

# Diagnosis Dan Prediksi Penyakit Ginjal Kronis Dengan Menggunakan Pendekatan Stacked-Generalization

Aji Thofiq Madani\*, Hery Sunandar, Sumiaty Adelina Hutabara

Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia  
Email: <sup>1,\*</sup>ajithoufiqmadani@gmail.com, <sup>2</sup>herysun1975@gmail.com, <sup>3</sup>sumiatyadelina@mail.com  
Email Penulis Korespondensi: ajithoufiqmadani@gmail.com

**Abstrak**—Ginjal merupakan organ penting yang berfungsi menjaga komposisi darah dengan mencegah penumpukan limbah dan mengendalikan keseimbangan cairan dalam tubuh. Penyakit ginjal kronis (PGK) merupakan masalah kesehatan masyarakat global dengan insiden gagal ginjal yang meningkat, prognosis yang buruk dan biaya pengobatan yang tinggi. Prevalensi PGK seiring meningkatnya jumlah penduduk usia lanjut yang memiliki gejala penyakit diabetes serta hipertensi, sekitar 1 dari 10 populasi dunia mengalami penyakit ginjal kronis pada stadium tertentu. Hasil systematic review dan meta analysys yang dilakukan oleh Hill et al, 2016, mendapatkan prevelensi global PGK sebesar 13,4%. Menurut hasil Global Burden Of Disease tahun 2010, penyakit ginjal kronis menyebabkan kematian peringkat ke-27 di dunia tahun 1990 dan meningkat menjadi urutan ke-18 pada tahun 2010. Sedangkan di Indonesia merupakan perawatan paling serius ranking kedua pembiayaan terbesar dari BPJS kesehatan setelah penyakit jantung. Berdasarkan data badan kesehatan dunia atau World Health Organization (WHO) memperlihatkan jumlah penderita gagal ginjal baik akut maupun kronis mencapai 50 % sedangkan yang diketahui dan mendapat pengobatan hanya 25% dan 12,5% yang terobati dengan baik. Stacked generalization adalah penumpukkan beberapa algoritma untuk menentukan algoritma mana yang lebih efektif, oleh karna penulis menggunakan algoritma decesion tree. Algoritma ini didapat dari mengklasifikasikan tiga algoritma yaitu decision tree, multilayer perceptron, stochastic gradient descent menggunakan penambangan aplikasi weka. Dari klasifikasi algoritma tersebut algoritma decision tree lebih efektif dari algoritma lainnya. Decision tree merupakan suatu pendekatan yang sangat populer dan praktis dalam machine learning untuk menyelesaikan permasalahan klasifikasi. Data mining merupakan solusi yang mampu menemukan kandungan informasi yang tersembunyi berupa pola dan aturan dari sekumpulan data yang besar agar mudah dipahami[3]. Hasil penelitian yang diusulkan adalah berupa model prediksi penyakit ginjal

**Kata Kunci:** Diagnosa; Prediksi; Penyakit Ginja Kronis; Algoritma Decesion Tree

**Abstract**—Kidneys are important organs that function to maintain the composition of the blood by preventing the buildup of waste and controlling fluid balance in the body. Chronic kidney disease (CKD) is a global public health problem with an increasing incidence of kidney failure, poor prognosis and high treatment costs. The prevalence of CKD along with the increasing number of elderly people who have symptoms of diabetes and hypertension, about 1 in 10 of the world's population has chronic kidney disease at a certain stage. The results of a systematic review and meta-analysis conducted by Hill et al, 2016, found that the global prevalence of CKD was 13.4%. According to the results of the 2010 Global Burden of Disease, chronic kidney disease caused death ranked 27th in the world in 1990 and increased to 18th in 2010. Meanwhile, in Indonesia, it is the second most serious treatment, ranking the second largest health insurance provider after heart disease. . Based on data from the World Health Organization (WHO) shows the number of patients with acute and chronic kidney failure reaches 50%, while only 25% are known and received treatment and 12.5% are well treated. Stacked generalization is the stacking of several algorithms to determine which algorithm is more effective, because the author uses a decision tree algorithm. This algorithm is obtained from classifying three algorithms, namely decision tree, multilayer perceptron, stochastic gradient descent using weka application mining. From the classification algorithm, the decision tree algorithm is more effective than other algorithms, Decision tree is a very popular and practical approach in machine learning to solve classification problems. Data mining is a solution that is able to find hidden information content in the form of patterns and rules from large data sets so that they are easy to understand [3]. The results of the proposed research are in the form of a prediction model for kidney disease

**Keywords:** Diagnosis; Prediction; Chronic Kidney Disease; Decesion Tree Algorithm

## 1. PENDAHULUAN

Ginjal merupakan organ penting yang berfungsi menjaga komposisi darah dengan mencegah penumpukan limbah dan mengendalikan keseimbangan cairan dalam tubuh. Penyakit ginjal kronis (PGK) merupakan masalah kesehatan masyarakat global dengan insiden gagal ginjal yang meningkat, *prognosis* yang buruk dan biaya pengobatan yang tinggi. *Prevalensi* PGK seiring meningkatnya jumlah penduduk usia lanjut yang memiliki gejala penyakit diabetes serta hipertensi, sekitar 1 dari 10 populasi dunia mengalami penyakit ginjal kronis pada stadium tertentu[1].

Hasil *systematic review dan meta analysys* yang dilakukan oleh Hill et al, 2016, mendapatkan *prevelensi global* PGK sebesar 13,4%. Menurut hasil *Global Burden Of Disease* tahun 2010, penyakit ginjal kronis menyebabkan kematian peringkat ke-27 di dunia tahun 1990 dan meningkat menjadi urutan ke-18 pada tahun 2010. Sedangkan di Indonesia merupakan perawatan paling serius ranking kedua pembiayaan terbesar dari BPJS kesehatan setelah penyakit jantung[1]. Berdasarkan data badan kesehatan dunia atau *World Health Organization* (WHO) memperlihatkan jumlah penderita gagal ginjal baik akut maupun kronis mencapai 50 % sedangkan yang diketahui dan mendapat pengobatan hanya 25% dan 12,5% yang terobati dengan baik[2]. Langkah yang paling tepat untuk mendapatkan keputusan terhadap diagnosa dan prediksi penyakit ginjal kronis diperlukan klasifikasi terhadap kriteria-kriteria yang menunjukkan indikasi jumlah penderita penyakit ginjal kronis, menggunakan pendekatan *stacked generalization*.

*Stacked generalization* adalah penumpukkan beberapa algoritma untuk menentukan algoritma mana yang lebih efektif, oleh karna penulis menggunakan algoritma decesion tree. Algoritma ini didapat dari mengklasifikasikan tiga algoritma yaitu decision tree, multilayer perceptron, stochastic gradient descent menggunakan penambangan aplikasi

weka. Dari klasifikasi algoritma tersebut algoritma decision tree lebih efektif dari algoritma lainnya, Decision tree merupakan suatu pendekatan yang sangat populer dan praktis dalam machine learning untuk menyelesaikan permasalahan klasifikasi. Data mining merupakan solusi yang mampu menemukan kandungan informasi yang tersembunyi berupa pola dan aturan dari sekumpulan data yang besar agar mudah dipahami[3]. Perlunya penerapan konsep untuk mendiagnosa dan memprediksi penyakit ginjal kronis perlu penelitian menggunakan perbandingan penelitian yang dilakukan oleh penelitian terdahulu [4], [5].

Menurut Hilda Amalia pada penelitian yang telah dilakukan sebelumnya yang berjudul “Perbandingan Metode Data Mining SVM dan NN untuk klasifikasi Penyakit Ginjal Kronis” menyimpulkan dari hasil penelitian yaitu dengan metode Neural Network (NN) diperoleh hasil 93,36% dan metode Support Vector Machine (SVM) diperoleh hasil 95,16%[6]. Penelitian selanjutnya yang telah dilakukan oleh Winda Widya Ariestya dan kawan-kawan yang berjudul” Decesion Tree Learning Untuk Pnentuan Jalur Kelulusan Mahasiswa” menyimpulkan dari hasil penelitian yaitu pada pengukuran kinerja kriteria information gain, gain ratio dan gini index memiliki nilai accuracy antara 0,85-1,00[7].

Untuk itu metode decision tree sangat cocok digunakan untuk mendiagnosa dan memprediksi banyaknya penderita penyakit ginjal kronis. Metode decision tree lebih efektif untuk mengetahui jumlah penderita dalam memprediksi resiko terkena penyakit ginjal kronis atau tidak ginjal kronis.

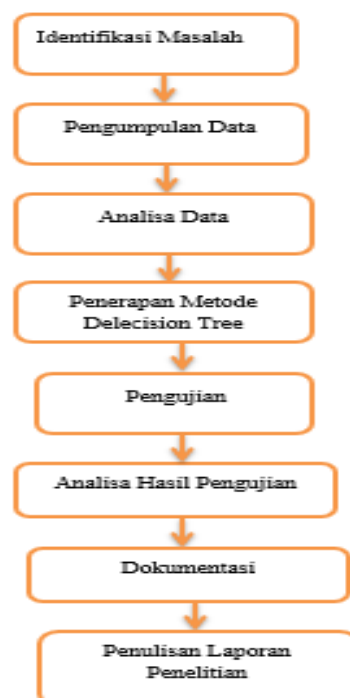
## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Jenis penelitian yang digunakan dalam penelitian ini merupakan penelitian *kuantitatif*. Jenis penelitian merupakan jenis penelitian yang spesifikasinya bersifat sistematis, terencana, dan terstruktur dengan jelas mulai dari awal hingga pembuatan desain penelitian, jenis penelitian *kuantitatif* dapat diartikan sebagai investigasi sistematis terhadap fenomena dengan mengumpulkan data yang dapat diukur melalui teknik statistik, matematika atau komputasi. Penelitian *kuantitatif* dalam penelitian ini dimaksudkan untuk mendapat gambaran dan keterangan-keterangan mengenai data yang digunakan dalam penelitian ini.

Metodologi yang digunakan pada penelitian ini adalah eksperimen. Penelitian eksperimen mencakup investigasi hubungan sebab – akibat menggunakan pengujian yang dikontrol sendiri. Cukup sering penelitian eksperimen ini mendapatkan kendala karena tidak cukupnya akses sampel, masalah etika dan lainnya, eksperimen biasanya dilakukan dalam pengembangan evaluasi dan pemecahan masalah.

Penelitian ini bertujuan untuk mendapatkan hasil dari memprediksi dan mendiagnosa penyakit ginjal kronis menggunakan pendekatan stacked generalization dengan melakukan implementasi klasifikasi dari beberapa algoritma data mining menggunakan alat penambangan weka 3.8, yaitu mengklasifikasikan algoritma decision tree, algoritma multilayer perceptron dan algoritma stochastic gradient descent. Dari tiga klasifikasi algoritma data mining tersebut dapat diambil kesimpulan penulis menggunakan algoritma decision tree. Karena penelitian yang diakui atau diterima harus mengikuti aturan yang ada, maka pada penelitian ini dilakukan menggunakan tahapan-tahapan seperti gambar di bawah ini:



Gambar 1. Tahapan Penelitian

## 2.2 Pendekatan Stacked Generalization Approach

*Generalisasi bertumpuk* adalah metode ansambel yang memungkinkan peneliti untuk menggabungkan beberapa algoritma prediksi yang berbeda menjadi satu dan mengklasifikasikan beberapa algoritma mana yang lebih efektif. Yaitu meliputi beberapa algoritma yang akan diklasifikasi antara lain *decision tree*, *multilayer perceptron* dan *Stochastic Gradient Descent (SGD)*. Sejak diperkenalkan pada awal 1990-an, metode ini telah berkembang beberapa kali menjadi sejumlah metode di antaranya adalah *Super Learner*. *Super Learner* menggunakan validasi silang untuk membangun kombinasi prediksi berbobot optimal dari perpustakaan algoritma kandidat. *Optimalitas* didefinisikan oleh fungsi tujuan yang ditentukan pengguna, seperti meminimalkan kesalahan kuadrat rata-rata atau memaksimalkan area di bawah kurva karakteristik operasi penerima. Meskipun sifatnya relatif sederhana, penggunaan *Super Learner* oleh ahli *epidemiologi* terhambat oleh keterbatasan dalam memahami detail *konseptual* dan teknis. Kami bekerja selangkah demi selangkah melalui dua contoh untuk mengilustrasikan konsep dan mengatasi masalah umum[8].

Pendekatan Generalisasi Bertumpuk untuk Diagnosis dan Prediksi, klasifikasi data dilakukan dengan tanda yang diperoleh pada Persamaan, tujuannya adalah untuk memilih *hyperplane* yang melakukan pemisahan titik sampel sebersih mungkin yang juga memaksimalkan pemisahan ke data *split* terdekat poin. Hal ini dicapai dengan pengenalan variabel *slack* yang positif dinilai, ke dalam kendala masalah *optimasi*

## 2.3 Metode Decision Tree (DT)

*Decision Tree* merupakan suatu pendekatan yang sangat populer dan praktis dalam *machine learning* untuk menyelesaikan permasalahan klasifikasi. Salah satu jenis dari *decision tree learning* adalah algoritma *Iterative Dichotomiser (ID3)*, pertumbuhan cabang pohon keputusan pada algoritma ID3 dilakukan sampai pohon tersebut mampu mengklasifikasikan sampel data secara akurat dengan tingkat kebenaran 100%. Adapun sampel data yang digunakan oleh ID3 memiliki beberapa syarat menurut [9]–[11], yaitu:

1. Deskripsi atribut nilai. Atribut yang sama harus mendeskripsikan tiap contoh dan memiliki jumlah nilai yang sudah ditentukan.
2. Kelas yang sudah didefinisikan sebelumnya. Suatu atribut contoh harus sudah didefinisikan, karena mereka tidak dipelajari oleh ID3.
3. Kelas-kelas yang diskrit. Kelas harus digambarkan dengan jelas, kelas yang dipecah-pecah menjadi kategori-kategori yang relatif, misalkan saja dikategori menjadi *hard, Flexible, soft*.
4. Jumlah contoh yang cukup. Karena pembangkitan induktif digunakan, maka dibutuhkan *test case* yang cukup untuk membedakan pola yang palid dari peluang suatu kejadian.

Algoritma pada metode ini menggunakan konsep dari *entropy* informasi. Dengan tujuan dengan mendefinisikan gain, pertama-tama digunakan ide dari teori informasi yang disebut *entropy*. *Entropy* mengukur jumlah dari informasi yang ada pada atribut dengan rumus:

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

Keterangan:

S : himpunan kasus

N : jumlah partisi S

Pi : proporsi dari Si terhadap S

Kemudian hitung nilai Gain :

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{s_i}{s} \cdot \text{Entropy} \quad (2)$$

Keterangan:

S : himpunan kasus

A : atribut

N : jumlah Partisi ke-i

[Si] : Jumlah Kasus Pada partisi ke-i

[S] : jumlah kasus dalam S

## 3. HASIL DAN PEMBAHASAN

### 3.1 Analisa Masalah

Analisa merupakan langkah awal dalam menyelesaikan dan mengidentifikasi sebuah permasalahan yang akan terjadi pada penentuan diagnosa dan prediksi penyakit ginjal kronis. Tahapan awal dalam penelitian ini adalah melakukan persiapan data yang didapat dari situs penyediaan data yaitu *UCI Repository*, data yang didapat merupakan data sekunder sebanyak 400 data yang terdiri dari 24 atribut dan satu label. Atribut-atribut yang digunakan adalah Umur, Tekanan Darah, *Specific Gravity*, *Albumin*, Gula Darah, Status Sel Darah Merah, Status *Pus Cell*, *Pus Cell Clumps*, Status Kehadiran Bakteri, Kadar Glukosa dalam Darah, *Blood Urea Nitrogen*, Kadar *Serum Creatinin*, Hasil *Sodium Blood Test*, Hasil *Pottasium Test*, Kadar *Hemoglobin*, *Packed Cell Volume*, Jumlah Sel Darah Putih, Jumlah Sel Darah Merah, Riwayat Hipertensi,

Riwayat Diabetes Mellitus, Riwayat Penyakit Jantung Koroner, Tingkat Nafsu Makan, Riwayat Pedal Idema, Riwayat Anemia dan satu tabel yaitu *Class*.

Data yang didapat tersebut disimpan kedalam bentuk tabel untuk mempermudah dalam proses pemasukan data, data yang didapat memiliki beberapa atribut yang memiliki nilai, dari proses normalisasi data didapat sebesar 400 data yang akan dilakukan dalam proses klasifikasi, dari proses klasifikasi menggunakan penambangan aplikasi weka dari *percentage split* 90% contoh data yang diklasifikasi dengan benar didapat sebesar 40 data. Data penderita penyakit ginjal kronis sebanyak 24 dan data tidak penderita ginjal kronis 16 data. Adapun data yang digunakan secara acak sebagai berikut:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe	ane	class
2	48	70	1.005	4	0	normal	abnormal	present	notpresent	117	56	3	111	2	11	32	6700	3	yes	no	no	poor	yes	yes	ckd
3	63	70	1.010	3	0	abnormal	abnormal	present	notpresent	380	60	2	131	4	10	32	4500	3	yes	yes	no	poor	yes	no	ckd
4	68	80	1.010	3	2	normal	abnormal	present	present	157	90	4	130	6	5	16	1100	2	yes	yes	yes	poor	yes	no	ckd
5	61	80	1.015	2	0	abnormal	abnormal	notpresent	notpresent	173	148	3	135	5	7	24	9200	3	yes	yes	yes	poor	yes	yes	ckd
6	48	80	1.025	4	0	normal	abnormal	notpresent	notpresent	95	163	7	136	3	9	32	6900	3	yes	no	no	good	no	yes	ckd
7	69	70	1.010	3	4	normal	abnormal	notpresent	notpresent	264	163	2	130	4	12	37	9600	4	yes	yes	yes	good	yes	no	ckd
8	73	70	1.005	0	0	normal	normal	notpresent	notpresent	70	163	0	125	4	10	29	18900	3	yes	yes	no	good	yes	no	ckd
9	73	80	1.020	2	0	abnormal	abnormal	notpresent	notpresent	253	163	4	138	5	10	33	7200	4	yes	yes	yes	good	no	no	ckd
10	56	90	1.015	2	0	abnormal	abnormal	notpresent	notpresent	129	163	6	131	4	9	29	6400	4	yes	no	no	good	no	no	ckd
11	48	80	1.005	4	0	abnormal	abnormal	notpresent	present	133	163	8	132	5	10	36	6200	4	no	yes	no	good	yes	no	ckd
12	59	70	1.010	3	0	normal	abnormal	notpresent	notpresent	76	163	15	135	7	7	22	3800	2	yes	no	no	poor	yes	yes	ckd
13	63	100	1.010	2	2	normal	normal	notpresent	present	280	163	3	143	3	13	40	9800	4	yes	no	yes	good	no	no	ckd
14	59	70	1.015	4	1	abnormal	normal	notpresent	notpresent	210	163	1	136	3	16	52	12500	5	no	no	no	good	no	no	ckd
15	71	70	1.010	3	0	normal	abnormal	present	present	219	163	3	133	4	10	33	5600	3	yes	yes	yes	good	no	no	ckd
16	73	100	1.010	3	2	abnormal	abnormal	present	notpresent	295	163	5	140	2	9	30	7000	3	yes	yes	yes	poor	no	no	ckd
17	71	60	1.015	4	0	normal	normal	notpresent	notpresent	118	163	5	136	4	11	35	15200	4	yes	yes	no	poor	yes	no	ckd
18	52	90	1.015	4	3	normal	abnormal	notpresent	notpresent	224	163	5	133	47	8	23	5000	2	yes	yes	no	good	no	yes	ckd
19	50	90	1.010	2	0	normal	abnormal	present	present	128	163	9	134	4	8	22	16300	2	no	no	no	poor	yes	yes	ckd
20	70	100	1.015	4	0	normal	normal	notpresent	notpresent	118	163	5	136	4	12	37	8400	8	yes	no	no	good	no	no	ckd
21	60	90	1.010	2	0	abnormal	normal	notpresent	notpresent	105	163	2	136	5	11	33	10500	4	no	no	no	good	no	no	ckd
22	60	60	1.010	3	1	normal	abnormal	present	notpresent	288	163	1	130	3	7	25	15200	3	yes	no	no	poor	no	yes	ckd
23	55	90	1.010	2	1	abnormal	abnormal	notpresent	notpresent	273	163	14	132	3	8	22	14600	2	yes	yes	no	poor	yes	yes	ckd
24	59	80	1.010	1	0	abnormal	normal	notpresent	notpresent	303	163	1	122	3	10	35	10900	4	no	yes	no	poor	no	no	ckd
25	83	70	1.020	3	0	normal	normal	notpresent	notpresent	102	163	2	115	5	8	26	12800	3	yes	no	no	poor	no	yes	ckd
26	23	80	1.025	0	0	normal	normal	notpresent	notpresent	70	163	1	150	4	17	52	9800	5	no	no	no	good	no	no	notckd
27	23	80	1.025	0	0	normal	normal	notpresent	notpresent	111	163	1	145	4	14	41	7200	5	no	no	no	good	no	no	notckd
28	23	80	1.020	0	0	normal	normal	notpresent	notpresent	99	163	1	142	4	17	46	4300	5	no	no	no	good	no	no	notckd
29	23	60	1.020	0	0	normal	normal	notpresent	notpresent	95	163	0	145	5	15	52	6300	4	no	no	no	good	no	no	notckd
30	24	70	1.025	0	0	normal	normal	notpresent	notpresent	140	163	0	140	4	16	48	5800	5	no	no	no	good	no	no	notckd
31	25	80	1.025	0	0	normal	normal	notpresent	notpresent	121	163	1	142	4	15	48	6900	5	no	no	no	good	no	no	notckd
32	25	70	1.020	0	0	normal	normal	notpresent	notpresent	88	163	0	136	4	13	48	7000	4	no	no	no	good	no	no	notckd
33	28	60	1.025	0	0	normal	normal	notpresent	notpresent	79	163	0	145	5	17	51	6500	5	no	no	no	good	no	no	notckd
34	29	80	1.020	0	0	normal	normal	notpresent	notpresent	83	163	0	139	3	17	40	9900	4	no	no	no	good	no	no	notckd
35	29	80	1.020	0	0	normal	normal	notpresent	notpresent	70	163	0	138	3	13	54	5400	5	no	no	no	good	no	no	notckd
36	30	80	1.020	0	0	normal	normal	notpresent	notpresent	131	16	1	147	3	14	45	9400	5	no	no	no	good	no	no	notckd
37	12	80	1.020	0	0	normal	normal	notpresent	notpresent	100	163	0	137	4	15	49	6600	5	no	no	no	good	no	no	notckd
38	15	80	1.025	0	0	normal	normal	notpresent	notpresent	93	163	0	136	3	16	50	6200	5	no	no	no	good	no	no	notckd
39	17	60	1.025	0	0	normal	normal	notpresent	notpresent	114	163	1	135	4	14	51	7200	5	no	no	no	good	no	no	notckd
40	20	70	1.020	0	0	normal	normal	notpresent	notpresent	123	163	1	135	3	14	44	5500	4	no	no	no	good	no	no	notckd
41	22	60	1.025	0	0	normal	normal	notpresent	notpresent	97	163	1	138	4	13	42	7900	6	no	no	no	good	no	no	notckd

**Gambar 2.** Dataset Penyakit Ginjal

### 3.1.1 Penerapan Metode Decision Tree

Pada tahapan ini dilakukan pengujian metode delection tree dengan penambangan aplikasi weka, yaitu melakukan klasifikasi setiap atribut yang digunakan untuk memprediksi orang dengan memasukan dataset yang sudah didownload oleh peneliti kemudian penulis melakukan pengujian dengan algoritma yang digunakan tersebut.

Pada tahapan ini dilakukan pengujian metode yang digunakan yaitu menghitung dan mendapatkan *rule-rule* yang ada pada algoritma *delection tree*. Akar pohon ditentukan dengan cara mencari nilai gain atribut yang paling tinggi menjadi akar pohon. Namun sebelum mencari nilai gain terlebih dahulu mencari nilai entropy.

Untuk yang pertama cari dahulu nilai entropy totalnya, yaitu:

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

$$= ((-24/40) * \log_2 (24/40)) + ((-16/40) * \log_2 (16/40)) = 0,970950594$$

Kemudian hitung nilai entropy dan gain dari masing-masing atribut, sebagai berikut;

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

1. Umur

$$\begin{aligned} \leq 30 \text{ th} &= ((-0/16)*\log_2(0/16) + (-16/16)*\log_2(16/16)) = 0 \\ >30 &= ((-24/24)*\log_2(24/24) + (-0/24)*\log_2(0/24)) = 0 \\ \text{Gain} &= 0,970950594 - (15/40)*0 - (1/40)*0 - (24/40)*0 = 0,970950594 \end{aligned}$$

2. Tekanan Darah

$$\begin{aligned} <60 &= ((-0/0)*\log_2(0/0) + (-0/0)*\log_2(0/0)) = 0 \\ 60 - 80 &= ((-16/32)*\log_2(16/32) + (-16/32)*\log_2(16/32)) = 1 \\ >80 &= ((-24/24)*\log_2(24/24) + (-0/24)*\log_2(0/24)) = 0 \\ \text{Gain} &= 0,970950594 - (0/40)*0 - (32/40)*1 - (24/40)*0 = 0,170950594 \end{aligned}$$

3. Specific Gravity

$$\begin{aligned} 1005 &= ((-5/5)*\log_2(5/5) + (-0/5)*\log_2(0/5)) = 0 \\ 1010 &= ((-11/11)*\log_2(11/11) + (-0/11)*\log_2(0/11)) = 0 \\ 1015 &= ((-6/6)*\log_2(6/6) + (-0/6)*\log_2(0/6)) = 0 \\ 1020 &= ((-2/2)*\log_2(2/2) + (-0/2)*\log_2(0/3)) = 0 \\ 1020 &= ((-0/16)*\log_2(0/16) + (-0/9)*\log_2(0/9)) = 0 \\ \text{Gain} &= 0,970950594 - (5/40)*0 - (11/40)*0 - (6/40)*0 \\ &= ((2/40)*0 - (16/40)*0) = 0,97095059 \end{aligned}$$

Dari perhitungan entropy di atas bahwa diketahui setibut yang mempunyai nilai nol tidak perlu dihitung namun untuk nilai gainnya maka harus tetap dihitung.

**Tabel 1.** Hasil Perhitungan Nilai Entropy Dan Gain

	Jumlah	Ckd	Notckd	Entropy	Gain
<b>Total</b>	40	24	16	0,97095059	
Umur( age)					0,97095059
	<=30 thn	16	0	16	0
	>30 thn	24	24	0	0
Tekanan Darah (bp)					0,17095059
	< 60 mm/hg	0	0	0	0
	60-80 mm/hg	32	16	16	1
	>80 mm/hg	8	8	0	0
Specific Gravity (sg)					0,97095059
	1.005	5	5	0	0
	1.010	11	11	0	0
	1.015	6	6	0	0
	1.020	2	2	0	0
	1.025	16	0	16	0
Albumin (al)					0,83377889
	0	17	1	16	0,32275696
	1	1	1	0	0
	2	7	7	0	0
	3	8	8	0	0
	4	7	7	0	0
	5	0	0	0	0
Gula Darah (su)					0,17095059
	0	32	16	16	1
	1	3	3	0	0
	2	3	3	0	0
	3	1	1	0	0
	4	1	1	0	0
	5	0	0	0	0
Status Sel Darah Merah (rbc)					0,22335687
	Normal	30	14	16	0,99679163
	Abnormal	10	10	0	0
Status Puss Cell (pc)					0,41997309
	Normal	24	8	16	0,91829583
	Abnormal	16	16	0	0
Pus Cell Clumps (pcc)					0,14649715
	Ada	7	7	0	0
	Tidak Ada	33	17	16	0,9993375

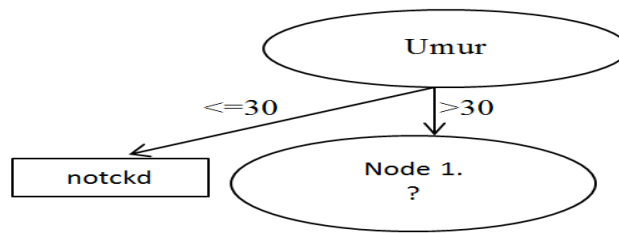
		Jumlah	Ckd	Notckd	Entropy	Gain
Bakteri (ba)						0,10059352
	Ada	5	5	0	0	
	Tidak Ada	35	19	16	0,9946938	
Kadar Glukosa Dalam Darah (bgr)						0,25155728
	<=200 mgs/dl	29	13	16	0,99226664	
	>200 mgs/dl	11	11	0	0	
Blood Urea Nitrogen (bu)						0,55522496
	<=50 mgs/dl	21	5	16	0,79185835	
	> 50 mgs/dl	19	19	0	0	
Kadar Serum Kreatinin (sc)						0,83377889
	<=1 mgs/dl	17	1	16	0,32275696	
	> 1 mgs/dl	23	23	0	0	
Hasil Sodium Blood Test (sod)						0,03810685
	<=20 mEq/l	2	2	0	0	
	< 120 mEq/l	38	22	16	0,98194079	
Hasil Potasium Test (pot)						0,17095059
	<=5 mEq/l	32	16	16	1	
	> 5 mEq/l	8	8	0	0	
Kadar Hemoglobin (hemo)						0,2812909
	< 10 gms	12	12	0	0	
	>10 gms	28	12	16	0,98522814	
Packed Cell Volume (pcv)						0,22335687
	<=30%	10	10	0	0	
	>30%	30	14	16	0,99679163	
Jumlah Sel Darah Putih (wc)						0
	<=500 cell/cum	0	0	0	0	
	>500 Cell/cum	40	24	16	0,97095059	
Jumlah Sel Darah Merah (rc)						0,41997309
	<=4milions/cmm	16	16	0	0	
	>4 milions/cmm	24	8	16	0,91829583	
Riwayat Hipertensi (htn)						0,55522496
	Ya	19	19	0	0	
	Tidak	21	5	16	0,79185835	
Riwayat Diabetes ( dm)						0,31274523
	Ya	13	13	0	0	
	Tidak	27	11	16	0,97511906	
Riwayat Penyakit Jantung Koroner (cad)						0,14649715
	Ya	7	7	0	0	
	Tidak	33	17	16	0,9993375	
Tingkat Nafsu Makan (appet)						0,2812909
	Baik	28	12	16	0,98522814	
	Kekurangan	12	12	0	0	
Riwayat Pedal Endema (pe)						0,25155728
	Ya	11	11	0	0	
	Tidak	29	13	16	0,99226664	
Riwayat Anemia (ane)						0,19653243
	Ya	9	9	0	0	
	Tidak	31	15	16	0,99924925	

Berdasarkan hasil perhitungan entropy dan gain diatas dapat dilihat bahwa atribut umur memiliki nilai tertinggi yaitu sebesar 0,970950594 maka atribut umur memiliki dua cabang yaitu umur <= 30tahun dan >30 tahun yang akan menjadi root atau pohon keputusan. Berikut adalah tabel dari atribut umur;

**Tabel 2.** Tabel Atribut Umur

Umur (age)	Jumlah	Ckd	Notckd
<=30 tahun	16	0	16
>30 tahun	24	24	0

Dari data diatas diketahui bahwa partisi  $\leq 30$  tahun memiliki nilai 0 pada kelas “ckd” dan partisi  $< 30$  tahun memiliki nilai 16 pada kelas “notckd” sedangkan pada partisi  $> 30$  tahun memiliki nilai 24 pada kelas “ckd” dan partisi  $> 30$  tahun memiliki nilai 0 pada kelas “notckd” maka dibuatlah pohon keputusan pada node 1 sebagai berikut;



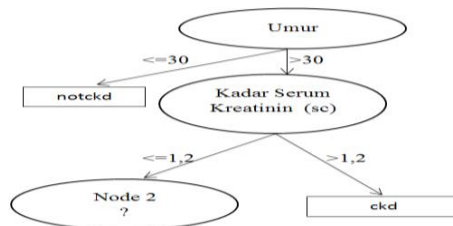
**Gambar 3.** Decision Tree Umur

Kemudian untuk mencari kelas dari cabang selanjutnya dengan menghitung entropy dan gain yang akan menjadi node pertama. Dari perhitungan entropy dan gain didapatkan nilai gain tertinggi yaitu atribut Kadar Serum Kreatinin (sc), berikut;

**Tabel 3.** Kadar serum Kreatinin (SC)

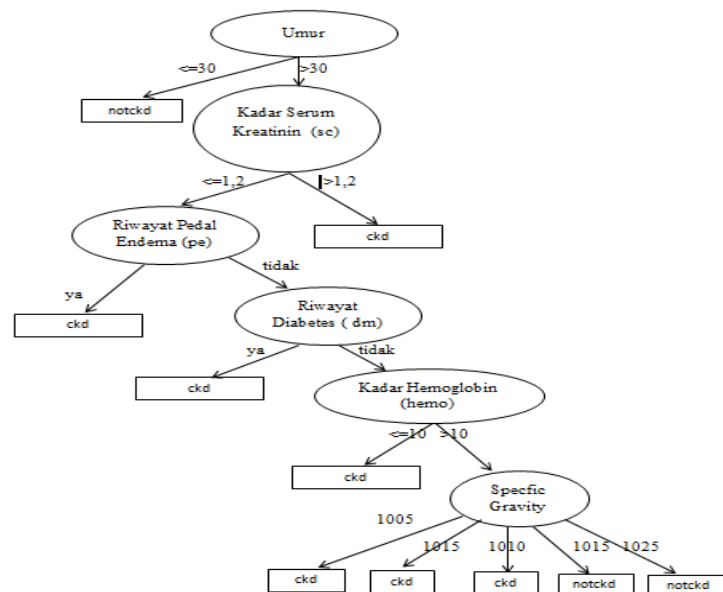
Kadar Serum Kreatinin (sc)	Jumlah	Ckd	Notckd
$\leq 1$ mgs/dl	17	1	16
$> 1$ mgs/dl	23	23	0

Dari perhitungan data diatas diketahui partisi atribut kadar serum kreatinin memiliki nilai 1 pada partisi  $\leq 1$  mgs/dl pada kelas “ckd” dan partisi  $\leq 1$  mgs/dl memiliki nilai 16 pada kelas “notckd” sedangkan pada partisi  $> 1$  mgs/dl pada kelas ckd memiliki nilai 23 dan partisi  $> 1$  mgs/dl memiliki nilai 0 pada kelas “notckd” maka dibuatlah pohon keputusan node 1 sebagai berikut;



**Gambar 4.** Decision Tree Node 1

Setelah dilakukan perhitungan keseluruhan, maka diperoleh pohon keputusan sebagai berikut;



**Gambar 5.** Pohon Keputusan (Decision Tree)

Dengan terbentuknya pohon keputusan (decision tree) yang terdapat pada gambar diatas dapat dibaca dengan rule sebagai berikut;

Jika umur = > 30 tahun

Dan kadar serum kreatinin = ya  
Dan riwayat pedal endema = ya  
Maka = ckd

Jika umur = > 30 tahun

Dan kadar serum kreatinin = tidak  
Maka = notckd

Jika umur =>30 tahun

Dan kadar serum kreatinin = ya  
Dan riwayat pedal endema = tidak  
Dan riwayat diabetes = ya  
Maka = ckd

Jika umur = > 30 tahun

Dan kadar serum kreatinin = ya  
Dan riwayat pedal endema = tidak  
Dan riwayat diabetes = tidak  
Dan kadar hemoglobin = <= 10  
Maka = ckd

Jika umur = >30 tahun

Dan kadar serum kreatinin = ya  
Dan riwayat pedal endema = tidak  
Dan riwayat diabetes = tidak  
Dan kadar hemoglobin = > 10  
Dan specific gravity = 1005  
Maka = ckd

Jika umur = >30 tahun

Dan kadar serum kreatinin = ya  
Dan riwayat pedal endema = tidak  
Dan riwayat diabetes = tidak  
Dan kadar hemoglobin = > 10  
Dan specific gravity = 1025  
Maka = notckd

Jika umur = < 30 tahun

Maka = not ckd

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dengan menggunakan pendekatan stacked generalization melalui metode algoritma decision tree dimana menghasilkan suatu pohon keputusan dapat diambil kesimpulan yaitu Banyaknya jumlah orang yang berumur di atas 30 tahun memiliki resiko penyakit ginjal kronis dan banyak dipengaruhi oleh faktor kadar serum kreatinin, riwayat pedal endema, riwayat diabetes, kadar hemoglobin dan spesfic gravity. Tingkat akurasi dengan algoritma decision tree menggunakan penambahan aplikasi weka memiliki persentasi yang akurat.

#### REFERENCES

- [1] Kemenkes RI, "Info datin ginjal," *Situasi Penyakit Ginjal Kron.*, pp. 1–10, 2017.
- [2] E. V. Hutagaol, "In light of another's word: European ethnography in the middle ages," *Jumantik*, vol. 2, no. 1, pp. 42–59, 2014.
- [3] J. L. Handarko, "Implementasi Fuzzy Decision Tree untuk Mendiagnosa Penyakit Hepatitis," vol. 4, no. 2, p. 158, 2015.
- [4] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*, 1st ed. Yogyakarta: Deepublish, 2020.
- [5] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1 : Kombinasi Metode Klastering dan Klasifikasi ( Kasus Pandemi Covid-19 di Indonesia )," *J. Media Inform. Budidarma*, vol. 4, no. 3, pp. 855–862, 2020.
- [6] H. Amalia, "Perbandingan Metode Data Mining Svm Dan Nn Untuk Klasifikasi Penyakit Ginjal Kronis," *J. Pilar Nusa Mandiri*, vol. 14, no. 1, pp. 1–6, 2018.
- [7] W. W. Ariestya, Y. E. Praptiningsih, and W. Supriatin, "Decision Tree Learning Untuk Penentuan Jalur Kelulusan Mahasiswa," *J. Ilm. FIFO*, vol. 8, no. 1, p. 97, 2016.
- [8] N. Singh and P. Singh, "A Stacked Generalization Approach for Diagnosis and Prediction of Type 2 Diabetes Mellitus," *Adv. Intell. Syst. Comput.*, vol. 990, pp. 559–570, 2020.
- [9] W. Fitriani and A. P. U. Siahaan, "Comparison Between WEKA and Salford System in Data Mining Software," *Int. J. Mob. Comput. Appl.*, vol. 3, no. 4, pp. 1–4, 2016.
- [10] Asmira, "Penerapan Data Mining untuk Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4,5 pada Bank BRI Unit Andounohu Kendari," *J. Sist. Komput. dan Sist. Inf.*, vol. 1, no. 1, pp. 22–28, 2019.
- [11] I. P. Sari and R. Harman, "Decission Tree Technique Dalam Menentukan Penjurusan Siswa Menengah Kejuruan," *J. Inf. Syst.*

*Res.*, vol. 1, no. 4, pp. 296–304, 2020.