

Perbandingan Kinerja Algoritma Naive Bayes, Support Vector Machine dan Random forest untuk Prediksi Penyakit Ginjal Kronis

Imaniar Ikko Mulya Rizky^{1a*}, Suhendro Yusuf Irianto^{2b}, Sriyanto^{3c}

^{abc}Program Studi Magister Teknik Informatika, Falkutas Ilmu Komputer, Institut Informatika dan Bisnis Darmajaya, Jl. ZA. Pagar Alam No.93, Gedong Meneng, Kec. Rajabasa, Kota Bandar Lampung, Lampung 35141, Indonesia.

*e-mail: imaniarikko.2221210046@mail.darmajaya.ac.id

suhendro@darmajaya.ac.id

sriyanto@darmajaya.ac.id

Abstract

Chronic kidney disease (CKD) is a serious condition that affects human health worldwide, leading to a gradual decline in kidney function and triggering severe complications, including kidney failure. Early prediction of CKD can provide significant benefits in preventing serious complications and enabling timely interventions. In recent decades, the use of algorithm-based predictive methods in data science has shown great potential in the field of healthcare, including disease prediction. This research aims to compare the performance of three algorithms, namely Naive Bayes, SVM, and Random Forest. The test results show that, with a data split of 70% for training and 30% for testing, the classification processing using the Naive Bayes algorithm achieved an accuracy of 97.14%, while the classification processing using the Support Vector Machine algorithm achieved 92.50%, and the Random Forest algorithm achieved 99.64%. This study demonstrates that the Random Forest algorithm performs the best in predicting Chronic Kidney Disease.

Keywords : *Chronic kidney disease; Naive Bayes; Support Vector Machine; Random forest; Predictions*

Abstrak

Penyakit ginjal kronis (PGK) merupakan kondisi serius yang memengaruhi kesehatan manusia di seluruh dunia dan dapat menyebabkan penurunan fungsi ginjal secara perlahan dan memicu komplikasi serius, termasuk gagal ginjal. Prediksi dini PGK dapat memberikan manfaat besar dalam mencegah komplikasi serius dan memungkinkan intervensi yang tepat waktu. Dalam beberapa dekade terakhir, penggunaan metode prediktif berbasis algoritma dalam ilmu data telah menunjukkan potensi besar dalam bidang kesehatan, termasuk prediksi penyakit. Penelitian ini bertujuan membandingkan kinerja tiga algoritma, yaitu Naive Bayes, SVM, dan Random Forest. Hasil pengujian menunjukkan bahwa pada split data (data training 0.7% dan data testing 0.3%), pemrosesan klasifikasi menggunakan algoritma Naive Bayes sebesar 97.14%, sedangkan pemrosesan klasifikasi menggunakan algoritma Support Vector Machine sebesar 92.50% dan untuk pemrosesan klasifikasi menggunakan algoritma Random Forest sebesar 99.64%. Penelitian ini menunjukkan bahwa algoritma Random Forest memiliki kinerja terbaik dalam memprediksi dalam Penyakit Ginjal Kronis.

Kata Kunci: *Naive Bayes; Support Vector Machine; Random forest; Prediksi*

1. PENDAHULUAN

Ginjal merupakan organ yang sangat penting dalam menjaga komposisi darah, mencegah penumpukan limbah atau kotoran, mengatur keseimbangan cairan dalam tubuh, menjaga tingkat elektrolit seperti sodium, potasium, dan fosfat tetap stabil, serta memproduksi hormon dan enzim yang membantu dalam mengendalikan tekanan darah, pembentukan sel darah merah, serta menjaga kesehatan dan kekuatan tulang (Amalia, 2018). Penyakit ginjal kronis (PGK) merupakan kondisi serius yang memengaruhi kesehatan manusia di seluruh dunia yang dapat menyebabkan penurunan fungsi ginjal secara perlahan dan memicu komplikasi serius, termasuk gagal ginjal (Mahardika Pratama et al., 2022).

Pertumbuhan penduduk yang pesat telah menyebabkan peningkatan jumlah kasus penyakit CKD. Menurut Global Burden of Disease, pada tahun 1990 penyakit CKD menempati peringkat ke-27, namun pada tahun 2010 naik menjadi peringkat ke-18 (Fadilla et al., 2018). Berdasarkan data Kementerian Kesehatan RI, pada tahun 2013, terdapat 499.800 orang atau 2 dari setiap 1.000 orang di Indonesia yang menderita penyakit ginjal kronis. Prevalensi penyakit ini

cenderung meningkat seiring bertambahnya usia (A'yuniyah et al., 2022). Hal – hal tersebut menunjukkan bahwa PGK merupakan kondisi yang membutuhkan perhatian serius, termasuk penanganan yang cepat melalui sistem prediksi yang akurat (Darwanto et al., 2021).

Prediksi dini PGK dapat memberikan manfaat besar dalam mencegah komplikasi serius dan memungkinkan intervensi yang tepat waktu. Dalam beberapa dekade terakhir, penggunaan metode prediktif berbasis algoritma dalam ilmu data telah menunjukkan potensi besar dalam bidang kesehatan, termasuk prediksi penyakit (Moreira & Namen, 2018). Dalam prediksi penyakit ginjal kronis (PGK), perbandingan kinerja algoritma menjadi sangat penting untuk memilih pendekatan terbaik. Tiga algoritma yang sering digunakan, yaitu Naive Bayes, Support Vector Machine (SVM), dan Random Forest, memiliki kriteria utama dalam pemilihan algoritma yang efektif, yaitu akurasi, kecepatan komputasi, dan kemampuan mengatasi masalah kompleks (Chou et al., 2023). Penelitian ini bertujuan membandingkan kinerja tiga algoritma, yaitu Naive Bayes, SVM, dan Random Forest, dalam memprediksi penyakit ginjal kronis. Untuk itu, kami akan menggunakan dataset yang mencakup atribut klinis dan historis yang relevan dengan PGK.

2. KERANGKA TEORI

2.1 Penyakit Ginjal Kronis (PGK)

Ginjal merupakan sepasang organ tubuh yang memiliki peran penting. Setiap ginjal memiliki ukuran sekitar 10-15 cm, berat sekitar 150 gram, dan terletak di sisi kanan dan kiri tubuh pada bagian belakang (Widayati, 2019).

Ginjal melakukan banyak fungsi vital tubuh (Widayati, 2019):

- a. Menyaring darah dan hasil metabolisme tubuh dengan membuang kelebihan cairan.
- b. Mengatur keseimbangan cairan dan elektrolit dalam seperti natrium dan kalium.
- c. Mengatur tekanan darah dan merangsang pembentukan sel darah merah.

Penyakit ginjal adalah kelainan yang mengenai organ ginjal yang timbul akibat berbagai faktor, biasanya timbul secara perlahan dan sifatnya menahun (Lisa Lolowang et al., 2021) dan ditandai dengan penurunan kemampuan ginjal dalam menyaring darah (Laju Filtrasi Glomerulus/LFG). Pada awalnya tidak ditemukan gejala yang khas sehingga penyakit ini sering terlambat diketahui sampai fungsi ginjal tersisa kurang dari 15% (Widayati, 2019).

Tanda dan gejala awal penyakit ginjal tidak spesifik. Mengetahui gejala awal ini sangat penting untuk mendapatkan penanganan yang tepat dan memperlambat perkembangan penyakit ginjal, jika tidak bisa dihentikan sepenuhnya (Kyneissia Gliselda, 2021).

Berikut adalah 10 tanda dan gejala awal yang mungkin muncul jika seseorang mengalami penyakit ginjal kronis (PGK) (Widayati, 2019):

- a. Perubahan pada pola buang air kecil .
Gangguan pada ginjal akan berpengaruh pada pola buang air kecil Anda seperti :
 1. Sering terbangun di malam hari untuk buang air kecil.
 2. Air seni berbuih dan bergelembung.
 3. Anda mungkin, lebih sering buang air kecil atau jumlah lebih banyak dari biasanya dan warna air seni lebih jernih.
 4. Atau Anda mungkin, lebih jarang buang air kecil atau jumlah lebih sedikit dari biasanya dan warna air seni lebih gelap.
 5. Air seni berdarah
 6. Merasakan nyeri saat dan kesulitan pada saat buang air kecil
- b. Pembengkakan pada setiap bagian tubuh.
- c. Lebih mudah terasa lelah.
- d. Gatal dan ruam pada kulit.
- e. Mual dan muntah.
- f. Sesak nafas atau nafas tersengal – sengal.
- g. Pusing dan sulit berkonsentrasi.
- h. Nyeri pada pinggang bagian bawah.
- i. Meriang.
- j. Bau mulut tidak sedap.

2.2 Data Mining

Data mining adalah proses pengambilan informasi data dalam jumlah besar dan kompleks yang sebelumnya tidak diketahui (Data et al., 2022). Informasi penting diperoleh melalui proses kompleks yang melibatkan kecerdasan buatan, teknik statistik, matematika, dan machine learning. Teknik-teknik ini digunakan untuk mengidentifikasi dan

mengeksktraksi informasi berharga dari database besar. Data mining adalah disiplin ilmu yang berkaitan dengan kecerdasan buatan dan rekayasa pengetahuan (Sudarsono et al., 2021).

Pertumbuhan berkelanjutan dalam penambangan data dan penemuan pengetahuan didorong oleh beberapa penemuan (Dany Prasetya et al., 2022):

- a. Pertumbuhan eksplosif dalam pengumpulan data sebagaimana dibuktikan oleh pemindaian supermarket.
- b. Menyimpan data di gudang data sehingga seluruh perusahaan memiliki akses ke stok data yang up-to-date dan otoritatif.
- c. Ketersediaan akses yang ditingkatkan ke data dari navigasi web dan intranet.
- d. Tekanan persaingan untuk meningkatkan pangsa pasar dalam perekonomian global.
- e. Pengembangan paket perangkat lunak penambangan data komersial siap pakai.
- f. Peningkatan daya komputasi dan kapasitas penyimpanan yang signifikan.

Hal terpenting dalam teknik data mining adalah aturan untuk menemukan pola frekuensi tinggi antar himpunan itemset yang disebut fungsi Association Rules (Aturan Asosiasi). Belakangan ini, data mining telah diterapkan dalam berbagai bidang, termasuk bisnis, pendidikan, dan telekomunikasi.

Tahapan – tahapan Data Mining (Sudarsono et al., 2021):

- a. Pembersihan data.
- b. Integrasi data.
- c. Transformasi data.
- d. Aplikasi data Teknik Mining.
- e. Evaluasi pola yang di temukan.
- f. Presentasi pola yang ditemukan untuk menghasilkan aksi.

2.3 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang dapat menjelaskan dan membedakan konsep atau kelas data. Tujuan klasifikasi adalah memperkirakan kelas suatu objek yang tidak memiliki label. Model yang digunakan dapat berupa aturan jika-maka (if-then), pohon keputusan (decision tree), atau jaringan saraf tiruan (neural network) (Triwidianti, 2021). Klasifikasi juga bisa didefinisikan salah satu teknik prediksi dalam data mining yang digunakan untuk membuat prediksi tentang nilai data berdasarkan informasi yang diketahui dari kumpulan data yang berbeda (T. Arifin & Ariesta, 2019). Salah satu masalah yang sering dihadapi dalam algoritma klasifikasi adalah penurunan akurasi saat menghadapi data yang tidak seimbang, di mana distribusi sampel antar kelas sangat tidak proporsional (Triwidianti, 2021).

Contoh tugas klasifikasi dalam bisnis dan penelitian meliputi (Dany Prasetya et al., 2022):

- a. Menentukan apakah transaksi kartu kredit tertentu adalah penipuan.
- b. Penempatan mahasiswa baru pada jalur khusus yang berkaitan dengan kebutuhan khusus.
- c. Mengevaluasi apakah aplikasi hipotek menimbulkan risiko kredit.
- d. Diagnosis adanya penyakit tertentu.
- e. Menentukan apakah wasiat itu ditulis oleh almarhum sendiri atau dipalsukan oleh orang lain.
- f. Menentukan Apakah Perilaku Keuangan atau Pribadi Tertentu Mengindikasikan Potensi Ancaman Teroris.

2.4 Naive Bayes Classifier

Naïve Bayes Classifier (NBC) adalah metode klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas dalam suatu klasifikasi (Apriliah et al., 2021). Dalam definisi lain, metode Naïve Bayes Classifier adalah metode prediksi sederhana berbasis probabilitas yang menggunakan teorema Bayes dengan asumsi independensi yang kuat antara fitur-fitur yang diamati (Moreira & Namen, 2018). Kelebihan dari Naïve Bayes Classifier adalah dapat digunakan dan dianggap cukup handal untuk mengolah data dan proses klasifikasi yang tidak dapat diselesaikan oleh metode SVM dalam permasalahan jumlah data set yang besar dengan fitur yang beragam (Chakraborty et al., 2022).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan:

$P(H|X)$: Probabilitas akhir bersyarat (conditional probability) suatu hipotesis H terjadi jika diberikan bukti (evidence) X terjadi.

$P(X|H)$: Probabilitas sebuah bukti X terjadi akan mempengaruhi hipotesis H.

$P(H)$: Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

$P(X)$: Probabilitas awal (priori) bukti X terjadi tanpa memandang hipotesis atau bukti yang lain.

2.5 Support Vector Machine

Support Vector Machine (SVM) adalah salah satu metode terbaik yang digunakan dalam masalah klasifikasi (Pratama et al., 2018). SVM berawal dari permasalahan klasifikasi dua kelas yang membutuhkan training set positif dan negatif. Tujuan SVM adalah menemukan hyperplane terbaik yang memisahkan kedua kelas dan memaksimalkan margin antara kelas-kelas tersebut. Dalam beberapa kasus, metode SVM linier tidak cukup untuk mengklasifikasikan data, maka fungsi kernel dikembangkan untuk mengklasifikasikan data nonlinier.

Untuk melakukan klasifikasi dengan algoritma Support Vector Machine (SVM) terdapat beberapa tahapan diantaranya yaitu (N. Arifin et al., 2021):

- Menentukan hyperplane atau garis pembatas antara dua support vector.
- Menentukan margin atau garis jarak antara support vector dan hyperplane.
- Pemetaan support vector ke dalam suatu kelas dalam class dimensi yang sama.

Pada penelitian menggunakan kernel Polynomial Degree. Sequential Training memiliki algoritme yang lebih sederhana dan waktu yang diperlukan lebih cepat. Adapun algoritme Sequential Training adalah sebagai berikut (Pratama et al., 2018):

- Menginisialisasi $\alpha_i = 0$ dan parameter lain, misalnya $\lambda = 0,5$, $\gamma = 0,01$, $c = 1$, IterasiMax = 100, dan $\varepsilon = 0,001$.

Kemudian menghitung matriks Hessian dapat dihitung dengan rumus persamaan (1).

$$D_{ij} = y_i y_j (K(x_i, x_j) + \alpha^2) \quad (1)$$

- Mulai dari data ke i sampai j, hitung menggunakan persamaan (2).

$$\begin{aligned} E_i &= \sum_{j=1}^n a_j D_{ij} \\ \delta\alpha_i &= \min \{ \max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i \} \\ \alpha_i &= \alpha_i + \delta\alpha_i \end{aligned} \quad (2)$$

- Langkah 2 dilakukan terus-menerus hingga kondisi iterasi maksimum tercapai atau max

$$(|\delta\alpha_i|) < \varepsilon$$

Selanjutnya didapatkan nilai support vector (SV), $SV = (\text{Threshold SV})$. Nilai Threshold SV didapatkan dari beberapa percobaan, biasanya digunakan threshold > 0 . Kemudian dilakukan proses testing untuk mendapatkan keputusan di mana fungsi keputusan dapat di hitung dengan persamaan (3).

$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (3)$$

di mana nilai b dengan persamaan (4).

$$-\frac{1}{2} \left[\sum_{i=1}^m \alpha_i y_i K(x_i, x^+) + \sum_{i=1}^m \alpha_i y_i K(x_i, x^-) \right] \quad (4)$$

2.6 Random forest

Algoritma berbasis tree adalah metode pembelajaran dalam machine learning yang digunakan untuk memecahkan masalah yang membutuhkan data yang cukup besar. Sekian dari banyaknya Algoritma berbasis tree ini adalah Algoritma Random Forest. Algoritma ini diusulkan oleh Tin Kam Ho pada tahun 1995, yang merupakan kombinasi dari x tree yang digabung dan dijadikan satu model (Labib et al., 2023). Random forest merupakan gabungan dari

masing-masing tree yang membentuk hutan (forest) dengan melakukan tahapan training data yang dimiliki dan mendapatkan hasil akhir yang berbentuk voting. Metode Random Forest juga untuk meningkatkan nilai akurasi. Metode ini bertujuan untuk membangun pohon keputusan yang terdiri dari root node, internal node, dan leaf node dengan mengambil atribut dan data secara acak sesuai ketentuan. Root node adalah simpul yang terletak di atas (akar dari pohon keputusan. Sedangkan internal node merupakan simpul percabangan, node ini mempunyai satu input dan minimal dua output. Leaf node/terminal node yaitu simpul terakhir yang tidak memiliki output dan hanya memiliki satu input (Nalatissifa et al., 2021). Perhitungan nilai entropy menggunakan rumus pada persamaan 1, dan nilai information gain pada rumus persamaan 2.

$$Entropy(\gamma) = - \sum_i p(c|\gamma) \log_2 p(c|\gamma) \quad (1)$$

Keterangan:

Y merupakan himpunan kasus

$p(c|Y)$ adalah proporsi nilai Y pada kelas c.

$$Information\ Gain(Y, a) = Entropy(\gamma) - \sum_{v \in Values(a)} \frac{|Y_v|}{|Y|} Entropy(Y_v) \quad (2)$$

Keterangan:

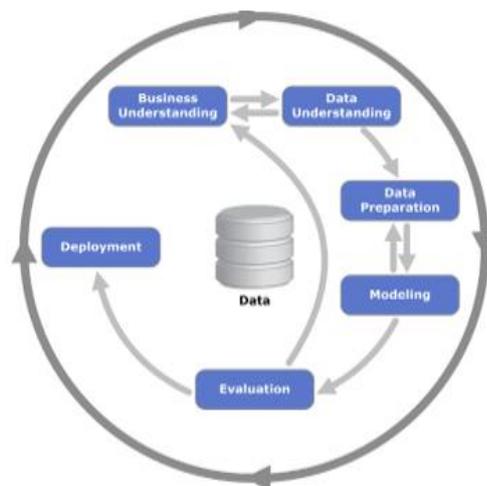
Values (a) adalah semua nilai yang mungkin pada himpunan kasus a.

Y_v merupakan subkelas dari Y dengan kelas v yang berkaitan dengan kelas a.

Y_a merupakan semua nilai yang sama dengan a.

2.7 Model Cross Standard Industry for Data Mining (CRISP-DM)

Pada penelitian ini, metode penelitian yang digunakan adalah model Cross Standard Industry for Data Mining (CRISP-DM). CRISP-DM memiliki enam fase siklus hidup proyek data mining (Nalatissifa et al., 2021). Adapun gambaran fase pada model CRISP-DM adalah sebagai berikut (Mahardika Pratama et al., 2022):



Gambar 1. Diagram CRISP-DM

Penjelasan mengenai fase pada model CRISP-DM adalah:

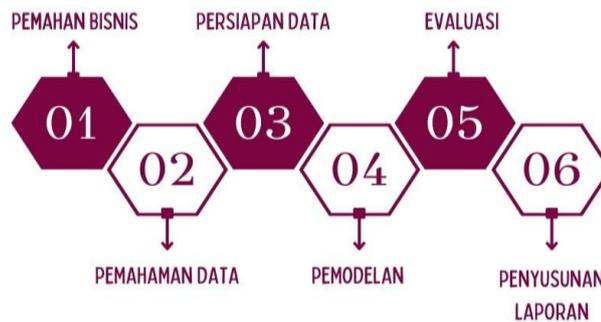
- Tahap Pemahaman bisnis (Business Understanding), Tahap ini merupakan tahap pemahaman terhadap substansi kegiatan data mining yang ingin dilakukan, kebutuhan perspektif penelitian. Kejadiannya yaitu menentukan sasaran/tujuan penelitian, memahami situasi penelitian, menerjemahkan tujuan penelitian dalam tujuan data mining.
- Tahap Pemahaman Data (Data Understanding), mengumpulkan data, mempelajari data agar data dapat dipahami kemudian digunakan pada penelitian,

- c) Tahap persiapan Data (Data Preparation), Mempersiapkan struktur basis datayang dapat memudahkan proses mining.
- d) Tahap Pemodelan (Modeling Phase), merupakan fase penentuan penggunaan teknik data mining, algoritma data mining, tools data mining, serta penentuan parameter dengan nilai optimal.
- e) Tahap Evaluasi (Evaluation Phase), Merupakan interpretasi hasil data mining pada proses pemodelan. Evaluasi bertujuan untuk menyesuaikan model yang diperoleh agar sesuai dengan sasaran yang akan dicapai.
- f) Tahap Penerapan (Deployment Phase) Merupakan tahap menyusun laporan pengetahuan yang diperoleh dari evaluasi proses data mining.

3. METODOLOGI

3.1 Tahapan Data Mining

Untuk memudahkan penelitian dan mencapai tujuan yang diinginkan, langkah-langkah berikut dibuat dalam tahapan penelitian:



Gambar 2. Tahapan Data Mining

Berikut penjelasan tahapan dalam penelitian data mining yang ditampilkan dalam Gambar 2:

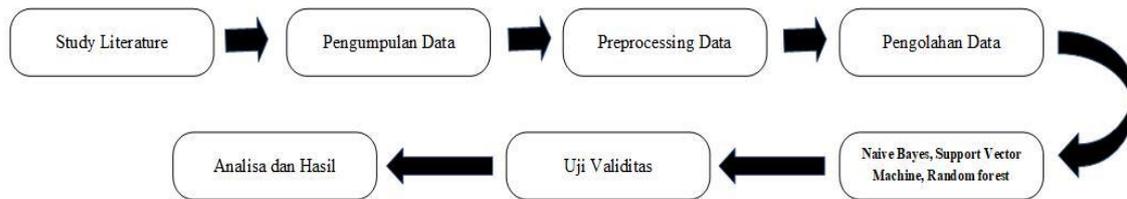
- 1) Fase pemahaman bisnis: Fase ini berfokus pada tujuan dan persyaratan proyek, serta merumuskan definisi masalah penambangan data untuk mencapai tujuan tersebut.
- 2) Fase pemahaman data: Pada fase ini, data dikumpulkan dan dilakukan analisis eksplorasi untuk memahami data, mengevaluasi kualitasnya, dan memilih subset data yang relevan.
- 3) Fase persiapan data: Tahap ini melibatkan persiapan data mentah awal, termasuk pemilihan kasus dan variabel yang diperlukan untuk analisis. Jika perlu, dilakukan transformasi data dan pembersihan untuk persiapan pemodelan.
- 4) Fase pemodelan: Pada fase ini, teknik pemodelan yang sesuai yaitu menggunakan algoritma Naive Bayes, Support Vector Machine dan Random forest dipilih dan diterapkan untuk memperoleh hasil yang maksimal.
- 5) Fase evaluasi: Model yang dikembangkan dievaluasi untuk mengukur kualitas dan efektivitasnya sebelum digunakan secara praktis. Tujuan penelitian ditentukan apakah tercapai atau tidak. Jika ada aspek penting yang belum teratasi, hasil penambangan data digunakan untuk pengambilan keputusan.
- 6) Tahap penyusunan laporan disertasi: Tahap ini melibatkan dokumentasi pekerjaan penelitian, mencatat hasil penelitian, dan menyusun laporan berbentuk argumentasi yang dapat digunakan sebagai literatur.

Dengan mengikuti tahapan tersebut, penelitian data mining dapat dilakukan secara sistematis dan menghasilkan laporan yang relevan untuk kepentingan bisnis atau penelitian.

3.2 Kerangka Kerja Penelitian

Gambar 1 dalam penelitian ini adalah kerangka pikiran yang disusun secara sistematis dan menggambarkan langkah-langkah proses yang akan diikuti dalam penelitian. Kerangka kerja ini akan membantu peneliti dalam

memperoleh data dan informasi yang dibutuhkan untuk memprediksi ketidakhadiran di tempat kerja. Penelitian ini menggunakan tiga algoritma klasifikasi, yaitu Naive Bayes, Support Vector Machine (SVM), dan Random Forest.



Gambar 3. Kerangka Kerja Penelitian

3.3 Pengolahan Data Awal

Data penelitian ini, terdiri dari 400 record, 25 atribut dan target class adalah penyakit ginjal kronis (PGK). Data bersumber dari UCI. Tahap pemahaman data yaitu dengan mendeskripsikan atribut - atribut yang terdapat pada dataset dan mendeskripsikan tipe nilai atribut pada dataset yang digunakan. Dibawah ini adalah tabel 1. Atribut Dataset

Tabel 1. Atribut Dataset

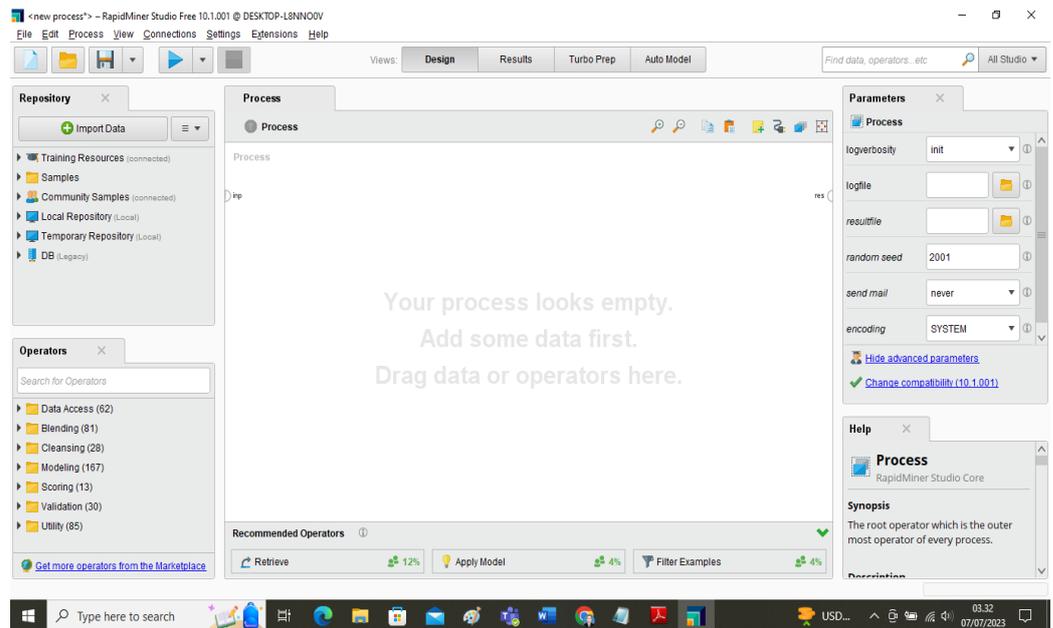
ATRIBUT	DESKRIPSI	NILAI
Age	Umur	Umur dalam tahun
Blood Pressure (bp)	Ukuran tekanan yang dihasilkan oleh darah	mm/Hg
Specific Gravity (sg)	Seriusnya suatu situasi atau kondisi	1.005,1.010,1.015,1.020,1.025
Albumin (al)	Protein plasma darah	0,1,2,3,4,5
Sugar (su)	Kadar gula darah	0,1,2,3,4,5
Red Blood Cells (rbc)	Pengukuran sel darah merah	Normal,abnormal
Pus Cell (pc)	Pengukuran sel nanah dalam urin	Normal,abnormal
Pus Cell clumps (pcc)	Pengukuran gumpalan sel nanah dalam urin	Present,notpresent
Bacteria (ba)	Pengukuran bakteri yang terdapat dalam tubuh	Present,notpresent
Blood Glucose Random (bgr)	Pengukuran kadar glukosa	mgs/dl
Blood Urea (bu)	Pengukuran kadar urea dalam darah	mgs/dl
Serum Creatinine (sc)	Pengukuran kreatin pada darah	mgs/dl
Sodium (sod)	Pengukuran natrium pada darah	mEq/L
Potassium (pot)	Pengukuran kalium pada darah	mEq/L
Hemoglobin (hemo)	Menghitung jumlah hemoglobin	gms
Packed Cell Volume	Parameter pengukuran sel darah	numeric
White Blood Cell Count (wc)	Menghitung jumlah total sel darah merah	Millions/cmm

Red Blood Cell Count (rc)	Menghitung jumlah total sel darah merah	Millions/cmm
Hypertension (htn)	Memiliki masalah hipertensi	Yes,no
Diabetes Mellitus (dm)	Memiliki masalah diabetes	Yes,no
Coronary Artery Disease (cad)	Coronary Artery Disease	Yes,no
Appetite (appet)	Memiliki masalah nafsu makan	Good,poor
Pedal Edema (pe)	Pengukuran cairan pada betis	Yes,no
Anemia (ane)	Pengukuran darah merah	Yes,no
Class (class)	Target Class	Ckd, notckd

4. HASIL DAN PEMBAHASAN

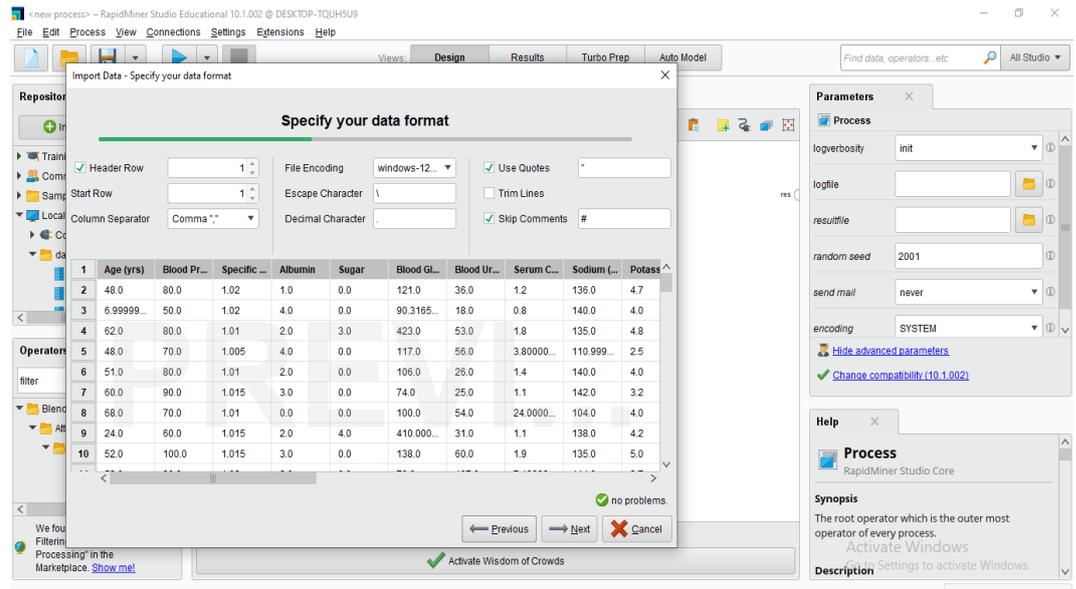
Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma, yaitu Naive Bayes, SVM, dan Random Forest, tool menggunakan software RapidMiner Studio, dalam memprediksi Penyakit Ginjal Kronis (PGK) dan studi ini dilakukan pada dataset yang tersedia untuk umum. Berikut adalah penerapan tool RapidMiner pada pengujian Algoritma Naive Bayes, SVM, dan Random Forest:

- [1] Untuk memulai, langkah pertama yang perlu dilakukan adalah membuka aplikasi RapidMiner terlebih dahulu.



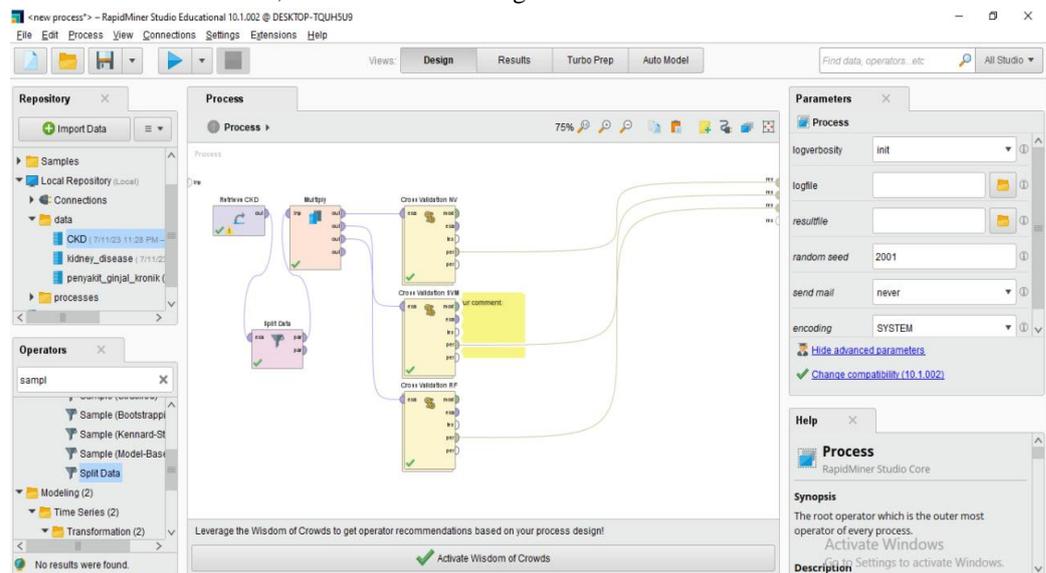
Gambar 4. Tampilan Rapidminer 10

- [2] Dataset yang akan digunakan dalam RapidMiner adalah dataset ckd.csv yang diperoleh dari kaggle.com. Untuk mengimpor dataset ini ke dalam RapidMiner, langkah-langkah pengaturan dataset dapat dilihat pada Gambar 5 di bawah ini.



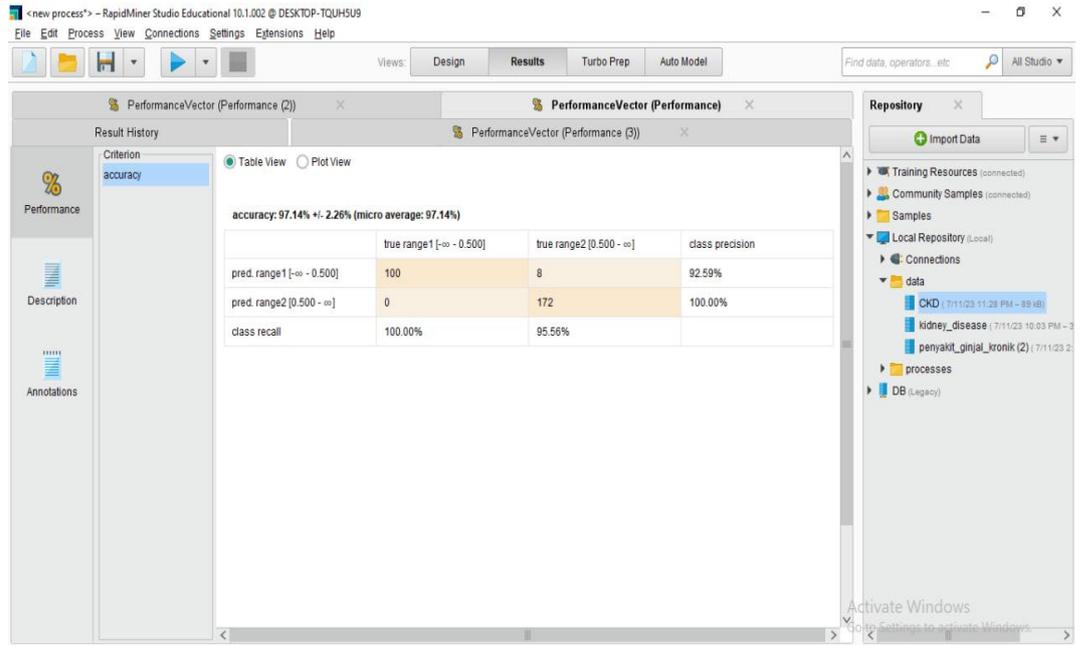
Gambar 5. Tampilan Pengaturan dan Proses Import Dataset ke dalam RapidMiner

- [3] Peneliti melakukan desain klasifikasi dengan menggunakan algoritma Naive Bayes, Support Vector Machine dan Random forest, seperti yang terlihat pada Gambar 5. Setelah dataset diimpor ke RapidMiner, langkah selanjutnya adalah menghubungkannya ke algoritma klasifikasi Naive Bayes, Support Vector Machine dan Random forest. Kemudian, hasilnya dihubungkan ke Apply Model, diikuti oleh Performance, dan terakhir dihubungkan ke Result.

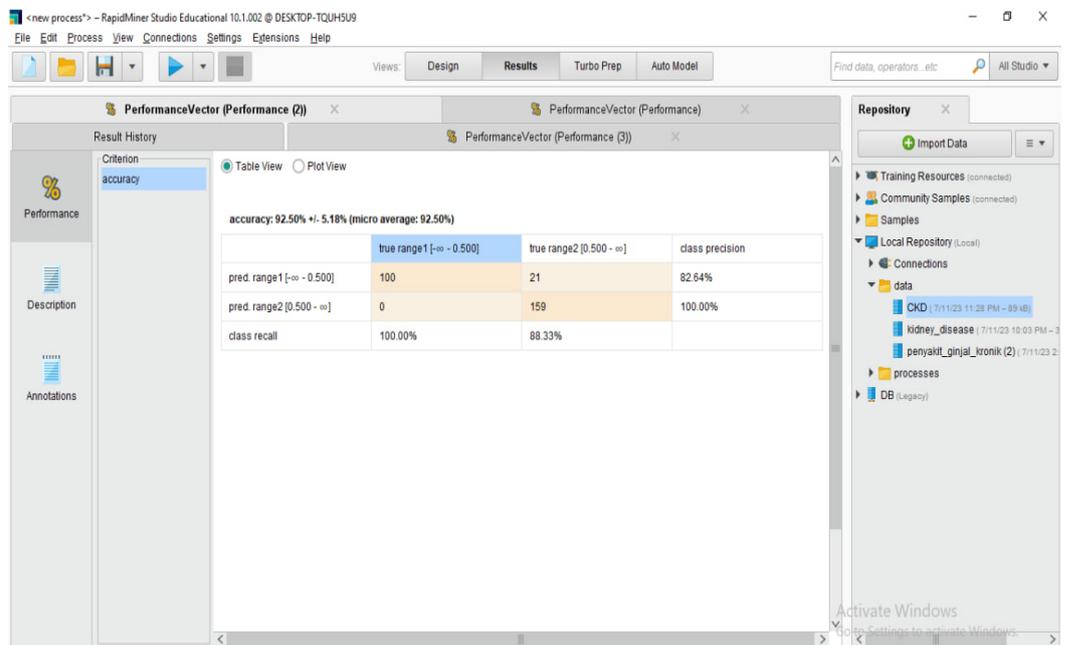


Gambar 6. Tampilan Design Klasifikasi Menggunakan Algoritma klasifikasi Naive Bayes, Support Vector Machine dan Random Forest

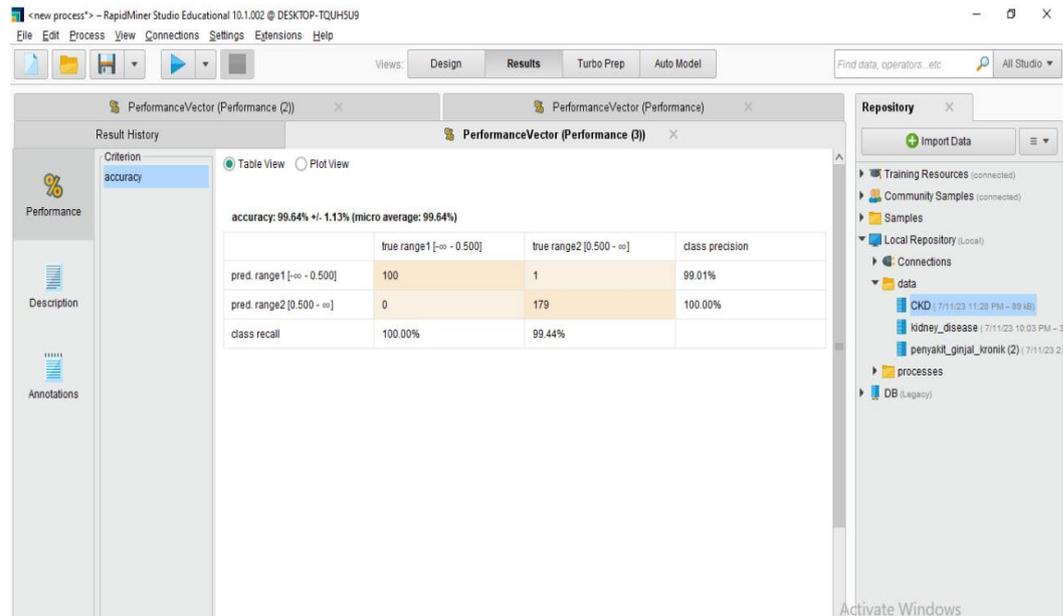
- [4] Setelah dilakukan pengujian, ditemukan bahwa pada split data (data training 0.7% dan data testing 0.3%) pemrosesan klasifikasi menggunakan algoritma Naive Bayes sebesar 97.14%, sedangkan pemrosesan klasifikasi menggunakan algoritma Support Vector Machine sebesar 92.50% dan untuk pemrosesan klasifikasi menggunakan algoritma Random forest sebesar 99.64% seperti yang tercatat pada Gambar 6, 7, 8 dan 9.



Gambar 7. Hasil Akurasi Dari Algoritma Naïve Bayes



Gambar 8. Hasil Akurasi dari Algoritma Support Vector Machine



Gambar 9. Hasil akurasi dari algoritma Random Forest

Tabel 2. Perbandingan Akurasi Nilai Kinerja Algoritma Naïve Bayes, Support Vector Machine dan Random Forest pada Prediksi Penyakit Ginjal Kronis

Algoritma	Akurasi
Naïve Bayes	97.14%
Support Vector Machine	92.50%
Random Forest	99.64%

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dengan menggunakan tools RapidMiner 10, ditemukan bahwa algoritma Random Forest memiliki nilai akurasi yang lebih tinggi sebesar 99.64% dibandingkan dengan Support Vector Machine sebesar 92.50% dan Naïve Bayes sebesar 97.14%. Hal ini disebabkan oleh performa komputasi yang lebih rendah pada Support Vector Machine dan Naïve Bayes dibandingkan dengan Random Forest. Untuk penelitian selanjutnya, disarankan untuk melakukan komparasi pada algoritma klasifikasi lainnya.

UCAPAN TERIMA KASIH

Alhamdulillah dengan rasa syukur dan terima kasih kepada Allah SWT, peneliti berhasil menyelesaikan penelitian ini. Saya ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua yang terlibat dalam penulisan penelitian ini. Dukungan, kontribusi, dan partisipasi Anda telah memainkan peran yang sangat penting dalam kesuksesan penelitian ini. Terima kasih atas waktu, pengetahuan, dan usaha yang Anda sumbangkan untuk membuat penelitian ini menjadi karya yang berarti dan bermanfaat.

DAFTAR PUSTAKA

A'yuniyah, Q., Tasia, E., Nazira, N., Pratama, P. F., Anugrah, M. R., Adhiva, J., & Mustakim, M. (2022). Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(1), 72. <https://doi.org/10.30865/json.v4i1.4781>

-
- Amalia, H. (2018). Perbandingan Metode Data Mining Svm Dan Nn Untuk Klasifikasi Penyakit Ginjal Kronis. *Maret*, 14(1), 1. www.bsi.ac.id
- Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *Sistemasi*, 10(1), 163. <https://doi.org/10.32520/stmsi.v10i1.1129>
- Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 6(2), 129. <https://doi.org/10.30998/string.v6i2.10133>
- Arifin, T., & Ariesta, D. (2019). Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization. *Jurnal Tekno Insentif*, 13(1), 26–30. <https://doi.org/10.36787/jti.v13i1.97>
- Chakraborty, S., Islam, S. H., & Samanta, D. (2022). Introduction to Data Mining and Knowledge Discovery. In *EAI/Springer Innovations in Communication and Computing*. https://doi.org/10.1007/978-3-030-93088-2_1
- Chou, L., Yang, H., & Hung, C. (2023). ScienceDirect Leptospirosis kidney disease : Evolution from acute to chronic kidney disease. *Biomedical Journal*, May, 100595. <https://doi.org/10.1016/j.bj.2023.100595>
- Dany Prasetya, F., Nugroho, H. W., & Triloka, J. (2022). Analisa Data Mining Untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Decision Tree C.45 Dengan Particle Swarm Optimization. *Prosiding Seminar Nasional Darmajaya, April 1989*, 198–209. <http://archive.ic>
- Darwanto, A. R. S., Taza Luzia Viarindita, & Yekti Widyaningsih. (2021). Analisis Regresi Logistik Binomial dan Algoritma Random Forest pada Proses Pengklasifikasian Penyakit Ginjal Kronis. *Jurnal Statistika Dan Aplikasinya*, 5(1), 1–14. <https://doi.org/10.21009/jsa.05101>
- Data, I., Untuk, M., Kinerja, M., Karyawan, K., Metode, M., & Linier, R. (2022). Implementasi Data Mining Untuk Memprediksi Kinerja. *02(1)*, 127–135.
- Fadilla, I., Adikara, P. P., & Setya Perdana, R. (2018). Klasifikasi Penyakit Chronic Kidney Disease (CKD) Dengan Menggunakan Metode Extreme Learning Machine (ELM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3397–3405. <https://www.researchgate.net/publication/323365845>
- Kyneissia Gliselda, V. (2021). Diagnosis dan Manajemen Penyakit Ginjal Kronis (PGK). *Jurnal Medika Utama*, 2(4).
- Labib, M., Zaidiah, A., & Yulistiawan, B. S. (2023). Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta). 155–166.
- Lisa Lolowang, N. N., Lumi, W. M. ., & Rattoe, A. A. (2021). KUALITAS HIDUP PASIEN GAGAL GINJAL KRONIS DENGAN TERAPI HEMODIALISA. *Jurnal Ilmiah Perawat Manado (Juiperdo)*, 8(02). <https://doi.org/10.47718/jpd.v8i01.1183>
- Mahardika Pratama, I. G. A., Astuti, L. G., Widiartha, I. M., Cahyadi Putra, I. G. N. A., Adi Pramarta, C. R., & Atmaja Darmawan, I. D. M. B. (2022). Diagnosis Penyakit Ginjal Kronis dengan Algoritma C4.5, K-Means dan BPSO. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 10(4), 371. <https://doi.org/10.24843/jlk.2022.v10.i04.p07>
- Moreira, L. B., & Namen, A. A. (2018). A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. *Computer Methods and Programs in Biomedicine*, 165, 139–149. <https://doi.org/10.1016/j.cmpb.2018.08.016>
- Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2021). Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja. *Jurnal Informatika Universitas Pamulang*, 5(4), 578. <https://doi.org/10.32493/informatika.v5i4.7575>
-

- Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2018). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa | Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(4), 1704–1708. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1351>
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, 4(1), 13–21. <https://doi.org/10.30813/jbase.v4i1.2729>
- Triwidianti, J. (2021). *Klasifikasi Penyakit Kanker Payudara Menggunakan Metode Seleksi Fitur dan Algoritma Decision Tree*No Title.
- Widayati, N. &. (2019). *Buku Panduan Mengenal Penyakit Ginjal Kronis dan perawatannya*. http://eprints.undip.ac.id/81430/1/Buku_Panduan_Mengenal_Penyakit_Ginjal_Kronis_dan_Perawatannya_Henni_Kusuma%2C_Suhartini%2C_Untung_Sujianto%2C_Chandra_Bagus_Ropiyanto%2C_Wahyu_Hidayati.pdf
-