

OPTIMASI ALGORITMA NAÏVE BAYES MENGGUNAKAN FORWARD SELECTION UNTUK KLASIFIKASI PENYAKIT GINJAL KRONIS

Miftahul Rizal¹, Muhamad Zakhy Syahaf², Satrio Rully Priyambodo³, Yudi Ramdhani⁴

Program Studi Tenknik Informasi, Teknik Informatika^{1,2,3,4}

Universitas Adhirajasa Reswara Sanjaya^{1,2,3,4}

miftahulrizal04@email.com¹, zakhysyahaf@email.com², satrioarul@email.com³, yudi@ars.co.id⁴

Abstrak

Penyakit Ginjal Kronis (PGK) ialah disfungsi ginjal liberal dan *ireversibel* di mana keahlian tubuh untuk mendukung badan mempertahankan metabolisme dan keseimbangan elektrolit tersendat. Penyakit ginjal kronis akan bertambah sebagai gagal ginjal terminal maupun ginjal tingkatan akhir penyakit dimana ginjal tidak sanggup bertugas lagi untuk melindungi materi ke dalam badan, membutuhkan pemulihan langkah- langkah lebih lanjut dari *dialisis* maupun *transplantasi* ginjal sebagai pemulihan pengganti ginjal. Penyakit ginjal adalah masalah kesehatan diseluruh dunia, yang memiliki Peningkatan *prevalensi* gagal ginjal dan *prognosis* buruk dan biaya yang tinggi. Penyakit ginjal kronis bisa menjadi sulit untuk disembuhkan, penyebab utama termasuk diabetes dan tekanan darah tinggi. Penelitian memperoleh dataset dari *UCI machine learning repository* dan terdapat banyak dataset, salah satunya ialah *dataset Risk Factor prediction of Chronic Kidney Disease*. Studi ini bermaksud untuk mengukur tingkat keakuratan penggolongan informasi data mining, studi ini menggunakan algoritma *C4.5*, *K-NN*, *Naïve bayes*, *Logistic regression*. Dari algoritma keempat akan dibandingkan pengklasifikasian dataset *Risk Factor prediction of Chronic Kidney Disease*, dataset yang diperoleh dari *UCI machine learning repository*. Dari hasil akurasi yang didapat dengan menggunakan Algoritma *C4.5* menghasilkan akurasi 90,45%, Algoritma *K-NN* menghasilkan kejituan 91,50%, Algoritma *Naïve Bayes* menghasilkan akurasi 92,92%, Algoritma *Logistic Regresion* menghasilkan akurasi 80,09%. sehingga bisa disimpulkan jika algoritma yang paling tidak akurat ialah algoritma *Naïve Bayes*.

Kata kunci : *C4.5*, *K-NN*, *Logistic regression*, *Naïve bayes*, *Penyakit ginjal kronis*.

Abstract

Chronic Kidney Disease (CKD) is progressive and irreversible kidney dysfunction in which the body's ability to support the body to maintain metabolism and electrolyte balance is hampered. Chronic kidney disease will develop into terminal kidney failure and end-stage kidney disease where the kidneys are no longer able to function to protect substances into the body, requiring further recovery steps from dialysis or kidney transplantation as a kidney replacement remedy. Kidney disease is a worldwide health problem, which has an increasing prevalence of kidney failure and a poor prognosis and high cost. Chronic kidney disease can be difficult to treat, leading causes include diabetes and high blood pressure. The research obtained datasets from the UCI machine learning repository and there are many datasets, one of which is the Risk Factor prediction of Chronic Kidney Disease dataset. This study intends to measure the level of accuracy of data mining information classification, this study uses the C4 algorithm. 5, K-NN, Naïve bayes, Logistic regression. The fourth algorithm will compare the classification of the Risk Factor prediction of Chronic Kidney Disease dataset, the dataset obtained from the UCI machine learning repository. From the accuracy results obtained using the C4 Algorithm. 5 produces 90.45% accuracy, K-NN Algorithm produces 91.50% accuracy, Naïve Bayes Algorithm produces 92.92% accuracy, Logistic Re-gression Algorithm produces 80.09% accuracy. so that it can be concluded that the most inaccurate algorithm is the Naïve Bayes algorithm.

Keywords : *C4.5*, *Chronic Kidney Disease*, *K-NN*, *Logistic regression*, *Naïve bayes*.

I. PENDAHULUAN

Penyakit ginjal kronis adalah disfungsi ginjal yang progresif serta ireversibel dimana keahlian badan buat mempertahankan metabolisme serta penyeimbang air serta elektrolit tersendat. Kandas ginjal kronis bisa tumbuh jadi penyakit ginjal stadium akhir ataupun penyakit ginjal stadium akhir kala ginjal tidak sanggup lagi menahan zat- zat di dalam badan, sehingga membutuhkan langkah penindakan lebih lanjut semacam mencuci darah ataupun transplantasi ginjal buat menghindari penyakit ginjal [1]. Penyembuhan PGK bisa dicoba dengan 2 metode, yakni mula- mula transplantasi ginjal serta hemodialisa ataupun pemurnian darah. Hemodialisis merupakan proses mensterilkan darah dengan mengumpulkan produk limbah. Pada penderita kandas ginjal kronik, hemodialisis bisa menghindari kematian [2].

Penyakit ginjal kronis adalah kasus kesehatan seluruh dunia, yang memiliki tingkat prevalensi dan kegagalan ginjal yang makin naik setiap tahunnya, prediksi awal yang buruk, dan biaya yang tinggi [3]. Penyakit ginjal ini disebut "penyakit diam" karena sering tidak menunjukkan gejala. Namun, penyakit ginjal kronis bisa menjadi sulit untuk disembuhkan. Penyebab utama termasuk diabetes dan tekanan darah tinggi [4]. Terkait dengan studi prediksi penyakit ginjal, khususnya penyakit ginjal kronis, sudah dilakukan penelitian oleh Parul Sinha dan Poonam Sinha pada tahun 2015 dengan menggunakan data sekunder dan metode *K-NN*. Studi ini dibandingkan dengan kajian tentang prediksi penyakit ginjal kronis dengan bantuan metode *K-NN* dan *SVM* untuk mempelajari metode baru dan melakukan analisis. Hasil studi menunjukkan bahwa metode *SVM* memiliki tingkat keakuratan sebesar 73.75% dan metode *K-NN* sebesar 78.75% [5].

Menurut Hermawati, *data mining* adalah proses yang menggunakan Teknik untuk analisis otomatis dan ekstraksi data[6]. *Data mining* adalah proses berulang dan interatif untuk mencari model atau pola yang lengkap, berguna, dan dapat dipahami dalam database yang sangat besar. Basis data pencarian tren atau pola yang diinginkan dalam *database* untuk mendukung pengambil keputusan di masa depan. Beberapa alat digunakan untuk mengidentifikasi pola ini dan membagikan analisis yang informatif dan mendalam untuk penyelidikan yang kemudian dapat dieksplorasi lebih lanjut dan digunakan oleh alat pendukung keputusan lainnya [7].

Klasifikasi adalah proses pengelompokan objek dengan fitur atau karakteristik yang sama ke dalam beberapa kelas. Pemrosesan dokumen secara otomatis bisa dilakukan dengan menggunakan fitur kata atau karakteristik yang muncul dalam materi pelatihan. Namun, dokumen yang besar dan banyak jumlahnya menyebabkan adanya banyak entri kata yang muncul. Oleh karena itu, untuk mempermudah proses klasifikasi, digunakan ringkasan yang memiliki jumlah kata yang lebih sedikit[8].

Naive Bayes merupakan metode yang baik untuk masalah klasifikasi karena mudah dipahami dan tidak memerlukan skema rumit untuk mengestimasi parameter. Ini membuat metode ini dapat diterapkan pada *dataset* yang besar dan kompleks[9]. *Naive Bayes* memiliki interpretasi yang mudah sehingga meskipun pengguna tidak memiliki keahlian dalam bidang teknologi klasifikasi, mereka masih dapat memahami metode ini. Algoritma ini juga memiliki tingkat kesalahan yang lebih rendah dibandingkan dengan metode klasifikasi lainnya. Penelitian ini, tidak hanya bertujuan untuk mencapai tingkat akurasi yang baik, tetapi juga untuk menentukan atribut model yang memiliki pengaruh dengan menerapkan seleksi fitur[10].

Pemilihan fitur adalah proses identifikasi fitur terpenting dalam data. Hal ini sangat penting karena ada beberapa atribut atau karakteristik relevan untuk masalah yang sedang dianalisis, dan beberapa fitur bahkan bisa mengurangi akurasi analisis. Karena itu, fitur yang tidak berguna atau menimbulkan kebisingan harus dihapus untuk meningkatkan akurasi. Ada beberapa teknik pemilihan fitur, seperti *forward selection*, *backward elimination*, dan *stepwise selection*. Teknik *forward selection* memberikan hasil yang lebih dan mengurangi waktu pemrosesan dibandingkan dengan *backward elimination* atau *stepwise selection* [11]. Tujuan dari penelitian ini adalah untuk menemukan algoritma klasifikasi penyakit ginjal yang tepat dan akurat dengan menggunakan klasifikasi yaitu algoritma C4.5, algoritma *K-NN*, algoritma *Naive Bayes* dan algoritma *regresi logistik*.

II. TINJAUAN PUSTAKA

1. Penyakit Ginjal Kronis

Gagalnya fungsi ginjal makin meningkat secara bertahap. Kasus tertentu, PGK dapat dideteksi pada tahap awal dan diobati untuk memperlambat atau menghentikan perkembangan penyakit, yang dapat mengarah pada perkembangan tahap ginjal. Beberapa kasus kerusakan ginjal dimulai dengan gangguan fungsi glomerulus menyebabkan darah mengeluarkan air dan molekul ke dalam urin, tetapi menahan protein. Penyebab penyakit ginjal yang paling umum adalah diabetes dan hipertensi, keduanya dapat menyebabkan nefron terbentuk di ginjal. Faktor risiko kedua adalah obesitas, yang dapat menyebabkan darah menggembung di ginjal. Faktor risiko ketiga adalah diabetes, yang dapat menyebabkan nefron terbentuk di ginjal. Faktor risiko keempat adalah diabetes, yang dapat menyebabkan nefron terbentuk di ginjal [12].

2. *Forward Selection*

Forward selection juga dikenal merupakan prosedur penting dalam persiapan data dalam penambangan data untuk mengoptimalkan kinerja dan mempercepat eksekusi suatu algoritma. Teknik memungkinkan pemilihan yang luas dari subset fitur dengan jumlah pilihan yang banyak, sehingga mengucualikan fitur yang tidak penting dan relevan untuk klasifikasi. *Forward selection* juga dikenal sebagai pemilihan fitur, pemilihan *subset*, pemilihan atribut, atau pemilihan *variabel*, juga dapat didefinisikan sebagai proses pemilihan atribut yang sesuai untuk digunakan dalam proses klasifikasi atau pengelompokan untuk tujuan pengecualian. Kompleksitas algoritma klasifikasi, menyetel akurasi algoritma klasifikasi, dan menentukan atribut yang mempengaruhi tingkat akurasi [13].

3. *Tools*

RapidMiner tools yang digunakan untuk penelitian, karena *RapidMiner* memiliki kemampuan bagus dan mudah digunakan, serta kemampuan grafis dan rendering yang terbatas. *RapidMiner* memudahkan pemakai untuk menggunakan perangkat lunak ini. Perangkat lunak ini adalah sumber terbuka dan dibuat menggunakan Bahasa Java yang dilisensikan di bawah Lisensi Publik GNU *RapidMiner* dapat berjalan di sistem operasi apapun. Melalui penggunaan Penambang cepat, tidak diperlukan keterampilan Pengkodean khusus dengan fungsionalitas penuh sudah disediakan. *RapidMiner* berspesialisasi dalam menggunakan penambangan data [14].

4. *Penelitian Terdahulu*

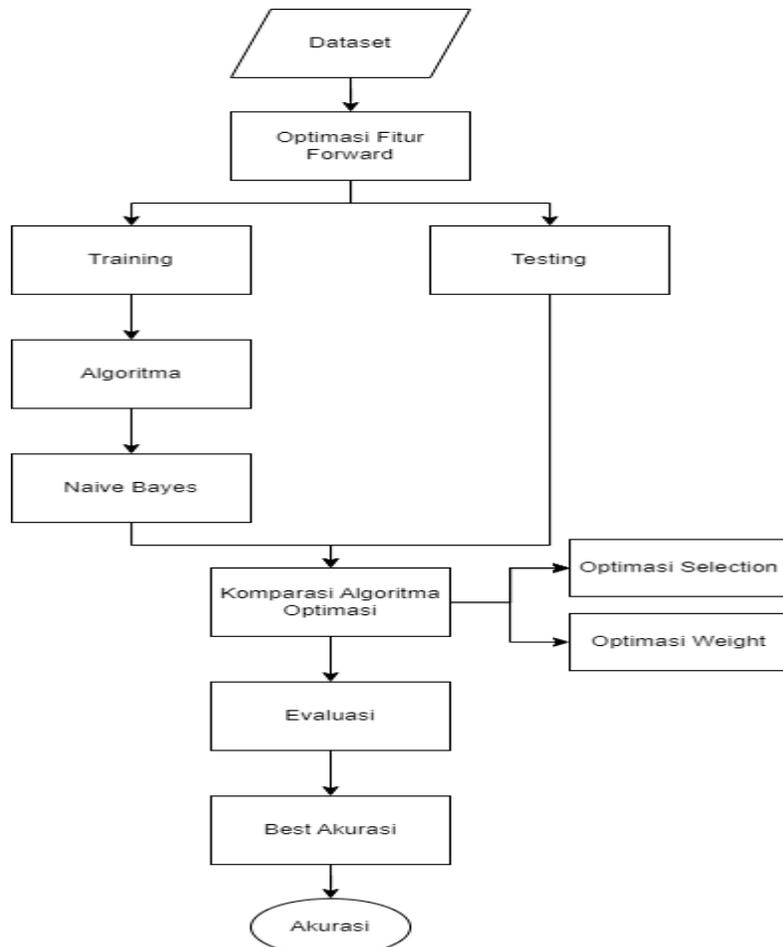
Tabel di bawah ini adalah Tabel 1 tentang penelitian dan pengembangan sebelumnya yang terkait dengan deteksi PGK. Jurnal berikut ini antara lain.

TABEL I
PENELITIAN SEBELUMNYA

Nama Jurnal	Penulis dan Tahun	Masalah	Hasil	Kesimpulan
Implementasi data mining untuk deteksi penyakit ginjal kronis (PGK) menggunakan <i>knearest neighbor</i> (K-NN) dengan <i>backward elimination</i>	(Gamada renda & Waspada, 2020)	Penyakit ginjal kronis adalah masalah kesehatan yang berkembang di seluruh dunia	Akurasi menggunakan metode <i>KNN</i> dengan eliminasi Fitur <i>backward</i> hasil 99,25%.	Saat memakai fitur <i>backward</i> akurasi dan kinerja yang dihasilkan lebih baik.
<i>classification of liver disease by applying random forest algorithm and backward elimination</i>	(Herliaw dkk, 2020)	Deteksi awal kanker, terutama terhadap kanker hati, penting untuk menghilangkan resiko kematian sangat tinggi disebabkan oleh penyakit ini.	Menggunakan <i>backward elimination</i> untuk mengoptimalkan akurasi dari algoritma <i>Random Forest</i> dan validasi <i>split validation</i> untuk memvalidasi permodelan. Hasil menunjukkan hasil akurasi 76.00% dan AUC 0.758	Hasil menunjukkan bahwa penelitian ini cukup baik untuk mengklasifikasi penyakit liver.

III. METODE PENELITIAN

Penelitian ini, algoritma *Naïve Bayes* digunakan dan dilakukan optimasi fitur dengan teknik *forward selection*. Tujuannya adalah untuk memprediksi faktor risiko penyakit ginjal kronis dengan menggunakan aplikasi *RapidMiner* versi 9.10.011



Gambar 1. Flowchart Optimasi Fitur Selection

1. Dataset

Penelitian ini, *dataset* diambil dari *UCI Repository* yang dikumpulkan melalui survei dan kuesioner yang dirancang untuk mengumpulkan informasi mengenai faktor risiko penyakit ginjal kronis. Data kesehatan pasien dikumpulkan melalui kuesioner di sekolah kedokteran di Bangladesh, termasuk informasi seperti usia, tekanan darah, hipertensi. Data tersebut disimpan dalam file CSV dan dibagi menjadi 68% untuk proses pelatihan dan 32% untuk proses pengujian.

TABEL II
KETERANGAN DATA

Attribute	Description
Blood pressure	Given in mm/mg
Specifity gravity	From 1005 to 10025 (the greater the danger).
Albumin	The scale runs from 0 to 5 (the higher the number, the better).
Sugar level	There are five severity degrees.
Red blood cells	Is the condition abnormal or normal? It is given in milligrams per liter.
Blood urea	A high level is not desirable.
Serum creatinine	A high level is not desirable.
Sodium	It is measured in milliequivalents per liter.
Potasium	It is measured in milliequivalents per liter.
Hemoglobin	A score of less than 15 indicates renal failure.
White blood cell count	This is numeric cell count

Red blood cell count	Should not be greater than or less than typical
Hypertension	It is binary (yes or no).
class	Whether diagnosed with CKD or not

2. Persiapan *Dataset*

Setelah data dikumpulkan, data harus diolah dengan melakukan preprocessing, karena data mungkin berisi data yang hilang atau tidak dibutuhkan. Pada tahap ini, kita mengisi data yang kosong dan mengubah beberapa data dari tipe string menjadi numerik

3. Validasi *Dataset*

Setelah proses *preprocessing* selesai, tahap selanjutnya adalah validasi data dengan membagi data menjadi data latih dan data uji menggunakan teknik *X cross-validation* untuk menguji performa terbaik. Algoritma dengan performa terbaik akan diambil untuk diuji.

4. Algoritma

C4.5: Algoritma *C4.5* adalah salah satu algoritma yang bisa digunakan untuk membuat pohon keputusan. Dari berbagai algoritma pohon keputusan, metode *ID3* dan *C4.5* yang dikembangkan oleh J. Ross Quinlan mungkin yang paling populer dan banyak digunakan dalam pembelajaran mesin. *C4.5* adalah peningkatan dari *ID3* yang dirilis pada tahun 1993. *C4.5* membagi data secara rekursif untuk membuat pohon keputusan. Algoritma ini mempertimbangkan atribut baik bertipe kategori maupun numerik. Dengan menghitung *return* dan *rate of return*, *C4.5* menentukan nilai tertinggi dari setiap atribut [15].

K-NN: Algoritma *KNN* merupakan metode untuk mengklasifikasikan objek berdasarkan data jarak belajar terdekat dari objek tersebut. Dalam algoritma ini, hasil dari setiap instance kueri diurutkan berdasarkan kelas terbanyak dan setelah dievaluasi akan menjadi kategori klasifikasi. Algoritma *KNN* mudah dipahami dan diimplementasikan, sehingga banyak digunakan dalam bidang kecerdasan buatan (AI) [16].

Naïve Bayes: Algoritma *Naive Bayes* merupakan metode untuk klasifikasi statistik yang memprediksi kemungkinan sebuah data masuk ke dalam kelas tertentu berdasarkan probabilitas. Algoritma ini didasarkan pada teorema *Bayes* karya *Thomas Bayes* dan terkenal dengan kecepatan dan akurasi yang tinggi saat mengolah database besar. *Naive Bayes* sering digunakan dalam machine learning karena memiliki akurasi tinggi dengan perhitungan yang sederhana [17].

Logistic Regression: Algoritma *logistic regression* adalah analisis regresi untuk memprediksi variabel dependen yang bernilai dikotomis, yaitu hanya memiliki dua nilai yang mewakili peristiwa seperti 0 atau 1. Ini berbeda dari regresi linier standar karena tidak ada asumsi hubungan linier antara variabel independen dan dependen. Regresi logistik adalah model prediktif dan pendekatan regresi linier biasa disebut *OLS (Ordinary Least Squares)*. Bedanya, dalam regresi logistik, peneliti memprediksi ukuran variabel dependen dalam skala dikotomis, seperti dua data nominal seperti ya atau tidak, baik atau buruk [18].

5. Optimasi Dataset

Tahap ini, optimasi data dilakukan dengan menggunakan dua metode optimasi, yaitu Fungsi *Select + Optimize Select*, untuk meningkatkan akurasi data sebelum algoritma memproses data. encapai akurasi yang lebih tinggi dan lebih baik dari sebelumnya.

6. Komparasi Algoritma

Pada tahap ini, algoritma *c4.5*, *k-NN*, *Naive Bayes*, dan *Regresi Logistik* dibandingkan berdasarkan akurasi klasifikasi mereka. Algoritma terbaik akan diuji menggunakan *teknik split validation dan cross validation*. Dengan *split ratio* mulai dari 0,5 sampai 0,9 dan *cross validation*.

7. Evaluasi

Penelitian ini bertujuan untuk menyajikan hasil data dengan akurasi yang terverifikasi. Model digunakan untuk mengevaluasi seberapa akurat setiap algoritma yang digunakan. Hasil klasifikasi akan diperiksa untuk

menentukan objek yang benar atau salah. Langkah-langkah untuk menampilkan hasil pengujian algoritma juga akan ditunjukkan [5].

IV. HASIL DAN PEMBAHASAN

Pada penelitian ini digunakan dataset memiliki 202 *record* dan 29 atribut dengan label atau target yang bernama PGK atau Non-PGK. Namun, ada masalah dengan data yang hilang karena beberapa objek atau kasus tidak memiliki data akibat kurangnya pemahaman atau informasi tentang objek tersebut. Walaupun biasanya tidak menjadi masalah besar jika hanya sedikit, namun jika persentase data yang hilang sangat tinggi, maka harus diambil langkah untuk memastikan apakah data statistik yang ada masih valid atau tidak. Penyebab umum dari masalah ini adalah responden tidak menjawab pilihan jawaban dalam survei.

1. Teknik *Preprocessing*

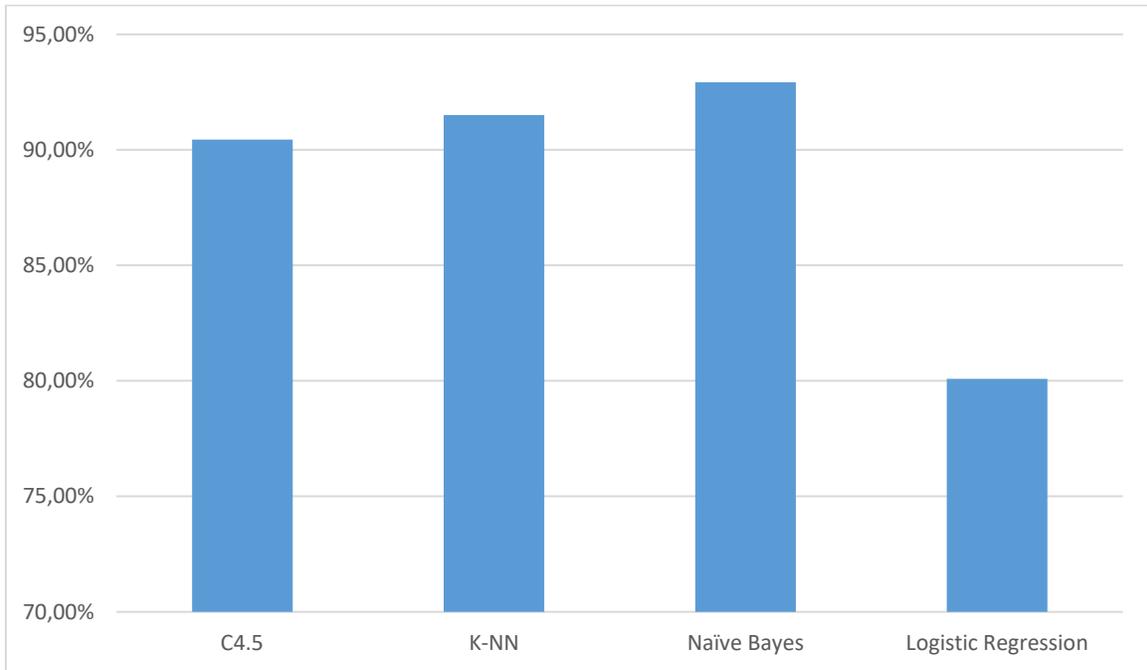
Preprocessing merupakan tahap penting dalam proses mining data pada penelitian ini. Sebelum melakukan pemrosesan data, data mentah harus diolah terlebih dahulu. Tahap ini melibatkan persiapan data seperti mengisi data yang kosong dan menghapus data yang tidak perlu.

2. Perbandingan Hasil Evaluasi Antara 4 Algoritma

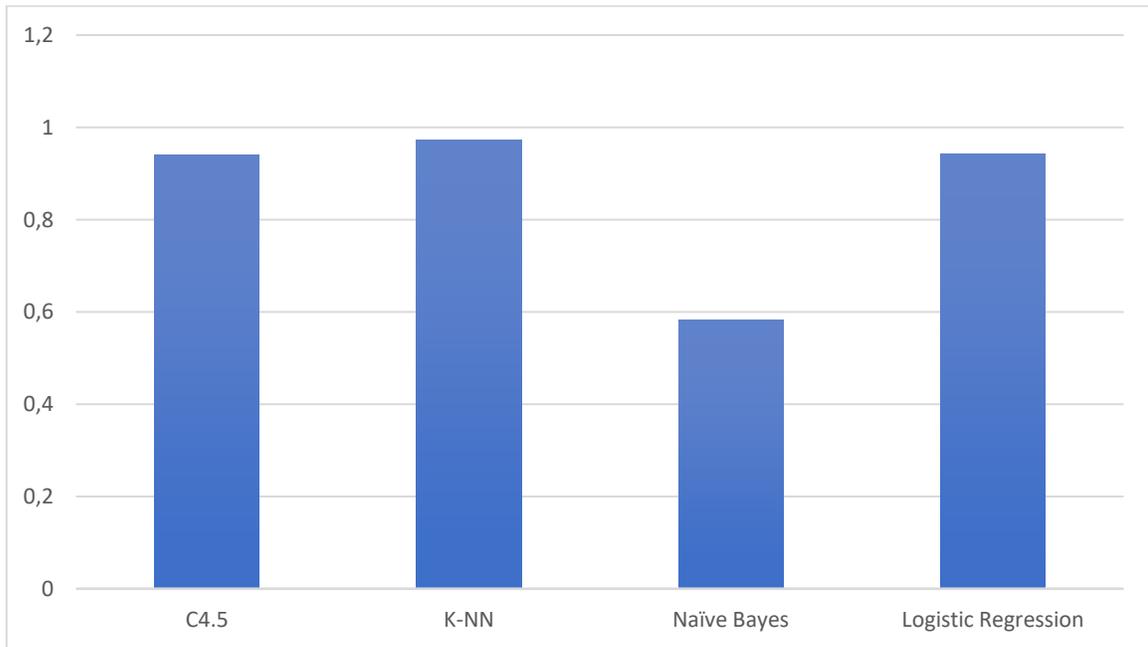
Pada penelitian ini menunjukkan bahwa hasil dari algoritma *Naive Bayes* memiliki hasil akurasi yang lebih baik dibandingkan dengan algoritma *regresi logistik*, *C4.5*, dan *K-NN*. Algoritma *C4.5* menghasilkan akurasi sebesar 90,45%, algoritma *K-NN* menghasilkan akurasi sebesar 91,50%, algoritma *Naive Bayes* menghasilkan akurasi sebesar 92,92%, dan algoritma *regresi logistik* menghasilkan akurasi sebesar 80,09%. kesimpulan dari penelitian ini bahwa algoritma *Naive Bayes* merupakan algoritma yang paling akurat.

TABEL III
PERBANDINGAN 4 ALGORITMA

Validation	Algoritma							
	C4.5		K-NN		Naive Bayes		Logistic Regression	
	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC
10 X Validation	90.45%	0.940	91.50%	0.973	92.92%	0.583	80.09%	0.943



Gambar 2. Grafik Perbandingan Akurasi 4 Algoritma

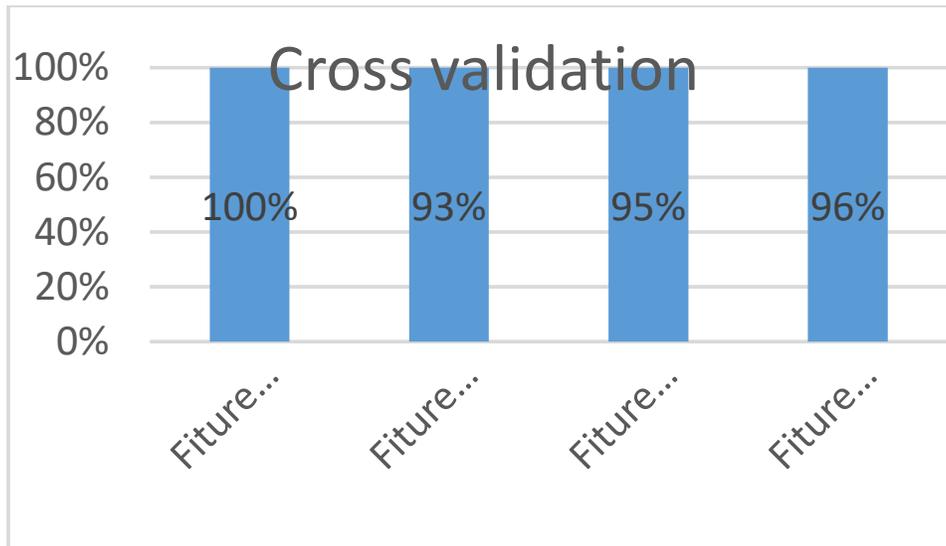


Gambar 3. Grafik Perbandingan Auc 4 Algoritma

Hasil klasifikasi yang diperoleh dari algoritma *Naive Bayes* belum sepenuhnya memuaskan, oleh karena itu, perlu dilakukan optimasi terlebih dahulu dengan menggunakan fitur *selection* + fitur *weighting*. Tujuannya adalah untuk membandingkan optimasi fitur dan menghasilkan data yang lebih baik, seperti yang terlihat pada tabel 3 dan 4.

TABEL IV
PENGUJIAN OPTIMASI DENGAN *CROSS VALIDATION*

Algoritma	Optimasi	Validasi	Akurasi	AUC
<i>Naïve Bayes</i>	<i>Fitur Selection + Optimize Selection</i>	<i>Cross Validation n X Vold</i>	100%	1.000
<i>Naïve Bayes</i>	<i>Fitur Selection + Optimize Selection (Evolutionary)</i>	<i>Cross Validation n X Vold</i>	93.00%	0.542
<i>Naïve Bayes</i>	<i>Fitur Weighting + Optimize Weight Evolutionary</i>	<i>Cross Validation n X Vold</i>	95.00%	1.000
<i>Naïve Bayes</i>	<i>Fitur Weighting + Optimize Weight PSO C</i>	<i>Cross Validation n X Vold</i>	95.50%	0.999



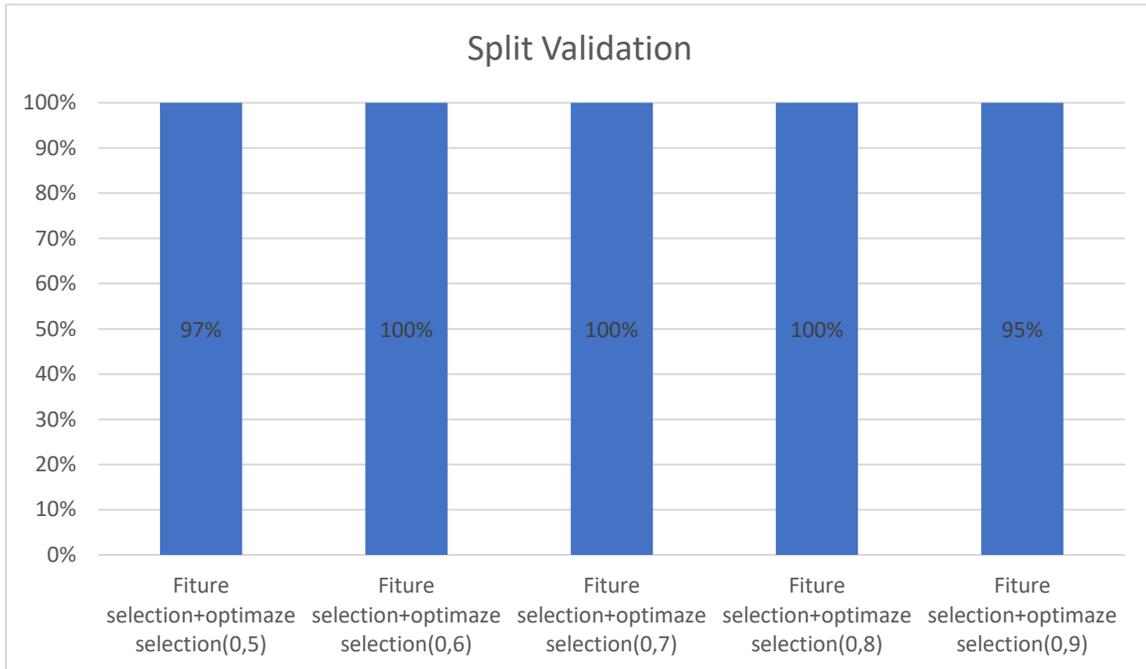
Gambar 4. Grafik Perbandingan *Fiture* Optimasi

Berdasarkan tabel 4, optimalisasi *fiture selection* dan *fiture weigthing* menghasilkan peningkatan nilai akurasi pada algoritma *Naïve Bayes*. Jika dibandingkan dengan fitur optimasi lainnya, fitur *optimize selection* dengan metode *optimaze selection* mencapai peningkatan akurasi yang lebih tinggi yaitu 100% Performa algoritma *Naïve Bayes* akan diuji setelah dioptimasi menggunakan fitur *optimize Selection*. Data divalidasi lebih lanjut menggunakan validasi split dengan rasio split 0,5 sampai 0,9. *RapidMiner* digunakan untuk mendapatkan hasil akurasi dan nilai AUC dan masing-masing rasio, Tabel di bawah ini. menunjukkan hasil pengujian kinerja dari algoritma *Naïve Bayes* setelah dioptimasi dengan fitur *optimize selection* dan metode *evolutionary*, seperti terlihat pada tabel 5.

TABEL V
OPTIMASI DATA DENGAN *SPLIT VALIDATION*

Algoritma	Optimasi	Validasi	Akurasi	AUC
<i>Naïve Bayes</i>	<i>Fitur Selection + Optimize Selection</i>	<i>Split Validation (0,5)</i>	96.97%	0.997

<i>Naïve Bayes</i>	<i>Fitur Selection + Optimize Selection</i>	<i>Split Validation (0,6)</i>	100%	1.000
<i>Naïve Bayes</i>	<i>Fitur Selection + Optimize Selection</i>	<i>Split Validation (0,7)</i>	100%	1.000
<i>Naïve Bayes</i>	<i>Fitur Selection + Optimize Selection</i>	<i>Split Validation (0,8)</i>	100%	1.000
<i>Naïve Bayes</i>	<i>Fitur Selection + Optimize Selection</i>	<i>Split Validation (0,9)</i>	95%	0.995



Gambar 5. Grafik Perbandingan *Fitur Optimasi* Menggunakan *Split Validation*

V. KESIMPULAN

Penyakit Ginjal Kronis adalah pembedahan ginjal yang *progresif* dan *ireversibel* dalam bentuk tubulus untuk tujuan meningkatkan metabolisme dan mengatur aktivitas ginjal. Gagal ginjal kronis dapat ditemukan di terminal ginjal atau stadium ginjal. Hal ini dikarenakan ginjal tidak dapat menggunakan bahan yang ada di dalam tubuh dan harus menggunakan pengobatan jangka panjang dari *dialisis* atau *transplantasi* ginjal sebagai pengobatan ginjal. *Insufisiensi* ginjal, gagal ginjal, penyakit ginjal stadium akhir yang disertai dengan komplikasi organ dan akhirnya menyebabkan kematian, penyakit ginjal stadium akhir yang disertai dengan komplikasi ginjal Sebagai bagian dari studi prediksi, data untuk kronis ginjal dianalisis dengan menggunakan 4 metode. Penelitian ini dilakukan menggunakan metode standar dan algoritma yang dipakai adalah *C4.5*, *KNN*, *Logistic Regression* dan *Naïve Bayes* untuk menganalisis hubungan antara keempat metode dan catatan PGK yang diambil dari *UCI repository* dataset tersebut, dan diolah menggunakan metode algoritma *Naive Bayes*, algoritma *C4.5*, algoritma *K-NN* dan algoritma *Logistic Regression*. Algoritma *C4.5* memiliki akurasi sebesar 90,45 persen, algoritma *K-NN* memiliki akurasi sebesar 91,50 persen, algoritma *Nave Bayes* memiliki akurasi sebesar 92,92 persen, dan algoritma *logistic regression* memiliki akurasi sebesar 80,09 persen. Karena itu, dapat dilihat bahwa algoritma terbaik adalah algoritma *Nave Bayes*. Terlihat bahwa penggunaan algoritma klasifikasi optimasi fitur berpengaruh signifikan terhadap peningkatan akurasi. Berdasarkan hasil penelitian yang diperoleh pada dataset penyakit ginjal kronis dapat diketahui bahwa optimasi algoritma *Naive Bayes* menggunakan *forward selection* dapat meningkatkan hasil akurasi pada algoritma *Naive Bayes* menjadi 100%.

REFERENSI

- [1] K. N. Siagian dan A. E. Damayanty, "Identifikasi Penyebab Penyakit Ginjal Kronik pada Usia Dibawah 45 Tahun di Unit Hemodialisis Rumah Sakit Ginjal Rasyida Medan Tahun 2015," 2015.
- [2] P. F. Wiliyanarti, "LIFE EXPERIENCE OF CHRONIC KIDNEY DISEASES UNDERGOING HEMODIALYSIS THERAPY," Mei 2019.
- [3] Kementerian Kesehatan RI, "Riset Kesehatan Dasar. Jakarta: Badan Penelitian dan Pengembangan Kesehatan Kementerian Kesehatan RI," 2013.
- [4] I. Yulianti, "IMPLEMENTASI SELEKSI ATRIBUT PADA ALGORITMA C4.5 MENGGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI PENYAKIT GINJAL KRONIS," 2018.
- [5] W. Yunus, "Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik," 2018.
- [6] Hermawati, Fajar, dan Astuti, "Data mining," 2013.
- [7] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," 2018.
- [8] N. Indriani, Ednawati Rainarli, dan Kania Evita Dewi, "Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen," Nov 2017.
- [9] Xindong Wu dan Vipin Kuma, "The top ten Algorithms in Data Mining," 2019.
- [10] M. F. Nugroho dan Setyoningsih Wibowo, "Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes," 2017.
- [11] Ladha dan Depp, "Feature Selection Methods And Algorithms," International Journal on Computer Science and Engineering (IJCSSE)," 2011.
- [12] Wasit Ginting, "Sistem Pakar Mendiagnosa Penyakit Ginjal Menggunakan Metode Case Based Reasoning," 2018.
- [13] Junta Zeniarja, Kiki Widia, dan Ramadhan Rakhmat Sani, "Penerapan Algoritma Naive Bayes dan Forward Selection dalam Pengklasifikasian Status Gizi Stunting pada Puskesmas Pandanaran Semarang," 2020.
- [14] Dwi Budi Srisulistiwati, Muhamad Khaerudin, dan Sri Rejeki, "SISTEM INFORMASI PREDIKSI PENJUALAN ALAT TULIS KANTOR DENGAN METODE FP-GROWTH (STUDI KASUS TOKO KOPERASI SEKOLAH BINA MULIA)"
- [15] E. E. Barito, T. Beng, dan D. Arisandi, "Jurnal Ilmu Komputer dan Sistem Informasi," 2022.
- [16] Asahar. D., Yanosma. K., dan Anggriani. Johar, "IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) DAN SIMPLE ADDITIVE WEIGHTING (SAW) DALAM PENGAMBILAN KEPUTUSAN SELEKSI PENERIMAAN ANGGOTA PASKIBRAKA," 2016.
- [17] F. Handayani, D. Feddy, dan S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," 2015.
- [18] A. Mulyo Widodo, Y. Salsabila Anggraeni, N. Anwar, A. Ichwani, dan B. Anggara Sekti, "Performansi K-NN, J48, Naive Bayes dan Regresi Logistik Sebagai Algoritma Pengklasifikasi Diabetes," 2021.