

BAYESIAN NETWORKS UNTUK PREDIKSI PENYAKIT GINJAL KRONIS

Idza Runianti Darusman¹, Ednawati Rainarli²

^{1,2} Jurusan Teknik Informatika, Universitas Komputer Indonesia

Jl. Dipati Ukur No. 112 – 116, Coblong, Bandung

E-mail : idzarunianti@gmail.com¹, ednawati.rainarli@email.unikom.ac.id²

ABSTRAK

Pada penelitian sebelumnya tentang prediksi penyakit ginjal kronis, Lambodar J. dkk telah menggunakan metode *naive bayes* untuk memprediksi penyakit ginjal kronis. Dari penelitian ini didapatkan hasil akurasi 95%. Setelah dilakukan wawancara dengan Dr. Eva M Hidayat didapatkan hasil bahwa ada beberapa variabel pada data penyakit ginjal kronis yang memiliki relasi dengan variabel lain. Hal ini sangat berlawanan dengan konsep independen yang menjadi syarat dalam penggunaan metode *naive bayes*. Untuk mengatasi masalah tersebut digunakan metode *bayesian networks*. Dalam penelitian ini struktur *bayesian networks* dibuat sesuai hasil wawancara dengan Dr. Eva M Hidayat dan studi pustaka. Estimasi parameter pada *bayesian networks* ini menggunakan *maximum likelihood estimation* (MLE). Untuk validasi menggunakan metode *5-fold cross validation*. Setelah dilakukan pengujian, didapatkan bahwa variabel *blood urea*, *serum creatine*, dan *pedal edema* memiliki pengaruh paling besar dalam prediksi penyakit ginjal kronis. Selain itu didapatkan nilai rata – rata akurasi 98% pada kondisi class pada dataset sudah seimbang. Hal ini membuktikan bahwa *bayesian networks* memiliki hasil prediksi yang baik dibandingkan dengan *naive bayes* yang memiliki akurasi 97%.

Kata Kunci : prediksi, penyakit ginjal kronis, *bayesian networks*, *naive bayes*, *maximum likelihood estimation*.

1. PENDAHULUAN

Ginjal merupakan organ yang berfungsi untuk menjaga komposisi darah dengan mencegah menumpuknya limbah dan menyeimbangkan cairan yang ada pada tubuh [1]. Menjaga kesehatan ginjal sangat penting karena ginjal merupakan salah satu organ penting dalam menjaga kesehatan tubuh. Gangguan pada ginjal ada banyak sekali salah satunya adalah ginjal kronis.

Penyakit ginjal kronis (PGK) merupakan kerusakan pada ginjal yang menyebabkan ginjal tidak dapat membuang racun dan produk sisa darah, dengan ditandai adanya protein dalam urin serta penurunan laju filtrasi glomerulus yang berlangsung

selama lebih dari 3 bulan [1] [2]. Menurut Guru Besar Penyakit Dalam Fakultas Kedokteran Universitas Indonesia (FKUI), Parlindungan Siregar dikutip dari Kompas.com penyakit ini diketahui orang umumnya ketika sudah mencapai stadium tiga sampai empat [3]. Keterlambatan penanganan serta pendeteksian penyakit ini mengakibatkan prevalensi kematian akibat penyakit ginjal kronis (PGK) ini tinggi sehingga dibutuhkan pendeteksian dini untuk mengurangi tingkat prevalensi kematian akibat PGK.

Pada penelitian sebelumnya, Lambodar Jena dan Narendra Ku. Kamila telah melakukan penelitian untuk mendeteksi penyakit ginjal kronis salah satunya dengan menggunakan *naive bayes*. Pada penelitian tersebut didapatkan bahwa tingkat akurasi sebesar 95% serta waktu yang dibutuhkan untuk membuat model serta pengujiannya selama 0.02 detik [4]. Namun setelah dilakukan wawancara dengan Dr. Eva M Hidayat, didapatkan bahwa ada beberapa variabel yang memiliki relasi dengan variabel yang lain sehingga penelitian dengan menggunakan *naive bayes* masih dapat ditingkatkan.

Naive bayes dikatakan ‘*naive*’ karena diasumsikan bahwa setiap fitur kondisional independen atau dapat dikatakan bahwa dalam penyelesaian kasusnya setiap fitur tidak memiliki hubungan [5], sehingga *naive bayes* tidak dapat menyelesaikan kasus – kasus tertentu khususnya pada kasus dimana setiap fiturnya membutuhkan korelasi satu sama lain. Adapun metode yang dapat digunakan untuk menyelesaikan masalah hubungan antar fitur ini, salah satunya dengan menggunakan *bayesian networks*.

Pada penelitian yang dilakukan oleh P.Fuster-Parra dkk tentang resiko penyakit jantung dengan menggunakan *bayesian networks* [6]. Dari penelitian tersebut didapatkan bahwa hasil akurasi yang didapatkan adalah 96%.

Oleh karena itu dalam penelitian ini akan dilakukan prediksi terhadap penyakit ginjal kronis dengan menggunakan *bayesian networks* dilihat dari fitur (variabel prediktor) yang sudah tersedia.

2. LANDASAN TEORI

2.1 Penyakit Ginjal Kronis

Penyakit ginjal kronis (PGK) merupakan penyakit ginjal dimana ginjal memiliki Glomerular

filtration rate (GFR) dibawah 60 mL/min/1.73m² selama lebih dari 3 bulan atau bukti terjadinya kerusakan ginjal selama lebih dari 3 bulan [2].

Pada ginjal kronis ini terdapat beberapa gejala – gejala awal yang akan dirasakan oleh penderita, seperti : tekanan darah tinggi, perubahan pada *urine*, adanya darah pada *urine*, kehilangan nafsu makan, mudah lelah, mual dan muntah, susah tidur, sakit kepala [2]. Adapun faktor terjadinya penyakit ginjal kronis, yaitu [2] :

1. Hipertensi
2. Diabetes Melitus
3. Riwayat keluarga yang memiliki penyakit PGK
4. Obesitas
5. Umur lebih dari 60 tahun

Ketika seseorang telah diduga terkena penyakit ginjal kronis, maka pasien harus melalui beberapa pemeriksaan fungsi ginjal seperti [2] :

1. Tes untuk albumin (tipe protein) dan/atau darah pada urin
2. Tes darah untuk melihat tingkat produk sisa dalam darah dan menghitung laju filtrasi glomerulus.
3. Tes tekanan darah jika penyakit ginjal ini disebabkan oleh tekanan darah tinggi dimana dapat merusak pembuluh darah pada ginjal.
4. *CT scan* untuk mengambil gambar ginjal dan jejak urin. Tes ini untuk melihat ukuran ginjal serta letak tumor atau batu ginjal.

Berdasarkan laju filtrasi glomerulusnya, penyakit ginjal kronis ini dibagi menjadi beberapa tingkatan [2] yang dapat dilihat pada Tabel 1

Tabel 1. Tingkatan Penyakit Ginjal Kronis

	Tingkat	Kondisi Ginjal
<i>Early Stages</i>	1	eGFR normal \geq 90mL/min/1,73m ²
	2	Penurunan eGFR sedikit antara 60-89mL/min/1,73m ²
<i>Middle Stages</i>	3a	Penurunan eGFR ringan-sedang antara 45-59mL/min/1,73m ²
	3b	Penurunan eGFR sedang-berat antara 30-44mL/min/1,73m ²
	4	Penurunan eGFR berat antara 15-29 mL/min/1,73m ²
<i>End Stages</i>	5	Gagal ginjal, penurunan eGFR < 15m L/min/1,73m ²

2.2 Probabilitas

Probabilitas merupakan kemungkinan kemunculan suatu peristiwa [7]. Teori probabilitas ada hubungannya dengan hasil yang berbeda [8]. Ketika eksperimen sudah terdefinisi dengan baik, maka kumpulan semua hasil disebut dengan ruang sampel [8].

Secara matematis, ruang sampel merupakan himpunan sedangkan hasilnya disebut dengan elemen dari himpunan [8]. Pada ruang sampel terbatas, setiap ruang sampel disebut dengan peristiwa dan bagian ruang sampel yang memiliki satu elemen disebut dengan peristiwa dasar [8]. Adapun Aturan peluang untuk kejadian A,B, dan \emptyset adalah sebagai berikut; [9]

$$1. 0 \leq P(A) \leq 1$$

Peluang suatu kejadian nilainya terletak diantara 0 dan 1 inklusif.

$$2. P(\emptyset) = 0$$

Peluang himpunan kosong adalah 0

$$3. P(A) \leq 1 - P(A)$$

Peluang komplemen suatu kejadian adalah 1 – peluang kejadian tersebut

$$4. P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

2.2.1 Probabilitas Bersyarat

Probabilitas bersyarat merupakan peluang kemunculan dimana suatu kejadian bergantung pada kejadian yang lainnya [10]. Probabilitas bersyarat dapat menggunakan *bayes theorem* sebagai pengambilan keputusannya.

Bayesian theorem ini merupakan cara untuk mengatasi ketidakpastian dengan menggunakan rumus *bayes* [11] yang dapat dilihat pada persamaan 1

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \quad (1)$$

Dimana :

$P(H|E)$: Probabilitas hipotesa (H) jika terdapat *evidence*

$P(E|H)$: Probabilitas munculnya *evidence* (E) jika diketahui hipotesa (H)

$P(H)$: Probabilitas kemunculan hipotesa (H) tanpa melihat *evidence* apapun

$P(E)$: Probabilitas kemunculan *evidence* (E)

2.3 Artificial Intelligence

Menurut *Oxford Dictionary Artificial Intelligence* (AI) merupakan teori dan perkembangan sistem komputer untuk dapat melakukan pekerjaan yang biasanya membutuhkan kecerdasan manusia seperti, persepsi visual, pengenalan suara, pembuat keputusan, dan penerjemahan antar bahasa [12]. Adapun pengertian *artificial intelligence* menurut para ahli, yaitu [13] :

1. H.A.Simon (1987) mengartikan “kecerdasan buatan (*artificial intelligence*) merupakan kawasan penelitian, aplikasi, dan intruksi yang terkait dengan pemrograman komputer untuk melakukan hal yang dalam pandangan manusia adalah cerdas”
2. Rich dan knight (1991) mengartikan “kecerdasan buatan (*artificial intelligence*) sebagai sebuah studi tentang bagaimana membuat komputer melakukan hal – hal yang pada saat ini dapat dilakukan dengan baik oleh manusia”

2.4 Bayesian Networks

Bayesian Networks atau biasa disebut sebagai *Belief Networks* atau *probabilistic causal networks* merupakan grafik struktural untuk mewakili relasi probabilitas antara sebagian besar variabel dengan melakukan inferensi probabilistik terhadap variabel – variabel tersebut [8].

Bayesian networks telah terbukti menjadi alat yang kuat untuk menemukan hubungan antar variabel [6]. *Bayesian networks* banyak digunakan sebagai penentu keputusan di bidang perbankan [14], bidang kesehatan [6], maupun penentu keputusan pada penggunaan transportasi [15].

Pada *bayesian networks* terdapat dua pembelajaran yang dilakukan, yaitu :

1. Structure Learning

Tahap pembelajaran ini merupakan tahap pengidentifikasian topologi [6] atau model pada *bayesian networks* yang nantinya akan menjadi acuan untuk pembuatan keputusan. Pada *structure learning* terdapat tiga pendekatan, yaitu [16] :

- a. *Search-and-score based method*
- b. *Constraint-based approaches*
- c. *Bayesian model average*

2. Parameter Learning

Tahap pembelajaran ini merupakan tahap menghitung estimasi pada setiap nilai yang ada pada variabel atau probabilitas bersyarat sesuai dengan topologi yang sudah ada [6]. Jika topologi tidak melalui tahap *structure learning* maka perhitungan estimasi dapat menggunakan dua metode, yaitu :

a. Maximum Likelihood Estimation (MLE)

MLE merupakan metode yang sering digunakan untuk mengestimasi nilai parameter dari data yang modelnya sudah diketahui. Suatu nilai parameter dikatakan bagus dilihat dari seberapa banyak kemungkinan untuk menghasilkan data yang diamati [17]. Pada variabel prediksi penyakit ginjal kronis terdapat dua jenis data, yaitu binomial dan multinomial. Pada data yang binomial maka perhitungan estimasi dapat dilihat pada persamaan 2 [17]

$$\theta = \frac{Na}{Na+Nb} \quad (2)$$

Sedangkan pada variabel yang datanya multinomial, maka perhitungan estimasinya dapat dilihat pada persamaan 3 [17]

$$\theta = \frac{Nk}{\sum_{\lambda} N_{\lambda}} \quad (3)$$

Pada variabel yang memiliki relasi dengan variabel lain, maka perhitungan estimasinya dapat dilihat pada persamaan 4 [17]

$$\theta_{xi|pxi} = \frac{N(P_{xi} \cap X_i)}{X_i} \quad (4)$$

b. Maximum a Posteriori Estimation (MAP)

3. METODE PENELITIAN

Metode penelitian yang dilakukan pada penelitian ini merupakan metode eksperimen karena metode ini merupakan salah satu metode kuantitatif dimana metode ini melakukan percobaan untuk melihat hasil dan menyelidiki kemungkinan sebab akibat [18]. Adapun tahapan dari metode penelitian ini yang dapat dilihat pada Gambar 1



Gambar 1. Metode Penelitian

3.1 Identifikasi Masalah

Identifikasi masalah merupakan tahap menemukan permasalahan yang terjadi. Pada penelitian ini masalah yang terjadi adalah pada data *chronic kidney disease* [19] terdapat variabel yang memiliki relasi terhadap variabel yang lain.

3.2 Pengumpulan Data

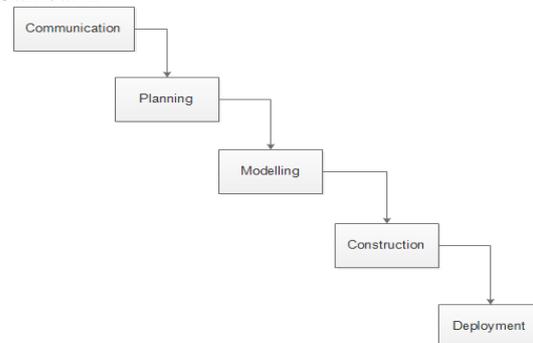
Pada tahap ini pengumpulan data dilakukan dengan cara melakukan wawancara dengan ahli yaitu Dr. Eva M Hidayat. Selain itu peneliti juga melakukan pengumpulan data dengan cara studi pustaka.

3.3 Analisis Data dan Kebutuhan Algoritma

Pada tahap ini peneliti melakukan analisis terhadap data masukan, *dataset* yang sudah dikumpulkan sebelumnya, serta keluaran data. Selain itu peneliti juga melakukan pemahaman konsep pada metode *Bayesian Networks*.

3.4 Pembangunan Perangkat Lunak

Pada tahap ini peneliti melakukan pembangunan perangkat lunak dengan menggunakan *Software Development Life Cycle* (SDLC) dengan menggunakan metode *waterfall* [13]. Tahap pembuatan perangkat lunak dapat dilihat pada Gambar 2



Gambar 2. Tahap Pembangunan Perangkat Lunak

3.5 Pengujian

Pada tahap ini dilakukan pengujian dengan menggunakan sistem yang sudah dibuat apakah hasil prediksi penyakit ginjal kronis sudah baik dan benar. Pengujian ini dilakukan dengan menguji fungsional pada sistem serta pengujian pada *dataset* dengan menggunakan metode *k-fold cross validation* dan *ROC curve*.

3.6 Penarikan Kesimpulan

Pada tahap ini peneliti melakukan penarikan kesimpulan dengan melihat dari hasil akurasi pada prediksi penyakit ginjal kronis.

4. HASIL DAN PEMBAHASAN

Sistem prediksi penyakit ginjal kronis yang akan dibangun terdiri dari beberapa proses utama, yaitu preprocessing, parameter learning dan validasi. Secara umum gambaran sistem yang akan dibangun dapat dilihat pada Gambar 3



Gambar 3. Gambaran Umum Sistem

Data masukan (data latih dan data uji) pada prediksi penyakit ginjal kronis ini didapatkan dari *chronic kidney dataset* [19] yang telah dilakukan validasi kepada Dr. Eva M Hidayat. Hasil dari validasi didapatkan bahwa dari 25 variabel yang ada hanya 13 variabel yang digunakan dimana 12 untuk memprediksi penyakit ginjal kronis dan 1 variabel merupakan *class* prediksinya. Variabel – variabel yang digunakan dapat dilihat pada Tabel 2

Tabel 2. Deskripsi Data Masukan

Variabel	Deskripsi	Type	Value
Age	Age	Numerical	Umur dalam tahun
BP	Blood pressure	Numerical	mm/Hg
AL	Albumin	Nominal	{0,1,2,3, 4,5}
BGR	Blood glucose random	Numerical	Mgs/dl
BU	Blood urea	Numerical	Mgs/dl
SC	Serum creatinine	Numerical	Mgs/dl
SOD	Sodium	Numerical	mEq/L
POT	Potassium	Numerical	mEq/L
HTN	Hypertenti on	Nominal	Yes, No
DM	Diabetic mellitus	Nominal	Yes, No
APPET	Appetite	Nominal	Good, Poor
PE	Pedal edema	Nominal	Yes, No
Class	Class	Nominal	Yes, No

4.1 Pre-processing

Pada tahap *preprocessing* data akan dilakukan pengkodean dan dicek apakah *class* pada data sudah seimbang atau belum, jika data belum seimbang maka *class* data akan diseimbangkan dengan menggunakan teknik *random oversampling* yang merupakan teknik yang dilakukan dengan cara melakukan penyalinan dan pengulangan data yang ada pada kelas minoritas sehingga keseimbangan data tercapai atau teknik menyalin data asli sebagai sample berulang dalam jumlah yang besar [20, 21]. Alur dari tahap *preprocessing* ini dapat dilihat pada Gambar 4



Gambar 4. Alur Preprocessing

Pada data nominal pengkodeannya adalah 1 dan 0. Sedangkan pada data *numerical* dilakukan pengkodean dengan cara mendapatkan hasil *range* dari studi pustaka [22, 23, 24, 25, 26] dan variabel *age* mengacu pada pengkodean yang ada pada penelitian P.Fuster-Parra [6]. Pengkodean variabel dapat dilihat pada Tabel 3

Tabel 3. Pengkodean Variabel

Variabel	Value	Keterangan
AL	1	Ada Kandungan albumin pada urin (1-5)
	0	Tidak ada kandungan albumin pada urin (0)
HTN	1	Menderita hipertensi (<i>yes</i>)
	0	Tidak menderita hipertensi (<i>no</i>)
DM	1	Menderita diabetes melitus (<i>yes</i>)
	0	Tidak menderita diabetes melitus (<i>no</i>)
APPET	1	Nafsu makan baik (<i>good</i>)
	0	Nafsu makan tidak baik (<i>poor</i>)
PE	1	Mengalami pembengkakan pada kaki (<i>yes</i>)
	0	Tidak mengalami pembengkakan pada kaki (<i>no</i>)
Age	1	< 18 tahun
	2	18 – 35 tahun
	3	35 – 55 tahun
	4	>55 tahun
BP	1	Blood Pressure Normal (< 85 mmHg)
	2	Blood Pressure High – Normal (85 mmHg – 89 mmHg)
	3	Blood Pressure High (≥ 90 mmHg)
SOD	1	Sodium Rendah (< 135 mEq/L)
	2	Sodium Normal (135

		mEq/L – 145 mEq/L)
	3	Sodium Tinggi (> 145 mEq/L)
POT	1	Potassium Rendah (≤ 3,5 mEq/L)
	2	Potassium Normal (3,6 mEq/L – 5,2 mEq/L)
	3	Potassium Tinggi (≥ 5,3 mEq/L)
BGR	1	Gula Darah Normal (< 100 mgs/dL)
	0	Gula Darah Tinggi (≥ 100 mgs/dL)
BU	1	Blood Urea Nitrogen normal (7mg/dL – 20 mg/dL)
	0	Blood Urea Nitrogen tinggi (≥ 20 mg/dL)
SC	1	Serum Creatine normal (0,8 mg/dL – 1,2 mg/dL)
	0	Serum Creatine Tinggi (> 1,2 md/dL)
CLASS	1	Terdiagnosa penyakit ginjal kronis (<i>ckd</i>)
	0	Tidak terdiagnosa penyakit ginjal kronis (<i>notckd</i>)

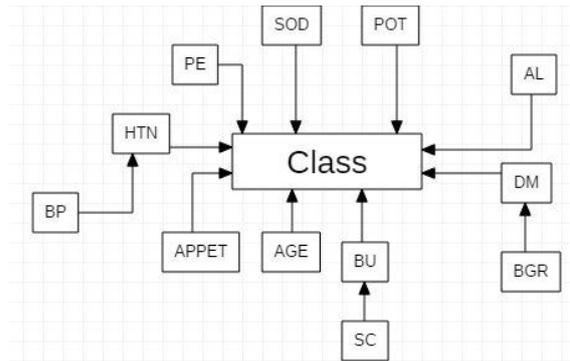
4.2 Learning Bayesian Networks

Setelah dilakukan tahap *preprocessing* maka tahap selanjutnya adalah tahap pembelajaran pada metode *bayesian networks*. Tahap pembelajaran dapat dilihat pada Gambar 5



Gambar 5. Alur Pembelajaran Bayesian Networks

Pada tahap *structure learning*, topologi sudah dibuat berdasarkan hasil wawancara dan studi pustaka. Topologi untuk prediksi penyakit ginjal kronis ini dapat dilihat pada Gambar 6



Gambar 6. Topologi Prediksi Penyakit Ginjal Kronis

Dilihat dari Gambar 5 maka perhitungan keputusan atau inferensi penyakit ginjal kronis ini dapat dilihat pada persamaan 5

$$P(class, age, appet, pe, al, bu, sc, dm, bgr, htn, pot, sod, bp) = P(age)P(appet)P(sod)P(pot)P(pe)P(al)P(htn|bp)P(dm|bgr)P(bu|sc)P(class|htn, dm, sod, pot, appet, age, al, pe, bu) \quad (5)$$

Tahap selanjutnya adalah tahap perhitungan estimasi nilai pada setiap variabel penentu prediksi penyakit ginjal kronis. Pada tahap ini digunakan metode *maximum likelihood estimation* (MLE). Pada variabel yang datanya berupa binomial, dengan menggunakan persamaan 2 maka didapatkan hasil estimasi pada variabel PE yang dapat dilihat pada Tabel 4

Tabel 4. Estimasi Nilai Pada Data Binomial

PE	P(PE)
1	0,18
0	0,82

Pada variabel yang datanya berupa multinomial dengan menggunakan persamaan 3, maka didapatkan hasil estimasi pada variabel SOD yang dapat dilihat pada Tabel 5

Tabel 5. Hasil Estimasi Pada Data Multinomial

SOD	P(SOD)
1	0,19
2	0,67
3	0,14

Pada variabel yang memiliki relasi lain maka perhitungan estimasi menggunakan persamaan 4. Hasil dari estimasi variabel DM yang memiliki relasi dengan variabel BGR dapat dilihat pada Tabel 6

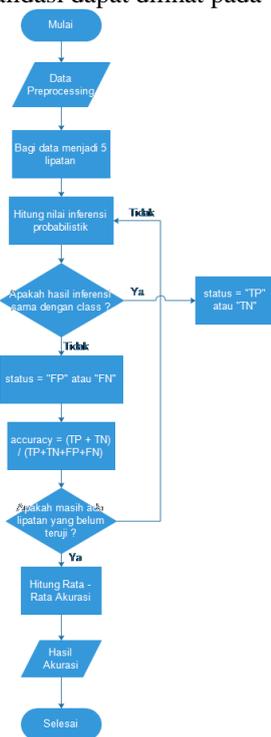
Tabel 6. Hasil Estimasi Pada Variabel Berelasi

BGR	P(DM BGR)	
	DM	
	1	0
1	0,14	0,86
0	0,36	0,64

4.3 Validasi

Setelah data dilakukan pengambilan keputusan dengan menghitung inferensi probabilistik dengan menggunakan persamaan 5, maka tahap selanjutnya adalah validasi dengan cara menghitung akurasi

dengan metode *k-fold cross validation* dan *ROC curve*. Alur validasi dapat dilihat pada Gambar 7



Gambar 7. Alur Validasi Prediksi Penyakit Ginjal Kronis

k-fold cross validation merupakan teknik yang biasa digunakan untuk mengestimasi kesalahan saat pengujian dimana estimasi ini dapat digunakan untuk melihat model terbaik dan memberikan gagasan tentang kesalahan pengujian model yang dipilih paling terakhir [27].

Perhitungan akurasi dengan menggunakan *ROC curve* yang merupakan teknik perhitungan akurasi dengan menggunakan grafik untuk memvisualisasikan, mengorganisir, dan menyeleksi klasifikasi yang tergantung pada performanya [28]. Pada *ROC curve* dibuat matriks untuk melihat perbandingan data asli dan data pengujian yang dapat dilihat pada Gambar 8

	POS	NEG	
POS	TP	FP	P + N
NEG	FN	TN	
	P	N	

Gambar 8. Matriks *ROC curve*

Pengujian pada *ROC curve* ada tiga kriteria yaitu akurasi, sensitivitas, dan spesivitas yang perhitungannya dapat dilihat pada persamaan 6 sampai persamaan 8

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad (6)$$

$$\text{Sensitivity} = \frac{TP}{P} \quad (7)$$

$$\text{Specificity} = \frac{TN}{N} \quad (8)$$

4.4 Pengujian

Pada tahap pengujian ini dataset dimasukkan dengan cara *import*. Sistem akan melakukan tahap

preprocessing ketika data dimasukkan sehingga data yang tersimpan pada database merupakan data hasil preprocessing. Setelah data dimasukkan maka tahap selanjutnya adalah perhitungan estimasi. Nilai estimasi yang di dapatkan akan menjadi nilai untuk perhitungan inferensi pada data uji.

4.4.1 Pengujian Akurasi

Pengujian akurasi dilakukan untuk mengetahui tingkat akurasi yang dilakukan oleh sistem menggunakan algoritma bayesian. Pengujian dilakukan menggunakan *k-fold cross validation* dengan jumlah $k = 5$ dimana dataset akan dibagi menjadi 5 bagian yaitu K1, K2, K3, K4, K5 dengan iterasi pengujian sebanyak 5 kali.

Jumlah data yang digunakan adalah 260 dengan masing – masing data (K) adalah 52 data dimana jumlah class prediksinya sama rata Percobaan pengujian dilakukan sebanyak lima kali, dimana tiap dataset akan secara bergantian menjadi data latih dan data uji. Skenario pengujian dapat dilihat pada Tabel 7

Tabel 7. Skenario Pengujian

No	Training	Testing
1	K1,K3,K4,K5	K2
2	K1,K2,K4,K5	K3
3	K1,K2,K3,K5	K4
4	K1,K2,K3,K4	K5
5	K2,K3,K4,K5	K1

Setelah dilakukan perhitungan dengan menggunakan persamaan 5 data uji dan dibandingkan dengan *class* pada data aslinya. Dengan menggunakan *k-fold cross validation* didapatkan hasil akurasi yang dapat dilihat pada Tabel 8

Tabel 8. Hasil Akurasi Pengujian

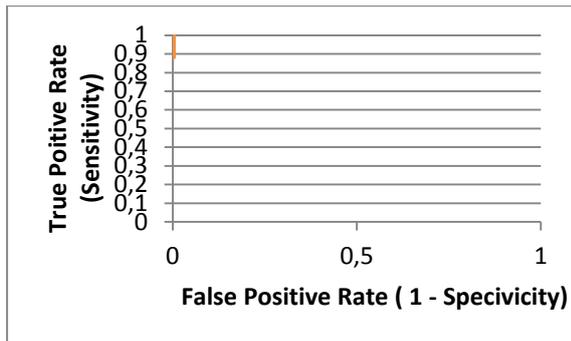
No	Testing	Akurasi
1	K2	0,981
2	K3	0,942
3	K4	0,981
4	K5	1
5	K1	1
Rata - Rata		0,981

Selain itu dengan menggunakan data uji yang sama dilakukan validasi dengan menggunakan metode *ROC curve*. Dengan menggunakan persamaan 7 dan persamaan 8, hasil pengujian dapat dilihat pada Tabel 9

Tabel 9. Hasil Pengujian *ROC curve*

No	Testing	Sensitivitas	Spesivitas
1	K2	0,962	1
2	K3	0,885	1
3	K4	0,962	1
4	K5	1	1
5	K1	1	1
Rata - Rata		0,962	1

Dilihat dari hasil pengujian pada dataset dengan menggunakan *ROC curve*, maka area dibawah kurva (AUC) pada pengujian prediksi penyakit ginjal kronis dapat dilihat pada Gambar 9



Gambar 9. Hasil Perhitungan ROC curve

5. PENUTUP

Setelah melakukan pengujian pada dataset dengan menggunakan *k-fold cross validation* didapatkan bahwa nilai rata – rata akurasi adalah sebesar 98,1%. Dapat disimpulkan bahwa hasil prediksi penyakit ginjal kronis dengan menggunakan *bayesian networks* meningkat dibandingkan dengan hasil prediksi penyakit ginjal kronis pada penelitian sebelumnya yang menggunakan metode *naive bayes*. Selain itu perhitungan akurasi dengan menggunakan metode ROC curve jika dilihat dari *area under curve* model yang digunakan untuk prediksi penyakit ginjal kronis ini tidak baik karena area dibawah kurva tidak diketahui. Hal ini dikarenakan pada tahap pembelajaran pembangunan struktur (*structural learning*) dibuat sudah pasti.

Adapun saran dari penelitian ini yaitu pada tahap *structural learning* sebaiknya dibangun menggunakan pendekatan yang sudah ada pada *bayesian networks* seperti *search-score based*, *constrain-based method*, serta *bayesian model averaging*. selain itu variabel yang digunakan untuk memprediksi penyakit ginjal kronis dapat digunakan semua variabelnya untuk melihat relasi yang lain antar variabelnya.

UCAPAN TERIMA KASIH

Terima kasih kepada Dr. Eva M Hidayat selaku validator penulis yang telah menyempatkan waktunya memberikan ilmu serta arahnya agar penulis dapat menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] InfoDatin, Situasi Penyakit Ginjal Kronis, Jakarta: Kementrian Kesehatan RI, 2017.
- [2] K. H. Australia, “All About Chronic Kidney Disease (CKD),” Kidney Health Australia, Mellbourn, 2015.
- [3] C. C. Amiranti, “Waspadai Gagal Ginjal The Silent Killer,” Kompas, [Online]. Available: <https://lifestyle.kompas.com/read/2017/02/27/070700623/waspadai.gagal.ginjal.the.silent.killer..> [Diakses 24 Maret 2018].
- [4] L. Jena dan N. Kamila, “Distributed Data Mining Classification Algorithm for Prediction

Chronic-Kidney-Disease,” *International Journal of Engineering Research in Management&Technology*, vol. IV, no. 11, pp. 110-118, 2015.

- [5] K. Collins-Thompson, “Naive Bayes,” Coursera, [Online]. Available: <https://www.coursera.org/learn/python-machine-learning/lecture/0XFms/naive-bayes-classifiers>. [Diakses 10 Mei 2018].
- [6] P.Fuster-Parra.et.al, “Bayesian network modelling : A case study of an epidemiologic system analysis of cardiovascular risk,” *Comput. Methods Programs Biomed.*
- [7] M. Spiegel dan L. J. Stephens, *Schaum's Outlines : Statistik Edisi Ketiga*, Jakarta: Erlangga, 2007.
- [8] R. E. Neapolitan, *Learning Bayesian Networks*, New Jersey: Pearson Prentice Hall, 2004.
- [9] S. Nugroho, *Dasar – Dasar Metode Statistika*, Jakarta: Grasindo, 2007, pp. 6, 53-54.
- [10] D. Afriyani, *Matematika Kelompok Teknologi, Kesehatan, dan Pertanian*, Bandung: Grafindo Media Pratama, 2008.
- [11] P. S. Dewi, “SISTEM PAKAR DIAGNOSIS PENYAKIT IKAN KOI,” *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. 4, no. 1, pp. 25-32, 2015.
- [12] O. University, “English Oxford Living Dictionaries,” Oxford University, [Online]. Available: https://en.oxforddictionaries.com/definition/artificial_intelligence. [Diakses 18 April 2018].
- [13] T. Harihayati, “Sistem Pakar Mendiagnosa Penyakit Umum Yang Sering Diderita Balita Berbasis Web Di Dinas Kesehatan Kota Bandung,” *Jurnal Komputer dan Informatika (KOMPUTA)*, vol. 1, no. 1, pp. 65-69, 2012.
- [14] M. W. Triki, “Bank Credit Risk: Evidence from Tunisia using Bayesian Networks,” *Journal of Accounting, Finance, and Auditing Studies*, vol. III, no. 3, pp. 93 - 107, 2017.
- [15] Q. Wang, H. Sun dan Q. Zhang, “A Bayesian Network Model on the Public Bicycle Choice,” *Hindawi*, vol. 2017, pp. 1 - 13, 2017.
- [16] C. Berzan, *An Exploration of Structure Learning in Bayesian Networks*, Medford: Tufts University, 2012.
- [17] N. Friedman, *Learning Bayesian Networks From Data*, California: U.C Berkeley, 1998.
- [18] Suryana, *Metode Penelitian*, Bandung: Universitas Pendidikan Indonesia, 2010.
- [19] M. Lichman, “UCI Machine Learning Repository,” 2013. [Online]. Available: <http://archive.ics.uci.edu/ml>. [Diakses 28 Februari 2017].

- [20] T. Hoens dan N. V. Chawla, *Imbalanced Datasets : From Sampling To Classifier*, Notre Dame: John Wiley & Sons, Inc., 2013.
- [21] Xionggao.et.al, "SROT: Sparse representation-based over-sampling technique for classification of imbalanced dataset," *IOP Conference Series : Earth and Environmental Science*, vol. 81, no. 1, pp. 1-10, 2017.
- [22] M. Clinic, "Blood Urea Nitrogen (BUN) test," Mayo Clinic, [Online]. Available: <https://www.mayoclinic.org/tests-procedures/blood-urea-nitrogen/about/pac-20384821>. [Diakses 27 Juni 2018].
- [23] M. Clinic, "Creatine Test," Mayo Clinic, [Online]. Available: <https://www.mayoclinic.org/tests-procedures/creatinine-test/about/pac-20384646>. [Diakses 27 Juni 2018].
- [24] M. C. Staff, "High Potassium (Hyperkalemia)," Mayo Clinic, [Online]. Available: <https://www.mayoclinic.org/symptoms/hyperkalemia/basics/definition/sym-20050776>. [Diakses 27 Juni 2018].
- [25] M. C. Staff, "Hyponatremia," Mayo Clinic, [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/hyponatremia/symptoms-causes/syc-20373711>. [Diakses 27 Juni 2018].
- [26] M. C. Staff, "Low Potassium (Hypokalemia)," Mayo Clinic, [Online]. Available: <https://www.mayoclinic.org/symptoms/low-potassium/basics/definition/sym-20050632>. [Diakses 27 Juni 2018].
- [27] M. A. L. Fernandez, *Cross Validation*, London: London of School Hygiene & Tropical Medicine, 2015.
- [28] T. Fawcett, "An introduction to ROC analysis," *Elsevier*, vol. 27, no. 12, pp. 861-874, 2006.
- [29] K/DOQI, *Clinical Practice Guidelines For Chronic Kidney Disease : Evaluation, Classification, and Stratification*, New York: National Kidney Foundation, Inc, 2002.