Journal of Soft Computing Exploration

Meningkatkan Akurasi Algoritma C4.5 Menggunakan Information Gain Rasio dan Adaboost untuk Klasifikasi Ginjal Kronis Penyakit

Aprilia Lestari[1], Alamsyah[2]
[1,2]Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang, Indonesia

Info Artikel

Sejarah artikel:

Diterima 2 Agustus 2020
Direvisi 20 Agustus 2020
Diterima 12 September 2020

Kata kunci:

Penambangan Data;
C4.5;
Peningkatan Rasio Keuntungan;
CKD;

ABSTRAK

Data informasi yang telah tersedia sangat banyak dan akan membutuhkan waktu yang sangat lama untuk mengolah data informasi yang jumlahnya banyak. Oleh karena itu, data mining digunakan untuk memproses data dalam jumlah besar. Metode data mining dapat digunakan untuk mengklasifikasikan penyakit pasien, salah satunya penyakit ginjal kronis. Penelitian ini menggunakan metode klasifikasi pohon klasifikasi dengan algoritma C4.5. Pada proses preprocessing dilakukan seleksi fitur untuk mengurangi atribut yang tidak meningkatkan hasil akurasi klasifikasi. Pemilihan fitur menggunakan rasio gain. Metode Ensemble menggunakan adaboost, yang dikenal dengan istilah boosting. Dataset yang digunakan oleh Chronic Kidney Dataset (CKD) diperoleh dari repositori UCI mesin pembelajaran. Tujuan dari penelitian ini adalah menerapkan information gain ratio dan adaboost ensemble pada dataset penyakit ginjal kronis menggunakan algoritma C4.5 dan mengetahui hasil akurasi algoritma C4.5 berdasarkan information gain ratio dan adaboost ensemble. Hasil yang didapatkan untuk iterasi default pada adaboost yaitu sebanyak 50 iterasi. Akurasi C4.5 stand-alone diperoleh 96,66%. Akurasi untuk C4.5 dengan menggunakan information gain ratio diperoleh 97,5%, sedangkan untuk metode C4.5 dengan menggunakan information gain ratio dan adaboost diperoleh 98,33%. 5 standalone diperoleh 96,66%. Akurasi untuk C4.5 dengan menggunakan information gain ratio diperoleh 97,5%, sedangkan untuk metode C4.5 dengan menggunakan information gain ratio dan adaboost diperoleh 98,33%. 5 standalone diperoleh 96,66%. Akurasi untuk C4.5 dengan menggunakan information gain ratio diperoleh 97,5%, sedangkan untuk metode C4.5 dengan menggunakan information gain ratio dan adaboost diperoleh 98,33%.

Penulis yang sesuai:

Aprilia Lestari
Departemen Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang,
Email: aprilialestari11@gmail.com

1. PERKENALAN

Seiring dengan perkembangan teknologi yang semakin pesat, masyarakat dapat dengan mudah mengakses informasi kapanpun dan dimanapun. Data informasi yang telah tersedia sangat banyak dan akan membutuhkan waktu yang sangat lama untuk mengolah data informasi yang jumlahnya banyak. Oleh karena itu, untuk mengolah data dalam jumlah besar digunakan teknik data mining. Data mining dapat diterapkan di berbagai bidang. Salah satunya dalam bidang kesehatan adalah memprediksi dan mengklasifikasikan suatu penyakit dari data rekam medis pasien. Metode data mining dapat digunakan untuk mengklasifikasikan penyakit pasien berdasarkan tingkat keparahan penyakitnya, salah satunya dengan mengklasifikasikan pasien dengan penyakit ginjal dan tidak. Penyakit ginjal kronis atau gagal ginjal merupakan masalah yang sangat serius di seluruh penjuru dunia, dimana ginjal mengalami kerusakan dan menjadi penyebab tidak maksimalnya fungsi ginjal [1].

Data mining digunakan untuk menentukan pola dalam pengetahuan data mining dan berguna dalam memecahkan masalah data di gudang data yang besar [2]. Istilah Data mining juga disebut sebagai penemuan pengetahuan. Data mining memiliki berbagai macam jenis metode, untuk itu pemilihan metode yang tepat akan tergantung dari tujuan dan prosesnya. Salah satu metode dalam data mining adalah klasifikasi. Metode klasifikasi memiliki input berupa kumpulan record, dimana setiap record ditandai dengan tuple (xy). X adalah atribut dan Y adalah atribut/target tertentu yang menunjukkan label kelas. Klasifikasi memiliki beberapa algoritma termasuk Naïve Bayes dan C4.5, masing-masing

yang memiliki akurasi yang berbeda [3]. Beberapa teknik yang ada dalam klasifikasi, pohon keputusan merupakan teknik klasifikasi yang sangat populer dan banyak digunakan.

Pohon keputusan adalah pendekatan yang paling kuat dalam penemuan ilmiah dan penambangan data, dan alat yang sangat efektif di berbagai bidang seperti ekstraksi data dan teks, ekstraksi informasi, pembelajaran mesin, dan pengenalan pola [4]. Salah satu teknik pohon keputusan yang paling populer adalah C4.5. Algoritma C4.5 merupakan salah satu algoritma yang dikembangkan oleh J. Ross Quinlan yang merupakan pengembangan dari algoritma ID3 (Iterative Dichotomiser 3) [5].

Penelitian ini dilakukan dengan menggunakan Dataset Penyakit Ginjal Kronis yang diperoleh dari repositori UCI mesin pembelajaran. Berikut ini adalah beberapa penelitian yang relevan dengan CKD. Pada penelitian [6] membandingkan dua algoritma yaitu C4.5 standalone dan C4.5 dengan Pessimistic pruning yang diterapkan pada dataset Chronic Kidney Disease. Standalone C4.5 memiliki akurasi 95% dan C4.5 dengan pemangkasan pressimistic menghasilkan akurasi 96.5%. Berdasarkan penelitian [7] membahas tentang prediksi penderita ginjal kronik menggunakan algoritma decision tree dan nave bayes. Dataset yang digunakan dalam penelitian ini adalah dataset penyakit ginjal kronik. Hasil dari penelitian ini adalah pohon keputusan yang menghasilkan akurasi sebesar 91% dan Naïve Bayes yang menghasilkan akurasi sebesar 86%. Dalam penelitian [8] yang menyatakan bahwa C4. Algoritma 5 memiliki akurasi tertinggi ketika diterapkan pada dataset Chronic Kidney Disease dibandingkan dengan algoritma Expectation Maximization (EM) dan Artificial Neural Network (ANN). Algoritma C4.5 menghasilkan akurasi sebesar 96,75%, EM 70% dan ANN 75%.

Untuk meningkatkan akurasi algoritma C4.5, penelitian ini menggunakan metode preprocessing dan metode ensemble. Pada proses preprocessing dilakukan seleksi fitur untuk mereduksi atribut yang tidak meningkatkan hasil akurasi klasifikasi. Pemilihan fitur menggunakan rasio gain. Metode Ensemble menggunakan adaboost yang juga dikenal sebagai boosting.

Tujuan dari penelitian ini adalah menerapkan Information Gain Ratio dan Adabost ensemble pada dataset Chronic Kidney Disease menggunakan algoritma C4.5 dan mengetahui hasil akurasi dari algoritma C4.5 berdasarkan Information Gain Ratio dan ensemble Adaboost.

2. METODE

2.1 Pemilihan Fitur

Seleksi fitur adalah proses untuk memilih subset dari atribut asli, sehingga ruang fitur berkurang secara optimal sesuai dengan kriteria tertentu. Pemilihan fitur yang bertujuan untuk mengurangi jumlah fitur tertentu yang berfokus pada data yang relevan dan meningkatkan kualitas sehingga seleksi fitur mampu bekerja lebih baik daripada proses yang didorong oleh fitur yang dipilih [11].

2.2.1 Rasio Perolehan Informasi

Rasio perolehan informasi adalah rasio perolehan informasi dengan informasi intrinsik. Untuk mengurangi bias terhadap atribut multi nilai dengan mengambil jumlah dan ukuran cabang dalam perhitungan saat memilih atribut. Ini berguna sebagai pertimbangan probabilitas logaritmik untuk mengukur dampak dari jenis perhitungan ini dalam kumpulan data.

2.2 Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh Quinlan sebagai versi perbaikan dari ID3. Pada ID3, induksi pohon keputusan hanya dapat dilakukan pada fitur tipe kategorikal (nominal/ordinal), sedangkan tipe numerik (internal/rasio) tidak dapat digunakan. Peningkatan yang membedakan algoritma C4.5 dari ID3 adalah dapat menangani fitur dengan tipe numerik, memangkas pohon keputusan, dan menurunkan set aturan. Algoritma C4.5 juga menggunakan kriteria gain dalam menentukan fitur yang merupakan pemutus simpul pada pohon yang diinduksi [12].

Dalam algoritma C4.5, membangun pohon keputusan hal pertama yang harus dilakukan adalah memilih atribut sebagai akar. Kemudian cabang dibuat untuk setiap nilai di root. Langkah selanjutnya adalah membagi kasus di cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus di cabang memiliki kelas yang sama [13].

$$GainRatio(A) = \frac{Gain(A)}{SplitEntropy(A)} \quad (1)$$

Untuk menghitung gain, digunakan Persamaan 2 sebagai berikut [15].

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (2)$$

Keterangan:

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A |S_{Saya}| =

Nomor Kasus di Partisi i |S| =

Nomor Kasus

Sedangkan untuk menghitung nilai entropi dapat dilihat pada Persamaan 3.

$$\text{Entropy}(S) = \sum_{i=1}^{n} - p_i * \log_2 p_i \quad (3)$$

Keterangan:

S = Himpunan kasus

n = Nomor partisi dari S

P_{Saya} = Proporsi S_{Saya} ke S, yang 2 dihitung dengan menggunakan Persamaan 4.

$$\log(X) = \frac{\ln(X)}{\ln(2)} \quad (4)$$

Entropi digunakan untuk menentukan node mana yang akan menjadi pemecah data pelatihan berikutnya. Nilai entropi yang lebih tinggi akan meningkatkan potensi klasifikasi. Yang perlu diperhatikan adalah jika entropi untuk node adalah 0 berarti semua data vektor berada pada label kelas yang sama dan node tersebut menjadi daun yang berisi keputusan (label kelas). Yang juga perlu diperhatikan dalam perhitungan entropi adalah jika salah satu elemen w_i bernilai 0 maka entropi dipastikan 0 juga. Jika proporsi semua elemen w_i sama, maka dapat dipastikan entropi bernilai [16].

Untuk menghitung Split Entropy Persamaan 5 digunakan sebagai berikut.

$$\text{SplitEntropy}_A(S) = -\sum_{i+1}^{n} \frac{|Si|}{|S|} * \log_2 \frac{|Si|}{|S|} \quad (5)$$

Keterangan:

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A |S_{Saya}| =

Nomor Kasus di Partisi i |S| =

Nomor Kasus

2.3 Peningkatan Adaptif (Adaboost)

Adaboost adalah bagian algoritma boosting dari ensemble learning yang digunakan untuk meningkatkan performa klasifikasi [17]. Menurut penelitian yang dilakukan oleh Nurzahputra & Muslim [18] menyatakan bahwa adaboost adalah bagian dari pembelajaran mesin yang diperkenalkan oleh Freud dan Schapire (1995) yang digunakan untuk meningkatkan aturan prediksi yang akurat dengan menyatukan banyak peraturan yang tidak akurat dan lemah.

Adaboost dan variannya telah berhasil diterapkan di beberapa bidang karena landasan teorinya yang kuat, prediksi yang akurat, dan kesederhanaan yang luar biasa. Langkah-langkah dalam algoritma adaboost adalah sebagai berikut.

A. Masukan: Kumpulan sampel penelitian dengan label {(xi,yi), ..., (xn,yn)}, algoritma pembelajaran komponen, jumlah rotasi T.

B. Inisialisasi: Bobot sampel pelatihan 1 = 1 , untuk semua i=1, ..., N

C. Lakukan untuk t=1, ..., T

1) Gunakan algoritma pembelajaran komponen untuk melatih komponen klasifikasi, h_T, untuk berat sampel pelatihan.

2) Hitung kesalahan pelatihan pada $h_t = \varepsilon_t \sum_i^N = 1 w_i^t, y_i \neq h_t(x_i)$

1) Tentukan bobot untuk pengklasifikasi komponen $h_t = \propto_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$

3) Perbarui bobot konstanta sampel pelatihan. $w_i^{t+1} = \frac{w_i^t \exp\{-\alpha_t y_i h_t(x_i)\}}{c_t}, i = 1, \ldots, N\ C_t$ adalah normalisasi

4) Keluaran: $f(x) = sign(\sum_t^T = 1\ \propto_t h_t(x))$

3. HASIL DAN PEMBAHASAN

Pada penelitian ini dibuat sistem berbasis web dengan menggunakan bahasa pemrograman Python untuk mengetahui hasil penerapan information gain ratio dan Adaboost Ensemble pada algoritma C4.5 dalam diagnosis penyakit ginjal kronis. Untuk membuatnya diperlukan data-data terkait diagnosis penyakit ginjal kronik yang akan digunakan sebagai sistem pengujian. Penelitian ini menggunakan dataset penyakit ginjal kronis yang diperoleh dari repositori pembelajaran mesin UCI. Data ini terdiri dari 24 atribut dan 1 kelas.

Pada tahap pengolahan data pengolahan data dilakukan sebelum algoritma diterapkan atau biasa disebut pre-processing. Dataset penyakit ginjal kronik yang diperoleh berupa file dengan ekstensi . arff, perubahan ekstensi file menjadi .xlsx untuk pemrosesan data.

3.1 Tahap Pembentukan

Tahap pemformatan berikut adalah pemformatan standar dalam dataset yang digunakan dalam penelitian. Misalnya pada atribut Rbc (Sel Darah Merah) dengan mengubah label pada atribut Rbc menjadi 0 untuk negatif (abnormal) dan 1 untuk positif (normal).

3.2 Penanganan Tahap Nilai yang Hilang

Penanganan missing value merupakan bagian dari pre-processing yang bertujuan untuk mengoptimalkan hasil penambangan. Nilai yang hilang dalam kumpulan data biasanya ditandai dengan simbol "?" Seperti pada Tabel 1 yang merupakan contoh dataset Penyakit Ginjal Kronik yang memiliki missing value.

Tabel 1. Nilai yang hilang dalam dataset penyakit ginjal kronis

Age	Bp	Sg	Al	Su	Rbc	Pc	Pcc	Ba
68	70	1.015	3	1	?	normal	Present	notpresent
68	70	?	?	?	?	?	notpresent	notpresent
68	80	1.010	3	2	normal	abnormal	Present	present
40	80	1.015	3	0	?	normal	notpresent	notpresent
47	70	1.015	2	0	?	normal	notpresent	notpresent
47	80	?	?	?	?	?	notpresent	notpresent
60	100	1.025	0	3	?	normal	notpresent	notpresent

Untuk mengganti missing value pada dataset menggunakan model perhitungan mean (rata-rata) dengan Persamaan 6.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (6)$$

A. Nilai untuk menggantikan nilai yang hilang pada atribut Sg:

$$\bar{x}(Sg) = \frac{\sum_1^{359.145} x_{(Sg)}}{353} = \frac{359.145}{353} = 1{,}017$$

B. Nilai untuk menggantikan nilai yang hilang pada atribut Al:

$$\bar{x}(Al) = \frac{\sum_1^{360} x_{(Al)}}{354} = \frac{360}{354} = 1{,}016949$$

C. Nilai untuk menggantikan nilai yang hilang dalam atribut Su:

$$\bar{x}(Su) = \frac{\sum_1^{158} x_{(Su)}}{351} = \frac{158}{351} = 0{,}450142 = 0$$

Dataset penyakit ginjal kronik yang telah diganti disajikan pada Tabel 2.

Tabel 2. Dataset penyakit ginjal kronis

Age	Bp	Sg	Al	Su	Rbc	Pc	Pcc	Ba
68	70	1.015	3	1	0.804	1	1	0
68	70	1.017	1.017	0.450	0.804	0.772	0	0
68	80	1.010	3	2	1	0	1	1
40	80	1.015	3	0	0.804	1	0	0
47	70	1.015	2	0	0.804	1	0	0
47	80	1.015	1.017	0.450	0.804	0.772	0	0
60	100	1.025	0	3	0.804	1	0	0

Pada tahap class balancing dilakukan dengan menerapkan algoritma SMOTE. Algoritma SMOTE diterapkan untuk membuat data baru lebih seimbang. Dataset awal German Credit memiliki 1000 sampel dengan 700 kelas loyal (baik) dan 300 kelas churn (buruk). Oleh karena itu perlu dilakukan penyeimbangan kelas dengan membuat data baru pada kelas churn. Dataset baru dari algoritma SMOTE menghasilkan 300 data kelas churn, sehingga terdapat 1300 data sampel baru. Hal ini dilakukan agar data dapat diklasifikasikan secara optimal. Tahap pemilihan atribut dilakukan dengan memilih atribut pada data yang digunakan. Pada tahap pemilihan atribut ini dilakukan reduksi dimensi pada data guna optimasi atribut yang akan mempengaruhi akurasi dari algoritma Naive Bayes. Reduksi dimensi pada atribut dilakukan dengan menggunakan Algoritma Genetika. Penghapusan atribut dilakukan satu persatu dari atribut yang memiliki nilai fitness terkecil dan akan di mining. Proses pemilihan atribut dan mining akan berhenti ketika hasil akurasi telah melebihi batas minimal yang ditentukan.

Setelah melalui tahap pre-processing, data baru akan melalui proses klasifikasi menggunakan algoritma Naive Bayes. Dari hasil yang diperoleh, terdapat peningkatan akurasi pada algoritma Naive Bayes dan algoritma Naive Bayes dengan menerapkan algoritma SMOTE dan pemilihan atribut pada Algoritma Genetika.

3.3 Tahap Implementasi Seleksi Fitur

Tahapan penerapan seleksi fitur adalah langkah pra-pemrosesan dalam data mining untuk memilih fitur dari atribut aslinya. Pada penelitian ini penerapan seleksi fitur pada dataset penyakit ginjal kronik bertujuan untuk menyeleksi atribut yang sesuai dengan kriteria tertentu untuk meningkatkan kualitas sehingga diperoleh hasil yang optimal. Hasil information gain ratio untuk masing-masing atribut CKD ditunjukkan pada Tabel 3.

Tabel 3. Hasil information gain ratio pada CKD

No	Attribute	Ratio
1	Age	0.06478486467333311
2	Blood Pressure	0.07449165865390706
3	Specific Gravity	0.29573829341251656
4	Albumin	0.2819094414967096
5	Sugar	0.07694929823654695
6	Red Blood Cells	0.05628074701652497
7	Pus Cell	0.07096759937984776
8	Pus Cell Clumps	0.02118141334366186
9	Bacteria	0.007827381958380508
10	Blood Glucose	0.17109797531713267
11	Blood Urea	0.1817205676125957
12	Serum Creatinine	0.36754848060905365
13	Sodium	0.17127694052839892
14	Potassium	0.1803325148203001
15	Hemoglobin	0.40690962861172
16	Packed Cell Volume	0.4000671660349939
17	White Blood Cell Count	0.123256384075207
18	Red Blood Cell Count	0.34560330527582805
19	Hypertension	0.24363083544933395
20	Diabetes Mellitus	0.21875835029559898
21	Coronary Artery Disease	0.07253163527515327
22	Appetite	0.22867128432500583
23	Pedal Edema	0.08596246562471399

Setelah dilakukan tahap perhitungan pemilihan fitur dengan menggunakan metode information gain ratio maka didapatkan hasil dari atribut terpilih. Hasil seleksi fitur menggunakan metode information gain ratio ditunjukkan pada Tabel 4.

Tabel 4. Hasil seleksi fitur menggunakan metode information gain ratio

Bp	Sg	Al	Bg	Bu	Sc	Sod	Hemo	Pcv	Rbcc	Htn	Dm
80.0	1.02	1.0	121.0	36.0	1.2	135.0	15.4	44.0	5.2	1.0	1.0
50.0	1.02	4.0	99.0	18.0	0.8	135.0	11.3	38.0	5.2	0.0	0.0
80.0	1.01	2.0	423.0	53.0	1.8	135.0	9.6	31.0	5.2	0.0	1.0
70.0	1.005	4.0	117.0	56.0	3.8	111.0	11.2	32.0	3.9	1.0	0.0
80.0	1.01	2.0	106.0	26.0	1.4	135.0	11.6	35.0	4.6	0.0	0.0

3.3 Tahap Penambangan Data

3.3.1 Implementasi Algoritma C4.5

Pada tahap ini model yang digunakan adalah dengan menerapkan algoritma C4.5 pada CKD. Data baru yang siap diolah dilakukan dengan berbagi data latih sebagai model dan data uji untuk mengukur kemampuan model yang terbentuk. Dalam penelitian ini distribusi data menggunakan metode random sub sampling. Dimana data training : data testing = 70% : 30% dibagi secara acak. Penerapan algoritma stand-alone C4.5 diperoleh akurasi sebesar 96,66% disajikan pada Tabel 5.

Tabel 5. Akurasi C4.5 stand-alone

Algorithm	Accuracy
C4.5	96,66%

3.3.1 Implementasi Algoritma C4.5 dan Rasio Perolehan Informasi

Pada tahap ini atribut asli dari dataset penyakit ginjal kronik terdiri dari 24 atribut dan 1 kelas, setelah diterapkan metode information gain ratio dengan memilih atribut terpilih 12 atribut. Dalam penelitian ini pendistribusian data menggunakan metode splitter yang terdapat pada sklearn library, yaitu metode random sub sampling. Sistem pembagian data dilakukan dengan metode sub-sampling random, dimana data dibagi menjadi 70% dan 30% dan pengambilan data dilakukan secara acak pada setiap eksekusi. Penerapan Information Gain Ratio pada data preprocessing menghasilkan akurasi C4.5 sebesar 97,5%, hasilnya disajikan pada Tabel 6.

Tabel 6. Akurasi C4.5 dan rasio perolehan informasi

Algorithm	Accuracy
C4.5 algorithm and Information Gain Ratio	97,5%

3.3.1 Implementasi Algoritma C4.5 dan Rasio Keuntungan Informasi dan Adaboost

Hasil dari pohon keputusan tersebut akan diketahui nilai gain dari atribut-atribut yang membentuk dataset tersebut. Dari nilai gain tersebut, masing-masing atribut diinisialisasi sebagai bobot awal dalam perhitungan adaboost. Setelah bobot inisialisasi diketahui, selanjutnya ditentukan iterasi dalam adaboost. Iterasi default di adaboost adalah 50 iterasi. Akurasi algoritma C4.5 berdasarkan information gain ratio dan adaboost dengan menggunakan subrandom sampling sebagai splitter adalah 98,33%. Hasil akurasi yang diperoleh disajikan pada Tabel 7.

Tabel 7. Akurasi C4.5 menggunakan rasio perolehan informasi dan adaboost

Algorithm	Accuracy
C4.5 algorithm using Information Gain Ratio and Adaboost	98,33%.

Hasil akurasi ini jauh lebih baik daripada hanya menggunakan algoritma C4.5 atau C4.5 berdasarkan information gain ratio saja.

4. KESIMPULAN

Penerapan information gain ratio dan adaboost ensemble merupakan kombinasi dari dua metode yang berguna untuk meningkatkan akurasi pada algoritma C4.5. Atribut asli dari dataset penyakit ginjal kronik terdiri dari 24 atribut dan 1 kelas, setelah diterapkan metode information gain ratio dengan memilih atribut terpilih 12 atribut. Iterasi default di adaboost adalah 50 iterasi. Akurasi C4.5 standalone adalah 96,66%, untuk C4.5 dengan information gain ratio 97,5%, sedangkan metode C4.5 berdasarkan information gain ratio dan adaboost adalah 98,33%. Jadi, dapat disimpulkan bahwa penggabungan metode information gain ratio dan adaboost dapat meningkatkan akurasi klasifikasi.

REFERENSI

[1] Boukenze, B., Mousannif, H., & Haqiq, A. (2016). Performa Teknik Data Mining untuk Prediksi Dalam Pelayanan Kesehatan Studi Kasus: Penyakit Gagal Ginjal Kronis. Jurnal Internasional Sistem Manajemen Basis Data (IJDMS), 8(3).

[2] Shajahaan, SS, Shanthi, S., & ManoChitra, V. (2013). Aplikasi Teknik Data mining untuk Memodelkan Data Kanker Payudara. Jurnal Internasional Teknologi Berkembang dan Rekayasa Lanjutan, 3(11): 362-369.

[3] Pranatha, AA (2012). Analisis Perbandingan Lima Metode Klasifikasi pada Dataset Sensus Penduduk. Jurnal Sistem Informasi, 4(2): 127-134.

[4] Neeraj, B., Girja, S., Ritu, DB, & Manisha, M. (2013). Analisis Pohon Keputusan pada Algoritma J48 untuk Data mining. Jurnal Internasional Penelitian Lanjutan dalam Ilmu Komputer dan Rekayasa Perangkat Lunak (JARCSSE), 3(6): 1114-1119.

[5] Muzakir, A., & Wulandari, RA (2016). Model Data mining sebagai Prediksi Penyakit Kehamilan dengan Teknik Decision Tree. Jurnal Ilmiah Informatika, 3(1): 19-26.

[6] Muslim, MA, Rukmana, SH, Sugiharti, E., Prasetiyo, B., & Alimah, S. (2018). Optimasi Particle Swarm Optimization Berbasis Algoritma C4.5 untuk Diagnosis Kanker Payudara. Konferensi Internasional tentang Matematika, Sains dan Pendidikan, 983(1): 012-063.

[7] Padmanaban, K. A & Parthiban, G. (2016). Menerapkan Teknik Machine Learning untuk Memprediksi Risiko Penyakit Ginjal Kronis. Jurnal Sains dan Teknologi India, 4 (2): 1-5.

[8] S, T., Bai, M., & Majumdar, J. (2017). Analisis dan Prediksi Penyakit Ginjal Kronis Menggunakan Teknik Data Mining. Jurnal Internasional Penelitian Teknik dalam Ilmu Komputer dan Teknik (IJERCSE), 4(9): 25-32.

[9] Gola, J., Britz, D., Staudt, T., Musim Dingin, M., Schneider, AS, Ludovici, M., & Mucklich, F. (2018). Klasifikasi struktur mikro tingkat lanjut dengan metode Data mining. Ilmu Material Komputasi, 148: 324-335.

[10] Nurzahputra, A., Safitri, AR, & Muslim, MA (2017). Klasifikasi Pelanggan pada Customer Churn Prediction Menggunakan Decision Tree. Prosiding Seminar Nasional Matematika. Semarang: Universitas Negeri Semarang: 717-722.

[11] Rodriguez-Galiano, VF, Luque-Espinar, JA, Chica-Olmo, M., & Mendes, MP (2018). Pendekatan Seleksi Fitur untuk Pemodelan Prediktif Pencemaran Nitrat Air Tanah: Evaluasi Metode Filter, Embedded, dan Wrapper. Ilmu Lingkungan Total, 624 (2018): 661-672.

[12] Prasetyo, E. (2014). Data mining: Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi Offset.

[13] Kusrini, & Luthfi, ET (2009). Algoritma Data Mining. Yogyakarta: CV Andi Offset.

[14] Neeraj, B., Girja, S., Ritu, DB, & Manisha, M. (2013). Analisis Pohon Keputusan pada Algoritma J48 untuk Data mining. Jurnal Internasional Penelitian Lanjutan dalam Ilmu Komputer dan Rekayasa Perangkat Lunak (JARCSSE), 3(6): 1114-1119.

[15] Quinland, J. Ross. (1986). Pengenalan Pohon Keputusan. Pembelajaran mesin. 1(1): 81-106

[16] Han, J. (2012). Konsep dan Teknik Data Mining. San Fransisco: Morgan Kauffman.

[17] Listiana, E., & Muslim, MA (2017). Penerapan Adaboost Untuk Klasifikasi Support Vector Machine Guna Mengingkatkan Akurasi Pada Diagnosa Chronic Kidney Disease. Prosiding Seminar Nasional Teknologi dan Informatika, 875-881.

[18] Nurzahputra, A., & Muslim, MA (2017). Peningkatan Akurasi pada Algoritma C4.5 Menggunakan Adaboost untuk Meminimalkan Risiko Kredit. Prosiding Seminar Nasional Teknologi dan Informatika. Kudus: Universitas Muria Kudus: 243-247