase *test* dimana model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Dan apabila akurasinya mencukupi, model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

Pada *decision tree* membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Setiap simpul daun pada *decision tree* menandai label kelas, dimana simpul yang bukan simpul akhir terdiri dari akar dan simpul internal yang terdiri dari kondisi tes atribut pada sebagian *record* yang mempunyai karakteristik yang berbeda. Simpul akar dan simpul internal ditandai dengan bentuk oval dan simpul daun ditandai dengan bentuk segi empat [2]. Berikut adalah struktur *decision tree* seperti yang ditunjukkan pada Gambar 2.



Gbr 2. Struktur *Decision Tree*

Ada beberapa konsep dalam *decision tree*, antara lain [2]:

1. Data dinyatakan dalam bentuk Tabel dengan *atribut* dan *record*.

2. *Atribut* menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan *tree*. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi *per-item* data yang disebut dengan target atribut.
3. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*.

Dalam pengembangan data mining teknik *classification* dengan *decision tree* dalam algoritma C4.5 terdapat beberapa tahapan sebagai berikut [3] :

1. Mempersiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus :

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i$$

Keterangan :

S = Himpunan kasus

n = jumlah partisi S

Pi = proporsi Si terhadap S

Kemudian hitung nilai gain menggunakan rumus

$$Gain(S,A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

Keterangan :

S = Himpunan Kasus

A = Fitur

n = Jumlah partisi atribut A

|Si| = Proporsi Si terhadap S

|S| = Jumlah kasus dalam S

3. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi.
4. Proses partisi *decision tree* akan berhenti saat :
 - a. semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam *record* yang dipartisi lagi
 - c. Tidak ada record didalam cabang yang kosong

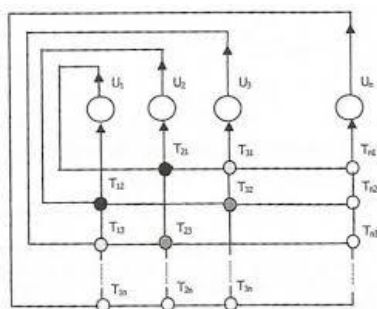
Neural networks atau *artificial neural networks* telah menjadi alat standar yang penting untuk *data mining* dan banyak digunakan untuk tugas-tugas *data mining* seperti *pattern classification*, *time series analysis*, *prediction* dan *clustering*. Dimana *Neural networks* atau *artificial neural networks* merupakan salah satu kelas dalam pemodelan kuantitatif yang populer dikalangan peneliti dan praktisi telah

digunakan selama 20 tahun terakhir dan telah berhasil diterapkan untuk menyelesaikan berbagai variasi masalah di hampir semua bidang bisnis, industri dan sains [4].

Salah satu model neural network yang biasa dipergunakan yaitu model *Hopfield Neural Network* merupakan jenis *neural networks* yang mampu menyimpan *memories* atau pola tertentu dengan cara yang mirip dengan otak. *Hopfield Neural Network* dikembangkan oleh John Hopfield (Ahli Fisika, pemenang nobel 1982). *Hopfield Neural Network* mempunyai sifat *associative memory*, ini berarti seperti ingatan yang mampu menyimpan informasi yang pernah diberikan dan menampilkan kembali informasi tersebut apabila diminta. *Hopfield Neural Network* juga bersifat *high-fault tolerance* yaitu kemampuan untuk memilih salah satu pola yang paling mirip dalam ingatannya apabila pola yang diberikan tidak sama seperti salah satu pola dalam pembelajaran. Hopfield merupakan jaringan syaraf tiruan yang terhubung penuh (*fully connected*), yang berarti bahwa setiap unit terhubung dengan setiap unit lainnya. Jaringan ini memiliki bobot – bobot yang simetris. Pada Jaringan Hopfield, setiap unit tidak memiliki hubungan dengan dirinya sendiri. Secara matematik hal ini memenuhi $w_{ij} = w_{ji}$ untuk $i \neq j$, dan $w_{ij} = 0$ untuk $i = j$.

Jaringan Hopfield biner mempunyai suatu lapisan unit pengolah. Setiap unit pengolah mempunyai sebuah nilai aktivitas atau kondisi yang bersifat biner. Keseluruhan jaringan mempunyai kondisi pada setiap saat. Kondisi tersebut berupa vektor dari bilangan 0 dan 1. Setiap anggota vektor sesuai dengan unit pengolah dalam jaringan. Jadi pada setiap kondisi jaringan dapat digambarkan dengan vektor sebagai berikut: $U = (U_1, U_2, U_3, \dots, U_n) = (+ + \dots + \dots +)$ Vektor ini mencerminkan sebuah jaringan yang terdiri dari n unit pengolah, dimana elemen ke i mempunyai kondisi u_i . Dalam notasi ini, + menggambarkan sebuah unit pengolah dengan nilai 1 dan -1 menggambarkan sebuah unit pengolah dengan nilai 0. [5]

Tiap hubungan mempunyai bobot. Bobot ini adalah suatu skalar yang berdasarkan pada kekuatan hubungan (*Connection Strength*). Kita misalkan T_{ji} merupakan bobot dari unit i ke unit j. Dalam jaringan Hopfield bobot T_{ij} dan T_{ji} mempunyai nilai yang sama sehingga $T_{ji} = T_{ij}$. [6]



Gbr 3. Jaringan Hopfield dengan Lingkaran Menunjukkan Nilai Bobot T_{ji} [6]

Dalam jaringan Hopfield, yang dipandang sebagai sistem pengingat asosiatif, proses belajar adalah proses penambahan ingatan (*memories*) dari pola yang dilatihkan. Dalam sistem ini prosedur pembelajaran adalah prosedur pembentukan bobot koneksi T_{ij} . Jika pada jaringan akan dilatihkan sejumlah m pola maka persamaan bobotnya adalah:

$$T_{ji} = \begin{cases} \sum_{p=1}^m P_{j,p} P_{i,p} & \text{Jika } i \neq j \\ 0 & \text{Jika } i = j \end{cases}$$

Dengan: T_{ji} = Bobot koneksi dari neuron i dengan neuron j m = Jumlah pola yang dilatihkan $P_{i,p}$ = Elemen ke i dari vector pola P yang dilatihkan Pada awalnya matriks bobot bernilai 0 ($T = 0$). Pemberian nilai bobot (proses pembelajaran) dilakukan dengan menerapkan persamaan diatas untuk sejumlah pola yang dilatihkan sehingga terbentuk suatu matriks bobot. Selanjutnya jika ada pola yang akan dilatihkan lagi, persamaan tersebut diterapkan dengan kondisi bobot yang telah terbentuk sebelumnya sebagai kondisi awal. Dalam jaringan Hopfield ini terdapat batasan kapasitas penyimpanan ingatan dari pola yang dilatihkan. Untuk jaringan dengan neuron sejumlah N, agar sistem stabil, batasan kapasitas pola yang diingat adalah $2N$. Awalnya pada jaringan ditetapkan kondisi untuk tiap unit pengolah. Prosedur pembaharuan diterapkan pada unit – unit dalam jaringan Hopfield. Pada satu saat, satu unit diperbaharui. Prosedur pembaharuan mempengaruhi kondisi tiap unit, kadang mengubah kondisinya dan kadang dibiarkan pada kondisi yang sama. Pembaharuan dari unit – unit pengolah berlanjut sampai tidak terjadi lagi perubahan. Ketika penjumlahan ini selesai dihitung, unit pengolah akan mengevaluasi apakah hasil penjumlahan lebih besar atau lebih kecil dari 0. Jika hasil penjumlahan lebih besar atau sama dengan 0, maka keluaran unit di set 1. Jika hasil penjumlahan lebih kecil dari 0, keluaran unit akan di set 0.

Tiap kondisi dari jaringan mempunyai sekumpulan energi. Nilai ini didefinisikan dengan:

$$E = -\frac{1}{2} \sum_j \sum_{i \neq j} T_{ji} U_j U_i$$

Dengan:

T_{ji} = Bobot koneksi dari neuron i dengan neuron j
 U_j = Unit pengolah pada neuron ke j

Persamaan ini dihubungkan dengan “energi” meskipun ini tidak menggambarkan energi yang nyata seperti pada sistem fisika. Fungsi energi dari persamaan diatas adalah sebuah fungsi objektif yang diminimalkan oleh jaringan. Pembaharuan dari jaringan Hopfield merupakan prosedur konvergen dimana energi dari keseluruhan jaringan akan menjadi semakin kecil. Pada akhirnya jaringan akan berada

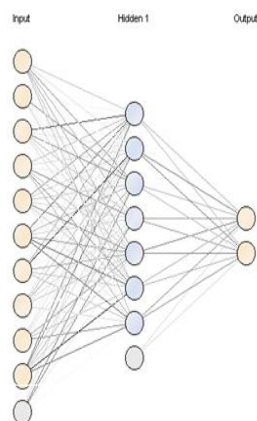
pada posisi stabil dimana pada kondisi ini energi berada pada nilai minimum [6].

III. PEMBAHASAN

Pembuatan model C4.5 dilakukan pada dataset yang terdiri dari 11 atribut yang merupakan atribut dari diagnosis penyakit liver dan *class* yang merupakan hasil akhir prediksi. Data kemudian di validasi agar proses pelatihan dapat berjalan dengan cepat dan mampu digunakan untuk melakukan pelatihan.

Model dari algoritma C4.5 yaitu berupa pohon keputusan, untuk dapat membuat pohon keputusan, langkah pertama adalah menghitung jumlah *class* yang terkena penyakit liver yang positif liver dan yang negatif liver dari masing-masing *class* berdasarkan atribut yang telah ditentukan dengan menggunakan data *training*. Kemudian menghitung *Entropy* (Total) menggunakan persamaan (2.1)

Data training yang digunakan untuk model Neural Network sama, namun nilai atribut merupakan data yang sudah dikonversi kedalam nilai numerik. Gambar 4.6 adalah neural net yang dihasilkan dari pengolahan data training dengan metode neural network adalah multilayer perceptron. Terdiri dari tiga layer, yaitu Input layer terdiri dari sepuluh neuron (sembilan neuron terdiri dari atribut dan satu neuron adalah bias), satu buah hidden layer yang terdiri dari delapan buah neuron, dan dua buah output layer yang merupakan hasil prediksi Positif Liver dan Negatif Liver, berikut model Neural Network dengan Framework Rapidminer versi 5.2.001. Sehingga jika model tersebut dijalankan maka akan didapatkan hasil seperti gambar 4 dibawah ini.



Gambar 4. Neural net yang dihasilkan dengan metode Neural Network

Untuk setiap data pada data training, dihitung input untuk neuron berdasarkan nilai input dan jaringan. Bobot awal untuk input layer, hidden layer, dan bias diinisialisasi secara acak biasanya berkisar - 0,1 s/d 1,0 (kusrini,2009.p,201). Nueron bias terdiri dari dua, yaitu pada input layer yang terhubung dengan neuron-neuron pada hidden layer, dan pada

hidden layer yang terhubung pada output layer. Setelah semua nilai awal diinisialisasi, kemudian dihitung input dari tiap neuron untuk membangkitkan output dengan 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).

Komparasi Model Decision Tree dengan Algoritma Neural Network

Hasil pengujian menyimpulkan bahwa Atribut yang mempengaruhi pengujian adalah 11 atribut. Hasil pengujian model decision tree dan algoritma Neural Network dapat dilihat pada tabel berikut:

TABEL I
PENGUJIAN ALGORITMA KLASIFIKASI DECISION TREE DAN NEURAL NETWORK

	Accuracy	AUC
C4.5	75.56%	0.898
Neural Network	74,17%	0,671

Dari hasil pengujian diatas, dengan dilakukan evaluasi baik secara *confusion matrix* maupun *ROC curve* terbukti bahwa pengujian yang dilakukan algoritma C4.5 memiliki nilai akurasi yang lebih tinggi dibanding menggunakan algoritma Neural Network. Nilai akurasi untuk model algoritma C4.5 sebesar 75,56% dengan nilai AUC 0,898 dan nilai akurasi algoritma Neural Network sebesar 74,17% dengan AUC 0,671. Berdasarkan nilai tersebut diperoleh selisih akurasi sebesar 1,39% dan selisih AUC sebesar 0,227%.

IV. KESIMPULAN

Algoritma Decision Tree dan Neural Network dapat diimplementasikan untuk membuat sebuah system dalam memprediksi penyakit liver. Algoritma Decision Tree memiliki nilai true positive lebih besar dibandingkan Neural Network, ini artinya tingkat memprediksi untuk penyakit liver lebih akurat dengan menggunakan Decision Tree dari pada Neural Network.

REFERENSI

- [1] Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*. London: Springer.
- [2] Muzakir, A., & Wulandari R.A., (2016). Model *Data Mining* sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik *Decision Tree*. *Scientific Journal of Informatics*; Vol. 3, No. 1
- [3] Larose, D. T. (2005). *Discovering Knowledge in Data An Introduction to Data Mining*. New Jersey: John Willey and Sons.
- [4] Widrow B., Rumelhart D.E., Lehr M.A., (1994), Neural networks: applications in industry, business and science, *Communications of the ACM*; 37:93-105

- [5] Azmi, Z., & Freizello, H., (2017), Implementasi Hopfield untuk Pengenalan Karakter Hiragana. Journal of Information System, Informatics and Computing; Vol.1, No. 1
- [6] Hermawan, A. (2006), Jaringan Saraf Tiruan Teori dan Aplikasi. Yogyakarta: Penerbit Andi.