

Pemodelan Prediktif Menggunakan Metode Ensemble Learning XGBoost dalam Peningkatan Akurasi Klasifikasi Penyakit Ginjal

R. Soelistijadi¹, Th. Dwiati Wismarini², Sri Eniyati³, Sunardi⁴

^{1,2,3}Program Studi Teknik Informatika, Universitas Stikubank, Semarang, Jawa Tengah, Indonesia

⁴Program Studi Teknologi Rekayasa Multimedia Grafis, Universitas Stikubank, Semarang, Jawa Tengah, Indonesia

E-mail: r.soelistijadi@edu.unisbank.ac.id¹, dwismarini@edu.unisbank.ac.id², eniyati@edu.unisbank.ac.id³, sunardi@edu.unisbank.ac.id⁴

Abstract

Chronic Kidney Disease (CKD) is a serious global health problem. However, information about how many people are affected by CKD in several countries is not very abundant and is sometimes not the same from one source to another. This research aims to increase accuracy in classifying CKD patients using the XGBoost ensemble learning method. The XGBoost model was drilled using the CKD dataset of 400 data records which were divided into training data and test data with a ratio of 70% used as training data and 30% as test data. Then an optimization technique is carried out, namely the parameter tuning process using a grid search method to find the best value using 5 parameters, namely *n_estimators*, *max_depth*, *learning_rate*, *Subsample*, *Colsample bytree*. The evaluation results using the confusion matrix, were obtained with an accuracy level of 99.16%, precision 98.17%, recall 99.16% and f1-score 99.16%. So the XGBoost algorithm implementing parameter tuning techniques is a good classification method that is good enough to be applied in CKD and Not CKD classification.

Keywords: CKD, Classification, XGBoost, Confusion Matrix

Abstrak

Penyakit Ginjal Kronis (PGK) merupakan masalah kesehatan global yang serius. Namun informasi tentang berapa banyak orang yang terkena PGK di beberapa negara tidak terlalu banyak dan kadang-kadang tidak sama antara satu sumber dengan sumber lainnya. Penelitian ini bertujuan untuk meningkatkan akurasi dalam mengklasifikasikan pasien PGK dengan menggunakan metode ensemble learning XGBoost. Model XGBoost dilatih menggunakan dataset PGK sebanyak 400 record data yang dibagi menjadi data latih dan data uji dengan perbandingan 70% digunakan sebagai data latih dan 30% sebagai data uji. Kemudian dilakukan teknik optimasi yaitu proses parameter tuning menggunakan metode pencarian grid untuk mencari nilai terbaik dengan menggunakan 5 parameter yaitu *n_estimators*, *max_depth*, *learning_rate*, *Subsample*, *Colsample bytree*. Dari hasil evaluasi menggunakan confusion matrix didapatkan hasil dengan tingkat accuration 99,16 %, precision 98,17 %, recall 99,16 % dan f1-score 99,16 %. Jadi algoritma XGBoost dengan mengimplementasikan teknik parameter tuning merupakan metode klasifikasi yang cukup baik untuk diterapkan pada kasus klasifikasi PGK dan Tidak PGK.

Kata kunci: PGK, Klasifikasi, XGBoost, Confusion Matrix

1. Pendahuluan

Seperti di negara-negara lain, banyak orang di Indonesia juga menderita penyakit ginjal kronis (PGK) yang serius. Namun informasi tentang berapa banyak orang yang terkena PGK di Indonesia tidak terlalu banyak dan kadang-kadang tidak sama antara satu

sumber dengan sumber lainnya. Sebuah penelitian yang dilakukan oleh Riset Kesehatan Dasar (Riskesdas) pada tahun 2018 mencatat bahwa sekitar 3,8 dari setiap 1000 orang di Indonesia menderita PGK, yang naik dari sekitar 2,0 dari setiap 1000 orang pada tahun 2013. Di Indonesia meskipun terdapat peningkatan dalam jumlah penderita PGK namun kesadaran akan PGK, diagnosa dini dan manajemen yang tepat masih merupakan tantangan besar. Oleh karena itu sangat diperlukan pengembangan metode yang lebih efektif dalam klasifikasi PGK [1].

Penelitian ini bertujuan untuk menggunakan metode ensemble learning XGBoost untuk memodelkan prediksi PGK dengan lebih akurat. Metode ini diharapkan dapat meningkatkan akurasi klasifikasi PGK dibandingkan dengan metode klasifikasi tradisional. Diharapkan dengan mengembangkan model prediktif yang lebih akurat maka diharapkan akan memungkinkan identifikasi dini PGK yang lebih tepat [2]. Model ensemble learning XGBoost merupakan salah satu tehnik yang akan digunakan dalam penelitian ini. Ensemble learning merupakan pendekatan yang menggabungkan beberapa model pembelajaran mesin untuk meningkatkan kinerja dan akurasi prediksi. Dalam konteks penelitian ini, XGBoost akan digunakan untuk memodelkan prediksi penyakit ginjal kronis (PGK) dengan lebih akurat [3]. Dengan kata lain, model ini dapat belajar dari berbagai algoritma dan kombinasi fitur untuk membuat prediksi yang lebih tepat tentang apakah seseorang memiliki PGK atau tidak [4].

Dengan mengembangkan model prediktif yang lebih canggih menggunakan XGBoost, diharapkan dapat memungkinkan identifikasi dini PGK dan intervensi yang lebih tepat. Artinya model ini diharapkan dapat membantu dokter dan profesional kesehatan untuk mendeteksi PGK pada tahap awal, sehingga dapat dilakukan tindakan pencegahan ataupun pengobatan yang lebih efektif untuk mencegah perkembangan penyakit dan meningkatkan prognosis pasien [5].

Manfaat dari penelitian ini adalah peningkatan kemampuan diagnosa dini PGK, yang pada gilirannya dapat mengarah pada perawatan yang lebih tepat waktu dan efektif bagi pasien PGK [6]. Lebih lanjut penelitian ini juga akan memberikan kontribusi pada pengembangan teknologi prediktif dalam bidang kesehatan, yang dapat meningkatkan kualitas hidup pasien PGK secara keseluruhan. Kerangka konseptual penelitian ini didasarkan pada penggunaan metode ensemble learning XGBoost dalam pemodelan prediktif untuk klasifikasi penyakit ginjal kronis. Diharapkan dengan pemrosesan data yang tepat dan pemilihan fitur yang relevan, diharapkan model prediktif yang dikembangkan dapat memberikan hasil yang lebih akurat [7].

Metode penelitian ini akan menggunakan pendekatan kuantitatif dengan mengumpulkan data klinis yang relevan. Data tersebut akan diproses dan dievaluasi menggunakan teknik ensemble learning XGBoost dan metrik evaluasi yang sesuai. Dalam hal ini ruang lingkup penelitian ini berfokus pada pengembangan dan evaluasi model prediktif untuk klasifikasi PGK, tanpa mempertimbangkan aspek-aspek lain dari manajemen PGK [8]. Dengan menggunakan metode ensemble learning XGBoost maka diharapkan dapat membuka jalan bagi perbaikan dalam manajemen penyakit ginjal kronis secara keseluruhan.

2. Metodologi Penelitian

Untuk memahami peran metode pembelajaran ensemble XGBoost dalam meningkatkan akurasi klasifikasi PGK maka dalam tinjauan pustaka ini akan diulas beberapa penelitian terkait metode klasifikasi dalam kesehatan, penggunaan XGBoost pada pemodelan prediktif, metode pembelajaran ensemble, dan Confusion Matrix.

a) Metode Klasifikasi dalam Kesehatan:

Metode klasifikasi adalah pendekatan yang digunakan untuk membedakan antara kategori-kategori tertentu dalam data. Klasifikasi artinya proses pembuatan model yg mengelompokkan objek berdasarkan atribut-atribut yg dimilikinya[9]. Dalam bidang

kesehatan, klasifikasi sering digunakan untuk memprediksi diagnosis penyakit berdasarkan fitur-fitur klinis. Metode klasifikasi tradisional, seperti regresi logistik dan pohon keputusan, telah digunakan secara luas dalam memprediksi suatu penyakit. Studi tentang klasifikasi PGK memberikan gambaran tentang penggunaan berbagai teknik klasifikasi dalam memprediksi PGK [10].

b) XGBoost dalam Pemodelan Prediktif

XGBoost atau Extreme Gradient Boosting adalah salah satu algoritma ensemble learning yang populer digunakan dalam pemodelan prediktif. XGBoost sendiri merupakan sebuah algoritma yang ditingkatkan dari gradient boosting decision tree yang efisien dalam membangun boosted trees [11]. Maulana dan Hadiana menjelaskan bahwa XGBoost memiliki performa yang baik dalam klasifikasi, bahkan mengungguli beberapa algoritma lain seperti Naïve Bayes dan Random Forest dalam konteks tertentu [12]. Selain itu Chen dan Guestrin juga menjelaskan bahwa XGBoost bekerja dengan mengkombinasikan prediksi dari beberapa model pohon keputusan yang lemah menjadi satu model yang kuat [13]. Keunggulan utama XGBoost meliputi efisiensi komputasi, kemampuan menangani data yang hilang, dan kemampuannya dalam mengurangi kesalahan bias dan varians. Penelitian menunjukkan bahwa XGBoost dapat menghasilkan model yang sangat akurat dan andal dalam berbagai aplikasi klasifikasi dan regresi, termasuk dalam deteksi suatu penyakit [7].

c) Ensemble Learning:

Ensemble learning merupakan pendekatan yang menggabungkan beberapa model pembelajaran mesin untuk meningkatkan kinerja prediksi. Metode ensemble learning telah terbukti efektif dalam berbagai aplikasi, termasuk klasifikasi medis. Studi tentang ensemble learning menjelaskan prinsip dasar dan keunggulan dari ensemble learning dalam menghasilkan model yang lebih akurat dan stabil [14].

d) Confusion Matrix :

Confusion matrix merupakan metode untuk mengukur kinerja model klasifikasi dengan membandingkan prediksi model dengan label sebenarnya. Matriks tersebut mewakili jumlah prediksi benar dan salah yang dibuat model untuk setiap kelas. Confusion matrix terdiri dari empat komponen utama yaitu : *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [15]. Dalam hal ini pengukuran kinerja model pada penelitian ini terdiri dari akurasi, presisi, recall, dan F1-score.

Accuracy sebagai indikator utamanya yaitu peningkatan akurasi prediksi model dalam identifikasi pasien PGK dan non-PGK. Sasaran akurasi ditetapkan berdasarkan standar klinis atau penelitian sebelumnya. Untuk menghitung akurasi digunakan persamaan berikut.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (1)$$

Precision untuk mengukur proporsi pasien yang diprediksi memiliki PGK yang diberikan oleh model adalah benar. Untuk menghitung presisi digunakan persamaan berikut.

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (2)$$

Recall untuk mengukur proporsi sejauh mana model dapat mengidentifikasi serta mengklasifikasi dengan benar semua kasus positif atau pasien PGK yang sebenarnya dengan persamaan berikut.

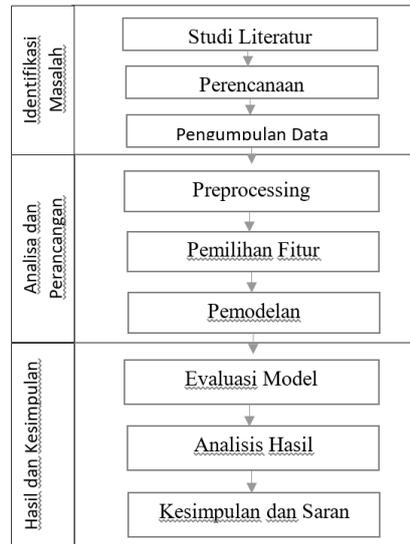
$$Recall = \frac{TP}{TP + FN} \times 100 \quad (3)$$

F1-Score merupakan harmonic mean dari presisi dan recall, yang memberikan gambaran keseluruhan tentang keseimbangan antara keduanya. Tujuannya yaitu

untuk meningkatkan F1-score sebagai ukuran komprehensif dari performa model dengan menggunakan persamaan berikut.

$$F1\text{-score} = \frac{2 * Precision * Recall}{Precision + Recall} \times 100 \quad (4)$$

Berikut tahapan penelitian model prediktif untuk mengklasifikasikan penyakit ginjal kronik (PGK) dengan metode pembelajaran ensemble XGBoost seperti terlihat pada Gambar 1.



Gambar 1. Tahapan penelitian

2.1. Identifikasi Masalah

Pada tahap ini peneliti melakukan tinjauan pustaka tentang PGK, metode klasifikasi dan penggunaan pembelajaran ensemble XGBoost dalam dunia kesehatan. Dalam hal ini peneliti melakukan analisis terhadap penelitian-penelitian sebelumnya untuk memahami pendekatan yang ada, kelemahan, serta peluang yang ada. Selanjutnya dilakukan identifikasi tujuan penelitian yang spesifik dan merumuskan rencana penelitian serta masalah penelitian yang ingin dipecahkan. Selain itu juga merancang kerangka penelitian konseptual yang mencakup penggunaan pembelajaran ensemble XGBoost dalam pemodelan prediktif untuk klasifikasi PGK. Langkah terakhir dalam identifikasi masalah ini yaitu melakukan pengumpulan data dengan cara mengidentifikasi sumber data klinis yang relevan, termasuk data pasien dengan riwayat PGK serta pemerolehan data klinis yang berasal dari data publik pada *web UCI*.

2.2. Analisa dan Perancangan

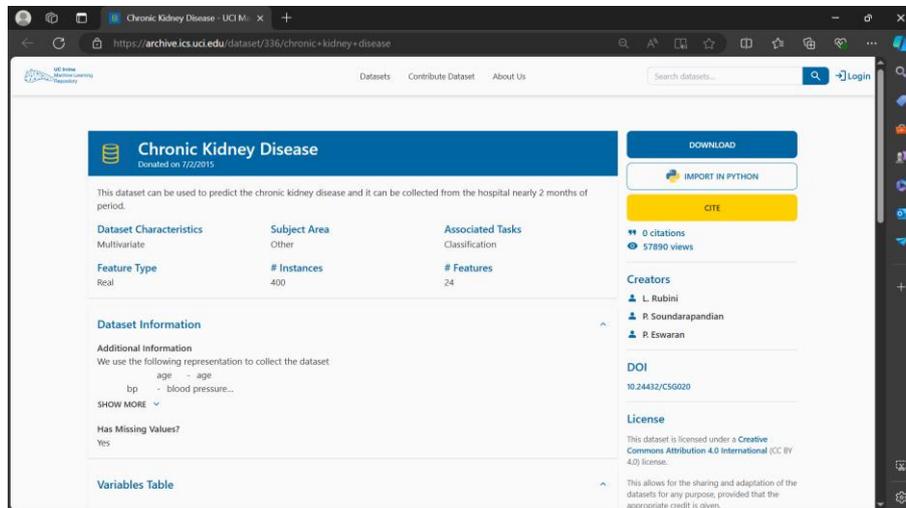
Tahap ini melibatkan preprocessing data dengan membersihkan nilai yang hilang (missing values) dan kemudian mengidentifikasi distribusi variabel, korelasi, dan pola yang relevan. Langkah selanjutnya dilakukan pemilihan fitur yaitu mengidentifikasi fitur paling relevan untuk prediksi PGK berdasarkan literatur dan analisis data. Seleksi menggunakan teknik seperti analisis variable penting (feature importance) untuk memilih fitur yang paling informatif. Langkah terakhir yaitu pemodelan prediktif untuk menggunakan data yang telah diolah untuk mengimplementasikan model pembelajaran ensemble XGBoost untuk klasifikasi PGK. Dalam hal ini, berbagai konfigurasi model seperti tuning parameter dan cross-validation untuk meningkatkan kinerja model.

2.3. Hasil dan Kesimpulan

Pada tahap terakhir, performa model dievaluasi menggunakan metrik yang sesuai seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Dalam hal ini, model divalidasi dengan menggunakan data uji terpisah untuk memastikan generalisasi yang baik. Kemudian dari analisis hasil diinterpretasikan hasil evaluasi untuk memahami kelebihan dan kelemahan model. Selanjutnya, diidentifikasi faktor-faktor yang berkontribusi terhadap kinerja model dan implikasinya terhadap praktik klinis. Terakhir, implikasi hasil terhadap diagnosis dini dan manajemen pasien dituangkan dalam bentuk kesimpulan dan saran, termasuk saran untuk penelitian atau pengembangan lebih lanjut di bidang ini.

3. Hasil Dan Pembahasan

Penelitian ini menggunakan dataset *Chronic Kidney Disease Database* (Penyakit Ginjal Kronis) dari UCI (University California Irving) yang dapat diakses secara online melalui website <https://archive.ics.uci.edu/datasets> seperti yang terlihat pada Gambar 2 dibawah ini.



Gambar 2. Data *Chronic Kidney Disease Database*

Data yang diambil dari website tersebut berformat CSV dan memiliki record dengan isi 400 baris seperti terlihat pada Gambar 3 di bawah ini.

id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wbcc	rbcc	htn	dm	cad	appet	pe	ane	classification	
0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	...	44	7800	5.2	yes	yes	no	good	no	no	no	no	no	no	no	ckd	
1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	...	38	6000	NaN	no	no	no	good	no	no	no	no	no	no	no	ckd	
2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	...	31	7500	NaN	no	yes	no	poor	no	yes	no	poor	no	yes	yes	ckd	
3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	...	32	6700	3.9	yes	no	no	poor	yes	yes	yes	yes	yes	yes	yes	ckd	
4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	...	35	7300	4.6	no	no	no	good	no	no	no	no	no	no	no	ckd	
...
395	395	55.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	...	47	6700	4.9	no	no	no	good	no	no	no	no	no	no	no	notckd
396	396	42.0	70.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	54	7800	6.2	no	no	no	good	no	no	no	no	no	no	no	notckd
397	397	12.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	...	49	6600	5.4	no	no	no	good	no	no	no	no	no	no	no	notckd
398	398	17.0	60.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	51	7200	5.9	no	no	no	good	no	no	no	no	no	no	no	notckd
399	399	58.0	80.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	53	6800	6.1	no	no	no	good	no	no	no	no	no	no	no	notckd

Gambar 3. Data *Chronic Kidney Disease Database*

Pada Gambar 3 *Dataset* tersebut terdapat variabel independen dan dependen. Variabel Independen terdiri dari 24 variabel yaitu *Age*, *Bp*, *Sg*, *Al*, *Su*, *Rbc*, *Pc*, *Pcc*, *Ba*, *Bgr*, *Bu*, *Sc*, *Sod*, *Pot*, *Hemo*, *Pcv*, *Wbcc*, *Rbcc*, *Htn*, *Dm*, *Cad*, *Appet*, *Pe*, *Ane* dan 1 variabel

Variabel Dependen terdiri dari 1 variabel yang berisi kan *Outcome (Class)*. Berikut tampilan dari Tabel 1 dan tabel 2 yang merupakan keterangan dari masing-masing Variabel Independen (*X*) dan Variabel Dependen (*Y*).

Tabel 1. Variabel Independen (*X*)

Kolom Dataset	Deskripsi	Kolom Dataset	Deskripsi
<i>Age</i>	Usia	<i>Sod (sodium)</i>	Sodium
<i>Bp (blood pressure)</i>	Tekanan darah	<i>Pot (potassium)</i>	Potassium
<i>Sg (specific gravity)</i>	Gravitasi Spesifik	<i>Hemo (hemoglobin)</i>	Hemoglobin
<i>Al (albumin)</i>	Albumin	<i>Pcv (packed cell volume)</i>	Kemasan Volume Sel
<i>Su (sugar)</i>	Gula	<i>Wbcc (white blood cell count)</i>	Jumlah sel darah putih
<i>Rbc (red blood cells)</i>	Darah merah	<i>Rbcc (red blood cell count)</i>	Jumlah sel darah merah
<i>Pc (pus cell)</i>	Sel Nanah	<i>Htn (hypertension)</i>	Hipertensi
<i>Pcc (pus cell clumps)</i>	Gumpalan sel nanah	<i>Dm (diabetes mellitus)</i>	Diabetes Mellitus
<i>Ba (bacteria)</i>	Bakteri	<i>Cad (coronary artery disease)</i>	Penyakit Arteri Koroner
<i>Bgr (blood glucose random)</i>	Glukosa Darah Acak	<i>Appet (appetite)</i>	Nafsu makan
<i>Bu (blood urea)</i>	Urea Darah	<i>Pe (pedal edema)</i>	Pedal Edema
<i>Sc (serum creatinine)</i>	Kreatinin Serum	<i>Ane (anemia)</i>	Anemia

Tabel 2. Variabel Dependen (*Y*)

Kolom Dataset	Deskripsi
<i>Outcome</i>	Ckd atau Tidak Ckd

Pada langkah selanjutnya dilakukan preprocessing terhadap data agar model dapat menggunakan data dengan performa yang maksimal. Hal ini mengingat dari Gambar 3 sebelumnya terlihat ada beberapa kolom yang perlu diubah menjadi tipe numerik karena masih bertipe string. Oleh karena itu langkah selanjutnya yaitu mengubah kolom target ke format numerik agar dapat diproses dengan menggunakan perintah `LabelEncoder()` seperti terlihat pada Gambar 4.



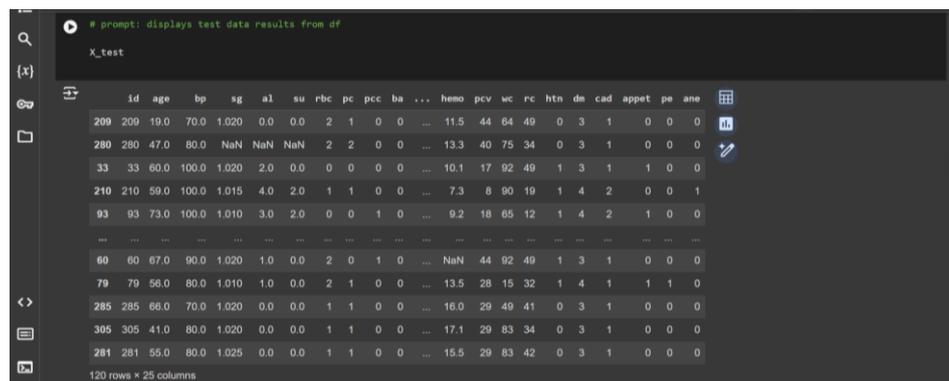
Gambar 4. Tampilan konversi kolom target ke format numerik menggunakan google colab.

3.1. Penerapan Model XGBoost

Setelah mengonversi kumpulan data dari kolom target ke format numerik maka dilanjutkan dengan menerapkan model XGBoost. Untuk itu sebelumnya data dipecah atau dipartisi menjadi dua bagian yaitu data latih dan data uji. Dalam penelitian ini digunakan split data dengan rasio 70:30 yaitu dengan membagi data menjadi 280 baris (70%) untuk data latih dan 120 baris (30%) untuk data uji, seperti yang ditunjukkan pada Gambar 5 dan Gambar 6.



Gambar 5. Data latih



Gambar 6. Data uji

Selanjutnya masuk ke tahap tuning parameter yang dalam penelitian ini menggunakan teknik pencarian grid untuk menyetel parameter algoritma XGBoost. Hasil yang diperoleh ditunjukkan pada Tabel 3 di bawah ini.

Tabel 3. Grid Search Parameter Tuning

	Grid Search Value	Nilai parameter terbaik
<i>n_estimators</i>	50, 100, 200	100
<i>max_depth</i>	3, 5, 7	3
<i>learning_rate</i>	0.1, 0.01, 0.001	0.01
<i>Subsample</i>	0.8, 0.9, 1.0	0.8
<i>Colsample bytree</i>	0.8, 0.9, 1.0	0.8

Hasil yang ditunjukkan pada Tabel 3 menunjukkan parameter optimal dan nilai parameter yang dapat meningkatkan kinerja algoritma pada percobaan yang dilakukan. Selanjutnya nilai parameter tersebut kemudian diterapkan pada proses pelatihan model XGBoost dengan menggunakan semua data latih.

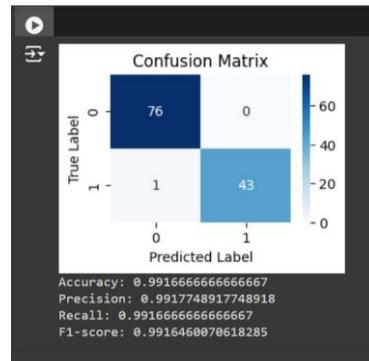
3.2. Evaluasi

Evaluasi penelitian merupakan proses menguji seberapa jauh model yang digunakan menghasilkan hasil yang akurat. Untuk menguji keakuratannya maka pada penelitian ini digunakan confusion matrix untuk skema pengujian dengan rasio data latih dan uji sebesar 70% : 30% dengan menggunakan *parameter tuning*. Berikut data nilai *confusion matrix* dari keseluruhan percobaan yang telah dilakukan.

Tabel 4. Nilai *confusion matrix*

True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)
76	0	1	43

Tabel 5 menunjukkan hasil confusion matrix dengan nilai *True Positive* sebesar 76, *False Positive* sebesar 0, *False Negative* sebesar 1, dan *True Negative* sebesar 43. Nilai-nilai tersebut dapat dihitung dengan menggunakan Persamaan 1, 2, 3, dan 4 untuk mendapatkan *Accuracy*, *Presicion*, *Recall*, dan *F1-score*. Untuk lebih jelasnya hasilnya berupa visualisasi yang dapat dilihat pada Gambar 7.



Gambar 7. Visualisasi Confusion Matrix

Berdasarkan Gambar 7 maka hasil dari evaluasi yang dilakukan menggunakan confusion matrix dapat dilihat pada tabel 5 berikut ini .

Tabel 5. Matrix evaluasi

Matrix evaluasi	Skor
Accuracy	0,9916666666666667
Precision	0,9917748917748918
Recall	0,9916666666666667
F1-score	0,99164600070618285

Dari Tabel 5 maka perhitungan menggunakan persamaan 1, 2, 3, dan 4 akan menghasilkan skor *Accuracy* sebesar 99,16%, *Precision* sebesar 98,17%, *Recall* sebesar 99,16, dan *F1-score* sebesar 99,16%.

4. Kesimpulan

Penelitian dilakukan dengan menggunakan algoritma XGBoost dengan menggunakan data set Penyakit Ginjal Kronis (PGK) sebanyak 400 record data. Dataset ini dibagi menjadi data latih dan data uji dengan perbandingan 70:30, dengan 70% digunakan sebagai data latih dan 30% sebagai data uji. Kemudian diterapkan teknik optimasi yaitu proses parameter tuning menggunakan 5 parameter dengan menggunakan metode pencarian grid untuk mencari nilai terbaik. Ke 5 parameter tersebut yaitu *n_estimators*, *max_depth*, *learning_rate*, *Subsample*, *Colsample bytree*. Dari hasil evaluasi menggunakan confusion matrix didapatkan hasil dengan tingkat akurasi 99,16 %, precision 98,17 %, recall 99,16 % dan f1-score 99,16 %. Jadi algoritma XGBoost dengan mengimplementasikan teknik parameter tuning merupakan metode klasifikasi yang cukup baik untuk diterapkan pada kasus klasifikasi PGK dan Tidak PGK. Selanjutnya saran dalam penelitian ini penggunaan dataset dianjurkan untuk menggunakan record data yang lebih banyak. Hal ini bertujuan untuk menghindari kasus over fitting pada Confusion Matrix.

Daftar Pustaka

- [1] Kementerian Kesehatan Ri, “Riskedas 2018”, Laporan, Riskedas 2018 Vol.44 No.8 Hal 181-222, 2018 [Online]. Available: [Http://Www.Yankes.Kemkes.Go.Id/Assets/Downloads/Pmk No. 57 Tahun 2013 Tentang Ptrm.Pdf](http://www.yankes.kemkes.go.id/assets/downloads/pmk_no_57_tahun_2013_tentang_ptrm.pdf).
- [2] M. M. El Sherbiny, E. Abdelhalim, H. El-Din Mostafa, Dan M. M. El-Seddik, “Classification Of Chronic Kidney Disease Based On Machine Learning Techniques,” *Indones. J. Electr. Eng. Comput. Sci.*, Vol. 32, No. 2, Hal. 945–955, 2023, Doi: 10.11591/Ijeecs.V32.I2.Pp945-955.
- [3] Y. Kale, S. Rathkanthiwar, P. Fulzele, Dan N. J. Bankar, “Xgboost Learning For Detection And Forecasting Of Chronic Kidney Disease (Ckd),” *Int. J. Intell. Syst. Appl. Eng.*, Vol. 12, No. 17s, Hal. 137–150, 2024.
- [4] A. Ogunleye Dan Q. G. Wang, “Xgboost Model For Chronic Kidney Disease Diagnosis,” *Ieee/Acm Trans. Comput. Biol. Bioinforma.*, Vol. 17, No. 6, Hal. 2131–2140, 2020, Doi: 10.1109/Tcbb.2019.2911071.
- [5] S. M. Ganie, P. K. D. Pramanik, S. Mallik, Dan Z. Zhao, “Chronic Kidney Disease Prediction Using Boosting Techniques Based On Clinical Parameters,” *Plos One*, Vol. 18, No. 12 December, Hal. 1–21, 2023, Doi: 10.1371/Journal.Pone.0295234.
- [6] S. K. Ghosh Dan A. H. Khandoker, “Investigation On Explainable Machine Learning Models To Predict Chronic Kidney Diseases,” *Sci. Rep.*, Vol. 14, No. 1, Hal. 1–15, 2024, Doi: 10.1038/S41598-024-54375-4.
- [7] Z. Salam Patrous, “Evaluating Xgboost For User Classification By Using Behavioral Features Extracted From Smartphone Sensors,” 2018.
- [8] F. Nateghi Haredasht, L. Viaene, H. Pottel, W. De Corte, Dan C. Vens, “Predicting Outcomes Of Acute Kidney Injury In Critically Ill Patients Using Machine Learning,” *Sci. Rep.*, Vol. 13, No. 1, Hal. 1–13, 2023, Doi: 10.1038/S41598-023-36782-1.
- [9] E. Susilowati, M. Kania Sabariah, And A. Akbar Gozali, “Implementasi Metode Support Vector Machine Untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter.” [Online]. Available: [Https://Www.Academia.Edu/33108996/Implementasi_Metode_Support_Vector_Machine_Untuk_Melakukan_Klasifikasi_Kemacetan_Lalu_Lintas_Pada_Twitter_Implementation_Support_Vector_Machine_Method_For_Traffic_Jam_Classificati on_On_Twitter](https://www.academia.edu/33108996/Implementasi_Metode_Support_Vector_Machine_Untuk_Melakukan_Klasifikasi_Kemacetan_Lalu_Lintas_Pada_Twitter_Implementation_Support_Vector_Machine_Method_For_Traffic_Jam_Classification_On_Twitter).
- [10] E. Listiana, R. Muzayanah, M. A. Muslim, Dan E. Sugiharti, “Optimization Of Support Vector Machine Using Information Gain And Adaboost To Improve Accuracy Of Chronic Kidney Disease Diagnosis,” *J. Soft Comput. Explor.*, Vol. 4, No. 3, Hal. 152–158, 2023 [Online]. Available: [Https://Shmpublisher.Com/Index.Php/JosceX/Article/View/218](https://shmpublisher.com/index.php/josceX/article/view/218).
- [11] T. Chen And C. Guestrin, “Xgboost: A Scalable Tree Boosting System,” In *Proceedings Of The Acm Sigkdd International Conference On Knowledge Discovery And Data Mining*, Association For Computing Machinery, Aug. 2016, Pp. 785–794. Doi: 10.1145/2939672.2939785.
- [12] M.D. Maulana, A.I. Hadiana, F.R. Umbara, 2023, “Algoritma Xgboost Untuk Klasifikasi Kualitas Air Minum”, *Jati (Jurnal Mahasiswa Teknik Informatika)*, Vol. 7 No. 5, Oktober 2023.
- [13] M. Syukron, R. Santoso, And T. Widiharih, “Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data”, [Online]. Available: [Https://Ejournal3.Undip.Ac.Id/Index.Php/Gaussian/](https://ejournal3.undip.ac.id/index.php/gaussian/).

- [14] Linggar Maretva Cendani, Adi Wibowo, 2022, Perbandingan Metode Ensemble Learning Pada Klasifikasi Penyakit Diabetes Jurnal Masyarakat Informatika (Jmasif) [Online]. Available: <https://ejournal.undip.ac.id/index.php/jmasif/article/view/42912>.
- [15] M. R.A Masud Dan M.R.H. Mondal, "Data-Driven Diagnosis Of Spinal Abnormalities Using Feature Selection And Machine Learning Algorithms," *Plos One*, Vol. 15, No. 2, Feb. 2020, [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0228422>.