KLASIFIKASI DIAGNOSA GAGAL GINJAL KRONIS MENGGUNAKAN ALGORITMA DECISION TREE ITERATIVE DICHOTOMISER 3 (ID3)

[1)]Muhammad Laili (1410651033), [2)]Hardian Oktavianto. [3)]Henny W.

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah

Jember Jl. Karimata No. 49 Jember Kode Pos 68121

Email: [1)]muhammadlaili806@gmail.com, [2)]ilham.saifudin@unmuhjember.ac.id.

ABSTRAK

Algoritma ID3 adalah algoritma data mining dengan mencari hasil klasifikasi menggunakan sebuah pohon keputusan pada penelitian ini meneliti data pasien yang terkena penyakit gagal ginjal kronis menggunakan sebuah pohon keputusan Tujuan dari penelitian ini adalah menerapkan algoritma id3 untuk mengetahui hasil klasifikasi data pasien yang terdampak penyakit gagal ginjal kronis pada RSUD Koesnadi Bondowoso Hasil dari penelitian ini adalah diperoleh nilai akurasi sebesar 80% dan nilai presisi sebesar 72,72% data diuji menggunakan beberapa fold dengan 5 skenario uji.

Kata kunci: Gagal ginjal kronis, Algoritma Decision Tree

I Pendahuluan

1. Latar Belakang

   Gagal ginjal kronik adalah gangguan fungsi ginjal yang bersifat progesif dan *irreversibel*. Gangguan fungsi ginjal ini terjadi ketika tubuh gagal untuk mempertahankan metabolisme dan keseimbangan cairan dan elektrolit sehingga mengakibatkan retensi urea dan sampah nitrogen lain dalam darah (Brunner & Suddarth, 2001). Pada gagal ginjal kronik terjadi penurunan laju filtrasi glomerulus kurang dari 60 ml/menit selama 3 bulan atau lebih yang *irreversibe*l dan didasari oleh banyak faktor (Arora, 2009).

   Oleh karena itu peneliti melakukan sebuah penelitian dengan mengimplementasi data dengan pengambilan data training 20 dan data testing 20 menggunakan data yang peneliti ambil dari RSUD Koesnadi Bondowoso,dengan kriteria masing-masing data yang berbeda meliputi variable,tipe data dan jumlah data,Hal ini bertujuan untuk menganalisa hasil akurasi dari kedua metode tersebut terhadap data yang diujikan.Berdasarkan pembahasan diatas peneliti mengambil judul "Klasifikasi Diagnosa Gagal Ginjal Kronis Menggunakan Algoritma *Decission Tree Itterative Dichotomiser 3* (Id3)**".**

2. Rumusan Masalah

   Berdasarkan latar belakang yang telah dijelaskan sebelumnya, permasalahan dalam tugas akhir ini adalah sebagai berikut :

   a. Bagaimana akurasi dan presisi dari penerapan ID3 untuk klasifikasi penyakit gagal ginjal kronis ?
   b. Bagaimana karakteristik model klasifikasi berdasarkan tree yang terbentuk ?

3. Tujuan

   Tujuan yang akan dicapai dari tugas akhir ini adalah:

   a. Mencari akurasi dan presisi dari penerapan ID3 untuk klasifikasi penyakit gagal ginjal kronis.
   b. Karakteristik yang di bentuk dari tree berupa hasil klasifikasi confusion maxtrik.

4. Manfaat

   Adapun manfaat dari tugas akhir ini adalah :

   Penelitian ini dapat memberikan informasi atau gambaran untuk peneliti selanjutnya sebagai bahan acuan dan referensi sehingga dapat dikembangkan dengan konsep metode dan materi yang lebih lanjut.

5. Batasan Masalah

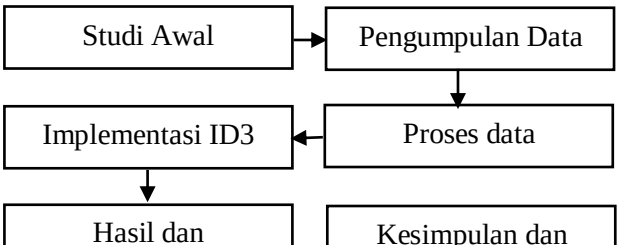
   Agar pembahasan ini dapat dilakukan secara terarah, maka perlu dilakukan batasan masalah meliputi :

   a. Tingkat akurasi dan algoritma yang dijadikan sebagai pengklasifikasian
   b. Sumber dataset dari RSUD Koesnadi Bondowoso, berupa data pasien diagnosa gagal ginjal kronis.
   c. Pengujian dilakukan dengan aplikasi RapidMiner
   d. Kriteria Gagal Ginjal Kronis yang digunakan meliputi : Tekanan darah, Hemoglobin, Faktor keturunan , Sesak napas, Usia.

II Metode Penelitian

1. Kerangka Penelitian

   Adapun metodologi penelitian yang dilakukan dalam pemanfaatan algoritma *Iterative Dichotomiser 3* (ID3) untuk mngidentifikasi pasien yang terdeteksi gagal ginjal kronis, memiliki langkah-langkah yang dapat digambarkan sebagai berikut:



Gambar 1. Tahapan Penelitian

2. Studi Awal
   Tahapan studi literatur yaitu mempelajari mengenai algoritma *Iterative Dichotomiser 3* (ID3) dan mencari informasi tentang penyakit Gagal ginjal kronis, serta memberi wawasan tentang penyakit gagal ginjal di masyarakat, baik lewat buku maupun paper dan lainnya, serta dataset sebagai bahan penelitian.

3. Pengumpulan Data
   Penelitian ini menggunakan dataset yang diambil dari RSUD Koesnadi Bondowoso,Dataset tersebut tentang pasien yang terkena penyakit gagal ginjal kronis,data berjumlah 250 pada dataset tersebut memiliki 5 atribut yaitu : Tekanan Darah, Hemoglobin, Factor Keturunan,Sesak Napas,Usia

4. Praproses Data
   Penelitian ini menngunakan data set yang di dapatkan secara langsung dari RSUD Koesnadi Bondowoso pada tahun 2018 dengan total jumlah data keseluruhan 250 pasien. Data ini mrncangkup 5 atribut yaitu: Tekanan Darah, Hemoglobin, Factor Keturunan,Sesak Napas,Usia yang akan di olah menggunakan mengenai algoritma *Iterative Dichotomiser 3* (ID3)

5. Pemanfaatan Algoritma Iterative Dichotomiser 3 (ID3)

   Dalam pemanfaatan algoritma ID3 peneliti menggunakan data training sebanyak 20 data sample.tahapan pertaman kita mencari Simpul Root atau titik awal dari Decision Tree kemudian tahap selanjutnya sitentukan simpul perantara setiap simpul perantara berhubungan dengan suatu pertanyaan atau pengujian, pohon yang terakhir adalah simpul *leaf* yaitu keputusan akhir atau kelas target untuk suatu pohon keputusan.

   Untuk menuntukan ketiga tahapan tersebut kita melakukan perhitungan terlebih dahulu dengan mencari nilai entropi dan gain pada data training yang tersedia untuk memudahkan perhitungan kita bisa menggunakan Microsoft excel dan membentuk table agar memudahkan kita menentukan entropi dan gain yang terbesar hingga yang terkecil.

III Implementasi

1. Implementasi Sistem
   Tahap implementasi sistem ini merupakan suatu tahapan menerapkan semua desain sistem yang telah dibuat kedalam tool sehingga dapat digunakan dengan mudah. Maka dilakukan implementasi sistem "Klasifikasi Diagnosa Gagal Ginjal Kronis Menggunakan Algoritma *Decission Tree Itterative Dichotomiser 3* (Id3)". Sistem ini menggunakan K-Fold Validation menggunakan tool Rapid Miner.

2. Gambaran datas
   Data penelitian yang digunakan yaitu data pasien diagnosa gagal ginjal kronis. *Data set* tersebut di ambil dari RSUD Koesnadi Bondowoso, pada data set tersebut memiliki 5 (Lima) atribut dan satu output yang terdiri dari dua kelas, yaitu "Terdiagnosa" dimana pasien terdiagnosa gagal ginjal kronis dan "Tidak" dimana pasien tidak terdiagnosa gagal ginjal. Berikut adalah gambaran atribut dalam dataset. Jumlah pasien diagnosa gagal ginja dalam data set berjumlah 250 orang. Untuk keterangan atribut di jelaskan pada tabel 4.1 di bawah ini.

NO	Nama Atribut	Parameter	Keterangan
1	Tekanan datah	Rendah Tinggi	Apakah tekanan darah rendah atau tinggi
2	Hemoglobin	Normal Kurang	Apakah hemoglobin normal atau kurang
3	Faktor keturunan	Ada Tidak	Apakah factor keturunan ada atau tidak.
4	Sesak napas	Ya Tidak	Apakah sesak napas ya atau tidak
5	Usia	Tua Dewasa	Apakah usia tua atau dewasa

Table 1 Keterangan Atribut

3. Praproses Data

Tahap ini akan mengubah dan menghapus parameter untuk menyesuaikan kebutuhan untuk klasifikasi yang biasa disebut normalisasi data. Dalam penelitian ini atribut yang mendapat preprocesing adalah atribut nama , karena nama tidak dipakai dalam klasifikasi.

4. Pemanfaatan algoritma ID3

Dalam pemanfaatan algoritma ID3 peneliti menggunakan *tool Rapid Miner* studio *version* 9.2 sebagai pengaplikasian. Sebelum data di masukan kedalam *Rapid Miner* data akan di simpan dalam format *excel*. Sebelumnya data akan dibagi menjadi 5 kategori data yaitu pembagian pengujian dengan data training dan testing yang berbeda-beda.

Table 2 skenario uji

P1	**P2**	**P3**	**P4**	**P5**
1-50	51-100	101-150	151-200	201-250
1-50	**51-100**	101-150	151-200	201-250
1-50	51-100	**101-150**	151-200	201-250
1-50	51-100	101-150	**151-200**	201-250
1-50	51-100	101-150	151-200	**201-250**

Dapat dilihat bahwa pada tabel diatas sudah terbagi menjadi beberapa bagian. Dalam tahap ini akan dilakaukan pengujian dengan data testing yang berbeda seperti P1 dan terletak pada kolom pertama dan pada P2 data testing terletak di kolom 2 dan begitu seterusnya hingga P5.

5. Hasil pengujian

Pada penelitian ini dilakukan scenario pengujian menggunakan cross fold validation dengan nilai k = 5 atau lima kali pengujian. Berikut hasil yang diperoleh.

a. Pengujian 1

Berdasarkan hasil pengujian P-1 dengan memasukan data uji dari 1 sampai 50 data Langkah selanjutnya adalah pengujian menggunakan *Confusion matrix* untuk mendapatkan hasil akurasi dan presisi.

		Klasifikasi	
		Positif	Negatif
Aktual	Positif	5	4
	Negatif	2	3

Gambar 2. Hasil confusion Matrix P1

Dari gambar 2 dapat dilihat bahwa dari 133 kasus P1 nilai akurasi algoritma id3 sebagai berikut:

$$Akurasi = \frac{5+3}{5+2+3+4} x100 = 57{,}14\%$$

$$Precison = \frac{5}{5+2} x100 = 71{,}42\%$$

b. Pengujian 2

Setelah dihitung dengan langkah dan cara sebelumnya seperti pada P-1 di dapatkan Confusion matrix untuk P-2 seperti pada gambar 2. Kemudian langkah selanjutnya menghitung nilai akurasi dan presisi untuk P-2 seperti pada P-1 sebelumnya.

		Klasifikasi	
		Positif	Negatif
Aktual	Positif	4	3
	Negatif	4	4

Gambar 3. Hasil confusion Matrix P2

Dari gambar 3 dapat dilihat bahwa dari 50 kasus P2 nilai akurasi algoritma id3 sebagai berikut:

$$Akurasi = \frac{4+4}{4+4+4+3} x100 = 53.33\%$$

$$Precison = \frac{4}{4+4} x100 = 50\%$$

c. Pengujain 3

Langkah selanjutnya adalah menghitung kembali untuk P-3 yaitu untuk mencari nilai akurasi dan presisi seperti pada cara P-1 dan P-2 sebelumnya.

		klasifikasi	
		Positif	Negatif
Aktual	Positif	6	2
	Negatif	2	5

Gambar 4. Hasil confusion Matrix P3

Dari gambar 4 dapat dilihat bahwa dari 50 kasus P3 nilai akurasi algoritma id3 sebagai berikut:

$$Akurasi = \frac{6+5}{6+2+5+2} x100 = 73{,}33\%$$

$$Precison = \frac{6}{6+2} x100 = 75\%$$

d. Pengujian 4

Setelah dihitung dengan langkah dan cara sebelumnya seperti pada P-1dan P-2 di dapatkan Confusion matrix untuk P-3 seperti pada gambar 4. Kemudian langkah selanjutnya menghitung nilai akurasi dan presisi untuk P-3 seperti pada P-1 dan P-2 sebelumnya.

		Klasifikasi	
		Positif	Negatif
Aktual	Positif	5	0
	Negatif	3	7

Gambar 5. Hasil confusion Matrix P4

Dari gambar 5 dapat dilihat bahwa dari 50 kasus P4 nilai akurasi algoritma id3 sebagai berikut:

$$Accurasi = \frac{5+7}{5+3+7+0} x100 = 80{,}00\%$$

$$Precison = \frac{8}{8+3} x100 = 72{,}72\%$$

e. Pengujian 5

Setelah dihitung dengan langkah dan cara sebelumnya seperti pada P-1,P-2 dan P-3 di dapatkan Confusion matrix untuk P-4 seperti pada gambar 4.8. Kemudian langkah selanjutnya menghitung nilai akurasi dan presisi untuk P-4 seperti sebelumnya.

		klasifikasi	
		Positif	Negatif
Aktual	Positif	5	5
	Negatif	2	2

Gambar 6. Hasil confusion Matrix P5

Dari gambar 6 dapat dilihat bahwa dari 50 kasus P4 nilai akurasi algoritma id3 sebagai berikut:

$$Accurasi = \frac{5+2}{5+2+5+2} x100 = 50{,}00\%$$

$$Precison = \frac{5}{5+2} x100 = 71{,}42\%$$

6. Hasil implementasi

Melalui perhitungan di atas di dapatkan hasil nilai rata-rata, minimum dan maximum untuk akurasi dan presisi P-1 sampai P-5 sebagai berikut:

1. Pengujian pertama untuk P-1 dengan memasukan data uji mulai nomor 1 sampai 50 mendapatkan akurasi sebesar 57,14% dan presisi 71,42%.
2. Pengujian kedua untuk P-2 dengan memasukan data uji mulai dari nomor 51 sampai 100 mendapatkan akurasi sebesar 53,33% dan presisi 50%.
3. Pengujian ketiga untuk P3 dengan memasukan data uji dari nomor 101 sampai 150 mendapatkan akurasi sebesar 73,33% dan presisi 75%.
4. Pengujian untuk P-4 dengan memasukan data uji dari nomor 151 sampai 200 mendapatkan akurasi sebesar 80,00% dan presisi 72,72%.
5. Pengujian untuk P-5 dengan memasukan data uji dari nomor 201 sampai 250 mendapatkan akurasi sebesar 50,00% dan presisi 71,42%.

Dari 5 kali pengujian menggunakan k fold cross valiation diperoleh hasil bahwa atribut "penurunan berat badan" selalu menjadi akar pada kelima pohon kuputusan, hal ini sejalan dengan hasil uji korelasi menggunakan correlation attribute evaluation dengan metode yang digunakan adalah ranker dan software bantu Weka. Berikut hasil pengukuran tingkat korelasi yang diperoleh.

Table 3 Hasil pengukuran korelasi atribut

rata-rata nilai korelasi	**Atribut**
0.5224	Faktor Keturunan
0.376	Hemoglobin
0.369	Tekanan Darah
0.223	Sesak Napas
0.1992	Usia

IV Hasil

1. Kesimpulan

Setelah melakukan analisis dan implementasi beserta pengujian dalam pemanfaatan algoritma ID3 untuk penyakit gagal ginjal kronis, maka didapatkan kesimpulan bahwa:

1. Hasil penerapan dengan menngunakan 5 uji data scenario memperoleh hasil seperti berikut :
   - P1 Akurasi 57,14% Presisi 71,42%
   - P2 Akurasi 53,335% Presisi 50%
   - P3 Akurasi73,33% Presisi 75%
   - **P4 Akurasi 80% Presisi 72,72%**
   - P5 Akurasi 50% Presisi 71%
2. Algoritma ID3 dapat menghasilkan tingkat akurasi dan presisi yang cukup tinggi dengan melakukan pengujian sebanyak 5 kali dengan data uji yang berbeda-beda, yaitu mendapatkan akurasi tertinggi sebesar 80%.
3. Dalam uji korelasi, atribut yang memiliki korelasi tertinngi sampai terendah atau atribut dengan pengaruh tertinggi sampai terendah yaitu:

rata-rata nilai korelasi	**Atribut**
0.5224	Faktor Keturunan
0.376	Hemoglobin
0.369	Tekanan Darah
0.223	Sesak Napas
0.1992	Usia

5.1 Saran

Berdasarkan penelitian yang telah di lakukan, diajukan saran penelitian sebagai berikut:

- Untuk pengembangan penelitian selanjutnya bisa dengan menambahkan metode optimasi pada data set sehngga memperoleh nilai akurasi lebih baik.

Daftar Pustaka

Adhiatma1 rief Tajally,Wahab Zulfachmi, '2014; Analisis Faktor-Faktor yang Berhubungan dengan Kejadian Gagal Ginjal Kronik Pada Pasien Hemodialisis Di RSUD Tugurejo Semarang

Arriawati, A. S. (2011) Klasifikasi merupakan proses untuk membedakan konsep atau kelas data .*Jurnal Informatika* vol. 8, No.1

Fahmia Nihaya Ika (2012) Hubungan Asupan Energi dan Protein dengan Status Gizi pada Penderita Gagal Ginjal Kronik yang Menjalani Hemodialisa Rawat Jalan di RSUD Tugurejo Semarang

Ipo Asri 2016, Hubungan Jenis Kelamin Dan FrekuensimHemodialisa Dengan Kualitas Hidup Pasien Gagal Ginjal Kronik Yang Menjalani Hemodialisa Di Rumah Sakit Umum Daerah Raden Mattaher Jambi

Juwita Lisavina 2018, Pengalaman Menjalani Hemodialisa Pada Pasien Gagal Ginjal Kronis

Kamil Insan, Rismia Agustina1 (2018) Gambaran Tingkat Kecemasan Pasien Gagal Ginjal Kronik Yang Menjalani Hemodialisis Di RSUD Ulin Banjarmasin

Mailani fitri (2017) Hubungan Dukungan Keluarga Dengan Kepatuhan Diet Pada Pasien Gagal Ginjal Kronik Yang Menjalani Hemodialisis

Mailani Fitri, Rika Fitri Andriani '2017' hubungan dukungan keluarga dengan kepatuhan diet pada pasien gagal ginjal kronik yang menjalani hemodialysis

Mariati Sulis (2013) Gambaran Makna Hidup Pasien Gagal Ginjal Kronik Yang Menjalani Hemodialisa

Obbie Kristanto. (2013). Penerapan Algoritma Klasifikasi Data Mining Id3 Untuk Menentukan Penjurusan Siswa SMAN 6. Fakultas Ilmu Komputer, 6–7. https://doi.org/10.1109/SICE.2007.4421147

Pangsibidang,G.S. (2016) Risiko Hipertensi, Diabetes, Dan Konsumsi Minuman Herbal Pada Kejadian Gagal Ginjal Kronik Di Rsup Dr Wahidin Sudirohusodo Makassar Tahun 2015

Prawitasari, Anita 2009. Classification and by Prediction DBD Tipes Algorithms Decision Tree. Universitas Gunadarma

Prof Dr. Soepomo 2014,Penerapan data mining prediksi penyakit ispa (Dengan *algoritma desicentree*)

Susanto, Sani, Ph.D, Suryadi, Dedy, 2010, Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data, Andi, Yogyakarta

Usman, H., & Akbar, P. S. (2003). Metodelogi Penelitian. Jakarta: PT Bumi Aksara