

Komparasi Metode Klasifikasi Data Mining untuk Prediksi Penyakit Ginjal

Sutri Handayani

Abstract— Data mining offers methodological and technical solutions for the prediction and classification of diseases, thus data mining has an important role in the health world. Main objectives of this study is to predict chronic kidney disease using classification algorithms such as Support Vector Machine and Algorithm C4.5. This study focuses on the discovery of appropriate algorithms for classification. From the results obtained, the classification algorithm C4.5 better than the SVM.

Intisari— Data mining menawarkan metodologi dan solusi teknis untuk prediksi dan klasifikasi suatu penyakit, dengan demikian data mining memiliki peranan yang penting dalam dunia kesehatan. Tujuan Utama dari penelitian ini adalah untuk memprediksi penyakit ginjal kronis dengan menggunakan algoritma klasifikasi seperti Support Vector Machine dan Algoritma C4.5. Penelitian ini berfokus pada penemuan algoritma yang tepat dan baik untuk klasifikasi. Dari hasil yang didapatkan, klasifikasi menggunakan algoritma C4.5 lebih baik dibandingkan dengan SVM.

Kata Kunci — ginjal, algoritma support vector machine, algoritma C4.5

I. PENDAHULUAN

Data mining adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang database. Keduanya memerlukan penyaringan melalui sejumlah besar material, atau menyelidiki dengan cerdas untuk mencari keberadaan sesuatu yang bernilai[1].

Ginjal adalah sepasang organ saluran kemih yang terletak di rongga retroperitoneal bagian atas. Bentuknya menyerupai kacang dengan sisi cekungnya menghadap ke media. Pada sisi ini, terdapat hilus ginjal, yaitu tempat struktur-struktur pembuluh darah, sistem limfatik, sistem saraf, dan ureter menuju dan meninggalkan ginjal.[2].

Tujuan utama dari penelitian ini adalah untuk mengetahui algoritma yang tepat dan baik dalam klasifikasi penyakit ginjal dengan menggunakan algoritma klasifikasi yaitu SVM (*Support Vector Machine*) dan algoritma C4.5.

II. KAJIAN LITERATUR

Dr. S. Vijayarani et al [3] Data mining adalah peran penting dalam beberapa aplikasi seperti organisasi bisnis, pendidikan lembaga, sektor pemerintah, industri perawatan kesehatan, ilmiah dan rekayasa. Dalam perawatan kesehatan

industri, data mining terutama digunakan untuk prediksi penyakit. Tujuan utama dari riset ini adalah untuk memprediksi penyakit ginjal menggunakan algoritma klasifikasi seperti Naïve Bayes dan Support Vector Machine. Penelitian ini bekerja terutama berfokus pada menemukan algoritma klasifikasi terbaik berdasarkan ketepatan klasifikasi dan waktu pelaksanaan faktor kinerja. Dari hasil percobaan teramati bahwa kinerja SVM lebih baik dari algoritma klasifikasi Naive Bayes.

Lakshmi.K.R et al[4] Tujuan utama dari penelitian ini adalah untuk mengembangkan model prediksi untuk penyakit ginjal dan untuk ,mendapatkan algoritma terbaik dari tehnik data mining medis. Dalam penelitian ini, tiga teknik data mining (Artificial Neural Networks, Keputusan pohon dan Regresi Logical) digunakan untuk memperoleh pengetahuan tentang interaksi antara variabel-variabel ini dan kelangsungan hidup pasien. Perbandingan kinerja tiga tehnik data mining digunakan untuk mengekstraksi pengetahuan dalam bentuk klasifikasi. Konsep yang diperkenalkan dalam penelitian ini telah diterapkan dan diuji menggunakan data yang dikumpulkan di situs dialisis yang berbeda. Berdasarkan hasil perhitungan, ANN disarankan untuk mendapatkan hasil yang lebih baik dengan akurasi dan kinerja.

Andrew Kusiak et al[5] Pada penelitian ini, data yang preprocessing, transformasi data, dan pendekatan data mining digunakan untuk memperoleh pengetahuan tentang interaksi antara banyak dari parameter yang diukur. Algoritma data mining digunakan untuk mengekstraksi pengetahuan dalam bentuk aturam-aturan keputusan, algoritma data mining memprediksi kelangsungan hidup pasien, Penelitian ini telah diterapkan dan diuji menggunakan data yang dikumpulkan diempat lokasi dialisis dan hasilnya dilaporkan.

A. Data Mining

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu [6]. Kaskas data mining meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatisasi yang dilakukan oleh data mining melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak digunakan[7]. Secara khusus, koleksi metode yang dikenal sebagai 'data mining' menawarkan metodologi dan solusi teknis untuk mengatasi analisis data medis dan konstruksi prediksi model[8].

Jurusan Teknik Informatika, STMIK Nusa Mandiri Jakarta, Jl. Kramat Raya No.25, Jakarta Pusat 10450 INDONESIA (telp: 021-31908575; fax: 021-31908565; e-mail: sutri.sfx@bsi.ac.id)

B. SVM

SVM adalah sebuah metode seleksi yang membandingkan parameter standar seperangkat nilai diskrit yang disebut kandidat set, dan mengambil salah satu yang memiliki akurasi klasifikasi terbaik[9]. SVM adalah salah satu alat yang paling berpengaruh dan kuat untuk memecahkan klasifikasi[10]. Support Vector Machines (SVM) adalah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi [11]. Dengan berorientasi pada tugas, kuat, sifat komputasi mudah dikerjakan, SVM telah mencapai sukses besar dan dianggap sebagai state-of-the-art classifier saat ini[12].

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:

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

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

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

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

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

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).

$$\min \tau(w) = 1/2\|w\|^2 \quad (4)$$

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

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya Lagrange Multiplier sebagaimana ditunjukkan pada persamaan 2.6:

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

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)

$$\text{Maximize: } \sum_{i=1}^l \alpha_i - \sum_{i=1}^l \alpha_i y_i \cdot y_j, x_i, x_j \quad (7)$$

Subject to:

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

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.

C. Algoritma C4.5

Algoritma C4.5 adalah hasil dari pengembangan algoritma ID3 (*Iterative Dichotomiser*) yang dikembangkan oleh Quinlan. Algoritma C4.5 atau pohon keputusan mirip sebuah pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan dengan mudah dapat dikonversi ke aturan klasifikasi. Secara umum keputusan pengklasifikasi pohon memiliki akurasi yang baik, namun keberhasilan penggunaan tergantung pada data yang akan diolah. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5[1] yaitu:

- 1) Mempersiapkan data training, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
- 2) Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai index entropy terendah. Sebelumnya dihitung terlebih dahulu nilai index entropy, dengan rumus:

$$\text{Entropy}(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j) \quad (9)$$

Keterangan:

i = himpunan kasus

m = jumlah partisi i

$f(i,j)$ = proposi j terhadap i

- 3) Hitung nilai gain dengan rumus:

$$\text{Entropy split} = - \sum_{i=1}^p \frac{n_i}{n} \cdot IE(i) \quad (10)$$

Keterangan:

p = jumlah partisi atribut

n_i = proporsi n_i terhadap i

n = jumlah kasus dalam n

- 4) Ulangi langkah ke-2 hingga semua record terpartisi
Proses partisi pohon keputusan akan berhenti disaat:
 - a) Semua tupel dalam record dalam simpul m mendapat kelas yang sama
 - b) Tidak ada atribut dalam record yang dipartisi lagi
 - c) Tidak ada record didalam cabang yang kosong.

D. Ginjal

Berdasarkan *National Kidney Foundation(NKF) Kidney Disease Outcome Quality Initiative (K/000) Guidelines*

Update tahun 2002, definisi penyakit ginjal kronis(GGK)[2] adalah :

1. Kerusakan ginjal >3 bulan, berupa kelainan struktur ginjal, dapat atau tanpa disertai penurunan laju filtrasi glomerulus(LFG) yang ditandai dengan kelainan patologi, dan adanya pertanda kerusakan ginjal, dapat berupa kelainan laboratorium darah atau urine, atau kelainan radiologi.
2. LFG <60 mL/menit/1,73 m² selama 3> bulan, dapat disertai atau tanpa disertai kerusakan ginjal.

E. Pengujian K-Fold Cross Validation

K-Fold Cross Validation adalah teknik validasi yang membagi data kedalam k bagian dan kemudian masing-masing bagian akan dilakukan proses klasifikasi. Dengan menggunakan K-Fold Cross Validation akan dilakukan percobaan sebanyak k. Tiap percobaan akan menggunakan satu data testing dan k-1 bagian akan menjadi data training, kemudian data testing itu akan ditukar dengan satu buah data training sehingga untuk tiap percobaan akan didapatkan data testing yang berbeda-beda. Data training adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakuratan hasil pembelajaran [6].

F. Validasi

Validasi adalah proses mengevaluasi akurasi prediksi dari sebuah model, validasi mengacu untuk mendapatkan prediksi dengan menggunakan model yang ada kemudian membandingkan hasil yang diperoleh dengan hasil yang diketahui[1]. Mengevaluasi akurasi dari model klasifikasi sangat penting, akurasi dari sebuah model mengindikasikan kemampuan model tersebut untuk memprediksi class target [13]. Untuk mengevaluasi model digunakan metode confusion matrix, dan kurva ROC (Receiver Operating Characteristic).

1. Confusion Matrix

Evaluasi kinerja model klasifikasi didasarkan pada pengujian objek yang diprediksi dengan benar dan salah, hitungan ini ditabulasikan Confusion Matrix[1]. Confusion Matrix adalah alat yang berguna untuk menganalisis seberapa baik classifier dapat mengenali tupel dari kelas yang berbeda[14]. Confusion matrix memberikan rincian klasifikasi, kelas yang akan diprediksi akan ditampilkan di bagian atas matrix dan kelas yang diobservasi ditampilkan di bagian kiri[1].

2. Kurva ROC

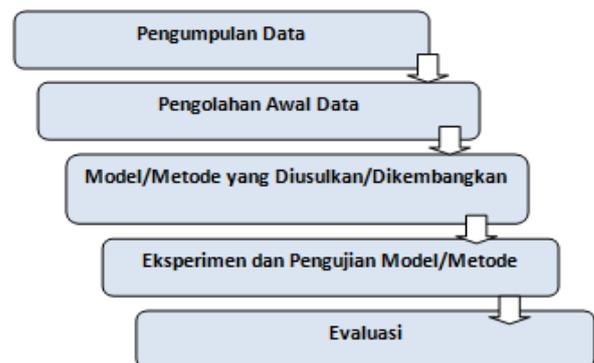
Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan confusion matrix. ROC adalah grafik dua dimensi dengan false positives sebagai garis horizontal dan true positives untuk mengukur perbedaan performansi metode yang digunakan[15]. ROC Curve adalah cara lain untuk menguji kinerja pengklasifikasian[1]. Kurva ROC digunakan untuk mengukur AUC (Area Under Curve). AUC dihitung untuk mengukur perbedaan performansi metode yang digunakan. Kurva ROC membagi hasil positif dalam sumbu y dan hasil negatif dalam sumbu x.

Sehingga semakin besar area yang berada dibawah kurva, semakin baik pula hasil prediksi. Nilai AUC dibagi menjadi beberapa kelompok[1]:

- a. 0.90 - 1.00 = *Excellent Classification*
- b. 0.80 - 0.90 = *Good Classification*
- c. 0.70 - 0.80 = *Fair Classification*
- d. 0.60 - 0.70 = *Poor Classification*
- e. 0.50 - 0.60 = *Failure*

III. METODE PENELITIAN

Metode yang digunakan adalah metode penelitian eksperimen dengan tahapan penelitian sebagai berikut:



Gbr.1 Tahapan Penelitian

Berdasarkan gambar diatas, berikut langkah yang dilakukan dalam proses penelitian:

1. Pengumpulan data
Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses.
2. Pengolahan awal data
Pada tahap ini dilakukan pengolahan data dari sumber data yang telah dikumpulkan meliputi penyeleksian data, data dibersihkan dan ditransformasikan ke bentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model.
3. Model yang diusulkan
Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Adanya jenis data dibagi menjadi data latihan (training data) dan data uji (testing data) juga diperlukan untuk pembuatan model.
4. Eksperimen dan pengujian model
Pada tahap ini model yang diusulkan akan diuji untuk melihat hasil berupa rule yang akan dimanfaatkan dalam pengambilan keputusan.
5. Evaluasi
Pada tahap ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakuratan model.

IV. HASIL DAN PEMBAHASAN

A. Tahap Pengumpulan Data

Teknik pengumpulan data ialah teknik atau cara-cara yang dapat digunakan untuk menggunakan data[16]. Dalam pengumpulan data terdapat sumber data, sumber data yang terhimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder [16]. Data yang diperoleh adalah data sekunder karena diperoleh dari UCI Repository. Jumlah pasien berjumlah 400 orang dengan 25 atribut diantaranya age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, anemia, class.

B. Tahap Pengolahan Awal Data

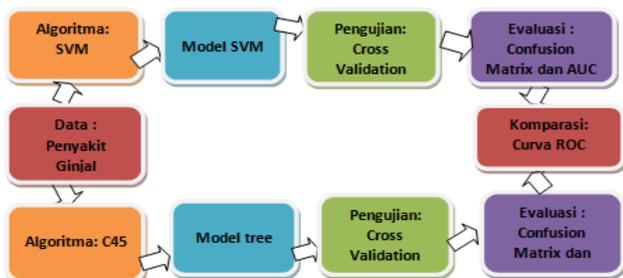
Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 400 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut [13].

1. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
2. Data integration and transformation, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam software Rapidminer.
3. Data size reduction and discretization, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informative.

Setelah dilakukan tehnik Validasi terhadap data penyakit ginjal, maka diperoleh sample data training seperti yang terlihat di Tabel 1.

C. Metode Yang Diusulkan

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



Gbr.2 Metode Usulan

D. Eksperimen dan Pengujian Metode

Metode penelitian eksperimen dapat diartikan sebagai metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu terhadap yang lain dalam kondisi yang terkendalikan[17]. Misalnya pengaruh ruang kerja terhadap produktivitas kerja karyawan.

Berikut tabel hasil data training sehingga didapatkan hasil akurasi tertinggi :

Tabel.2
Hasil Data Training

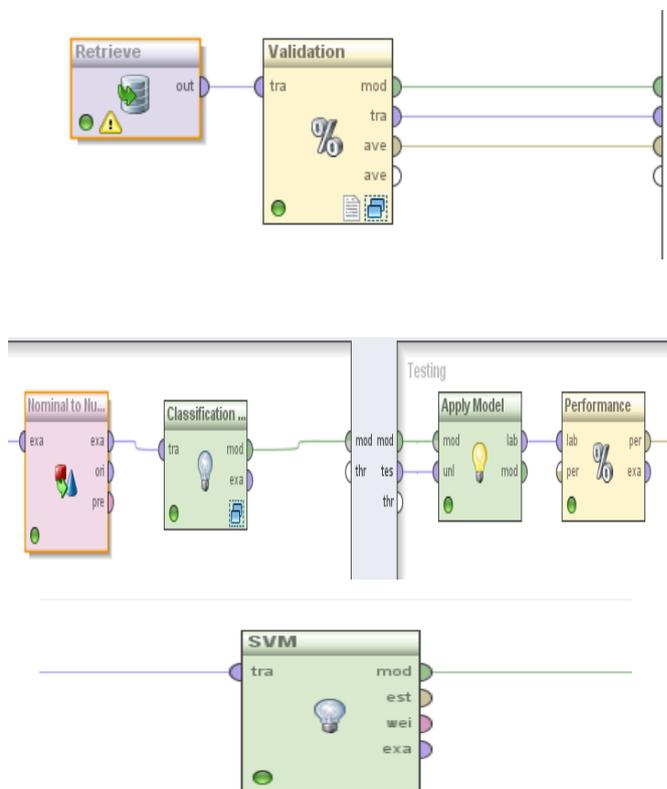
Hasil SVM	Hasil Algoritma C4.5
95.00%	97.25%

E. Evaluasi dan Validasi Hasil

Hasil dari pengujian model yang dilakukan untuk memprediksi penyakit Ginjal dengan support vector machine dan support vector machine berbasis particle swarm optimization untuk menentukan nilai accuracy dan AUC. Metode pengujiannya menggunakan cross validation dengan desain modelnya sebagai berikut:

1) Hasil Pengujian Model Support Vector Machine

Pada penentuan hasil penyakit diabetes menggunakan algoritma Support Vector Machine berbasis pada framework RapidMiner sebagai berikut:



Gbr.3 Model Pengujian Validasi Support Vector Machine

Nilai accuracy, precision, dan recall dari data training dapat dihitung dengan menggunakan RapidMiner. Hasil pengujian dengan menggunakan model Support Vector Machine pada tabel 3.

a) *Confusion Matrix*

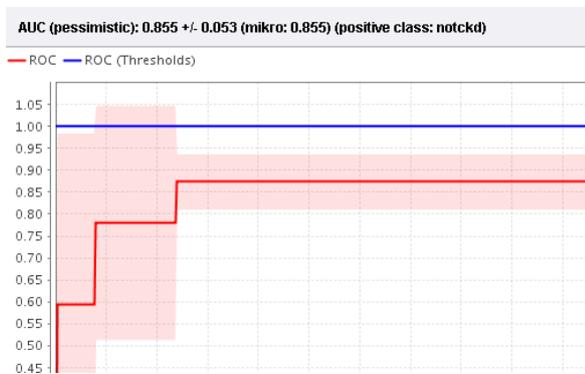
Berikut tabel confusion matrix untuk metode svm. Diketahui tingkat akurasi 95%. Dari data keseluruhan sebanyak 400 data, 246 data diprediksi sesuai dan 16 data diprediksi notckd tetapi ternyata masuk kedalam klasifikasi normal, dan sebanyak 4 data diprediksi normal tetapi masuk kedalam klasifikasi notckd dan sebanyak 134 data diprediksi sesuai yaitu ckd.

TABEL 3
MODEL CONFUSION MATRIX UNTUK METODE SVM

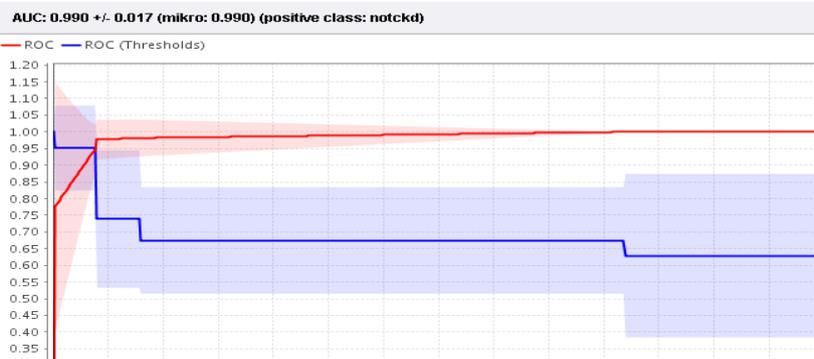
accuracy: 95.00% +/- 3.54% (mikro: 95.00%)			
	true ckd	true notckd	class precision
pred. ckd	246	16	93.89%
pred. notckd	4	134	97.10%
class recall	98.40%	89.33%	

b) *Kurva ROC*

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bisa dilihat pada Gambar 4 yang merupakan kurva ROC untuk algoritma Support Vector Machines. Kurva ROC pada gambar 4 mengekspresikan confusion matrix dari Tabel 3 Garis horizontal adalah false positives dan garis vertikal true positives.



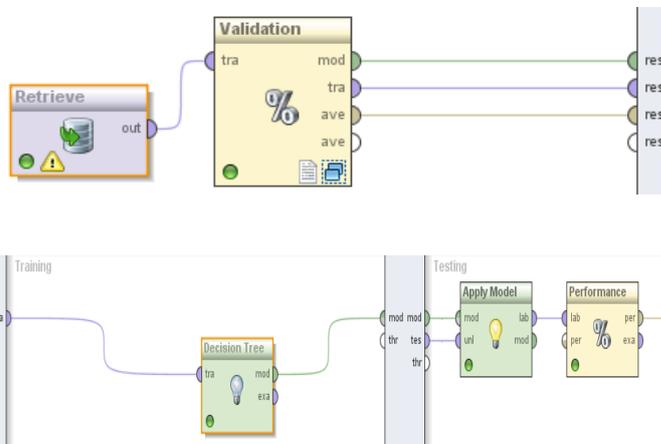
Gbr.4 Kurva ROC dengan Metode Support Vector Machines



Gbr.6 Kurva ROC dengan Metode Algoritma C4.5

2) Hasil Pengujian Algoritma C4.5

Pada penelitian penentuan hasil penyakit Ginjal menggunakan algoritma C4.5 pada framework Rapid Miner sebagai berikut:



Gbr.5 Model pengujian algoritma C4.5

Nilai accuracy, precision, dan recall dari data training dapat dihitung dengan menggunakan RapidMiner. Hasil pengujian dengan menggunakan algoritma C4.5 pada tabel 4...

a) *Confusion Matrix*

Berikut tabel confusion matrix untuk algoritma C4.5. Diketahui tingkat akurasi 97.25%. Dari data keseluruhan sebanyak 400 data, 242 data diprediksi sesuai dan 3 data diprediksi notckd tetapi ternyata masuk kedalam klasifikasi normal, dan sebanyak 8 data diprediksi normal tetapi masuk kedalam klasifikasi abnormal dan sebanyak 147 data diprediksi sesuai yaitu ckd.

TABEL 4
MODEL CONFUSION MATRIX UNTUK METODE ALGORITMA C4.5

accuracy: 97.25% +/- 2.36% (mikro: 97.25%)			
	true ckd	true notckd	class precision
pred. ckd	242	3	98.78%
pred. notckd	8	147	94.84%
class recall	96.80%	98.00%	

b) Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bisa dilihat pada Gambar 6 yang merupakan kurva ROC untuk algoritma C4.5. Kurva ROC pada gambar 6 mengekspresikan confusion matrix dari Tabel 4 Garis horizontal adalah false positives dan garis vertikal true positives.

F. Analisa Hasil Pengujian

Dari hasil pengujian diatas, baik evaluasi menggunakan confusion matrix maupun ROC curve terbukti bahwa hasil pengujian algoritma C4.5 memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma SVM. Nilai akurasi untuk model algoritma C4.5 sebesar 97.25% dan nilai akurasi untuk model algoritma SVM sebesar 95.00% dengan selisih akurasi 2.25%

TABEL.5
PERBANDINGAN SELURUH PENGUJIAN

	Pengujian SVM	Pengujian C4.5
Accuracy	95.00%	97.25%
Precision	96.90%	95.15%
Recall	87.33%	98.00%
AUC	0.855	0.990

Tabel tersebut menunjukkan bahwa, tingkat akurasi menggunakan algoritma C4.5 adalah sebesar 97.25%, dan dapat dihitung untuk mencari nilai accuracy, sensitivity, specificity, ppv, dan npv hasilnya pada persamaan dibawah ini:

TABEL.6
MODEL CONFUSION MATRIX UNTUK ALGORITMA C4.5

Accuracy .97.25% +/-2.36%(mikro:97.25%)			
	True ckd	True notckd	Class precicion
Pred.ckd	242	3	98.78%
Pred.notckd	8	147	94.84%
Class recall	96.80%	98.00%	

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{242 + 147}{242 + 147 + 8 + 3} = 0.9725$$

$$Sensitivity = \frac{TP}{TP + FN} = \frac{242}{242 + 3} = 0.9877$$

$$Specificity = \frac{TN}{TN + FP} = \frac{147}{242 + 8} = 0.588$$

$$PPV = \frac{TP}{TP + FP} = \frac{242}{242 + 8} = 0.968$$

$$NPV = \frac{TN}{TN + FN} = \frac{147}{147 + 3} = 0.98$$

V. KESIMPULAN

- Langkah-langkah untuk menggunakan algoritma C4.5 dalam memberikan klasifikasi penyakit Ginjal untuk kondisi ckd dan notckd sebagai berikut :
 - Mengumpulkan data-data penyakit tulang punggung yang meliputi atribut age, blood pressure, specific gravity, albumin,sugar ,red blood cells, pus cell,pus cell clumps,bacteria,blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, anemia
 - Membagi data-data yang dikumpulkan menjadi data training dan testing
 - Melakukan perhitungan terhadap data training dan menguji model didapatkan hasil dihasilkan akurasi 97.25 %
- Perbandingan dari hasil pengujian maka didapatkan: akurasi 95.00 % dengan menggunakan algoritma SVM. 97.25% menggunakan algoritma C4.5. Jadi dapat disimpulkan untuk pengujian menggunakan Algoritma C4.5 lebih tinggi tingkat akurasinya dibandingkan dengan SVM.

UCAPAN TERIMA KASIH

Terima kasih kepada kakak Ningsi, adik Susi, adik Yongki, Lasamilu Buton yang telah memberikan inspirasi,semangat dan Do'a bagi penulisan jurnal ilmiah ini.

REFERENSI

- Gorunescu, F. (2011). Data Mining Concepts, Models and Techniques. Berlin Heidelberg: Springer Verlag.
- Prof. DR. dr. M. Farid Aziz, SpOG (K), dr. Julianto Witjaksono, MOG, SpOG(K) Onk, dr. H. Imam Rasjidi, SpOG(K) Onk, Panduan pelayanan medik: model interdisiplin penatalaksanaan kanker serviks dengan gangguan ginjal, Jakarta : 2008
- Dr. S. Vijayarani, Mr.S.Dhayanand, "Data mining classification algorithms for kidney disease prediction," International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 4, August 2015
- Lakshmi. K.R, Nagesh. Y and VeeraKrishna. M, (2014) Performance Comparison Of Three Data Mining Techniques For Predicting Kidney Dialysis Survivability, International Journal of Advances in Engineering & Technology, Mar, Vol. 7, Issue 1 , pg no. 242-254
- Andrew Kusiak, Bradley Dixonb, Shital Shaha, (2005) Predicting survival time for kidney dialysis patients: a data mining approach, Elsevier Publication, Computers in Biology and Medicine 35. page no 311-327
- Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining Practical Machine Learning Tools And Techniques. Burlington, Usa: Morgan Kaufmann Publishers.
- Moertini, V. S. (2002). Data Mining Sebagai Solusi Bisnis. Integral, Vol. 7 No. 1, April 2002 , 4.
- Bellazzi, R., & Zupanb, B. (2008). Predictive Data Mining In Clinical Medicine: Current Issues And And Guidelines. International Journal Of Medical Informatics 7 7 , 81-97.
- Dong, Y., Xia, Z., Tu, M., & Xing, G. (2007). An Optimization Method For Selecting Parameters In Support Vector Machines. Sixth International Conference On Machine Learning And Applications , 1-6.

- [10] Burges, C. J. (1998). A Tutorial On Support Vector Machines For Pattern Recognition. Boston: Kluwer Academic Publishers
- [11] Maimon, O. (2010). Data Mining And Knowledge Discovery Handbook. New York Dordrecht Heidelberg London: Springer.
- [12] Huang, K., Yang, H., King, I., & Lyu, M. (2008). Machine Learning Modeling
- [13] Vercellis, C. (2009). BusinessIntelligent: Data Mining and Optimization for Decision Making. Southern Gate: John Willey & Sons Inc.
- [14] Han, J., & Kamber, M. (2007). Data Mining Concept and Techniques, San Fransisco: Mofgan Kaufan Publisher.
- [15] Liao. (2007). Recent Advances in Data Mining of Enterprise Data: Algorithms and Application. Singapore: World Scientific Publishing.
- [16] Riduwan. (2008). Metode dan Teknik Menyusun Tesis. Alfabeta. (Bandung).
- [17] Sugiyanto. (2008). Metode Penelitian Kuantitatid Kualitatif dan R&D. Bandung: Alfabeta.



Sutri Handayani. Padang Brahrang, 27 November 1988. Pendidikan Terakhir S2 Pasca Sarjana STMIK Nusa Mandiri Jakarta, saat ini mengajar sebagai Dosen di STMIK Nusa Mandiri Jakarta, Program S1 Program Studi Teknik Informatika.