

# Perbandingan Kinerja Algoritma K-Nearest Neighbor dan Algoritma Random Forest Untuk Klasifikasi Data Mining Pada Penyakit Gagal Ginjal

Salmon<sup>1,\*</sup>, Azahari<sup>2</sup>, Hanifah Ekawati<sup>3</sup>

<sup>1</sup>Prodi Sistem Informasi, STMIK Widya Cipta Dharma, Samarinda, Indonesia

<sup>2</sup>Prodi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda, Indonesia

<sup>3</sup>Prodi Bisnis Digital, STMIK Widya Cipta Dharma, Samarinda, Indonesia

Email: <sup>1,\*</sup>salmon@wicida.ac.id, <sup>2</sup>azahari@wicida.ac.id, <sup>3</sup>hanifah@wicida.ac.id

Email Penulis Korespondensi: salmon@wicida.ac.id

Submitted: 19/12/2024; Accepted: 29/12/2024; Published: 30/12/2024

**Abstrak**—Penyakit gagal ginjal merupakan salah satu penyakit kronis yang banyak dijumpai di seluruh dunia. Kondisi ini terjadi ketika ginjal kehilangan kemampuan untuk menyaring limbah dan cairan berlebih dari darah. Penyakit gagal ginjal merupakan kondisi serius yang terjadi ketika fungsi ginjal menurun secara signifikan atau berhenti sama sekali. Penyakit gagal ginjal memiliki dampak yang luas pada kesehatan fisik, mental, dan sosial pasien. Oleh karena itu, penanganan dini dan pendekatan holistik sangat diperlukan untuk meminimalkan dampaknya. Dalam bidang kesehatan, kemajuan teknologi telah memungkinkan pengolahan data medis secara lebih efektif melalui penerapan data mining. Data Mining adalah proses menggali dan menganalisis data dalam jumlah besar untuk menemukan pola, hubungan, atau informasi berharga yang sebelumnya tidak diketahui. Klasifikasi dalam Data Mining adalah proses mengelompokkan atau mengkategorikan data ke dalam kelas atau label tertentu berdasarkan atribut atau fitur yang dimilikinya. Pada klasifikasi sendiri terdapat berbagai macam algoritma didalamnya seperti Algoritma K-Nearest Neighbor (KNN) dan Random Forest (RF). Algoritma K-Nearest Neighbor (KNN) dan Random Forest (RF) merupakan dua algoritma yang banyak digunakan dalam tugas-tugas klasifikasi. Maka dari itu, pada penelitian ini akan dilakukan proses perbandingan terhadap kinerja algoritma K-Nearest Neighbor dan algoritma Random Forest. Perbandingan kinerja algoritma data mining untuk mengevaluasi dan menentukan algoritma mana yang paling efektif dan efisien dalam menyelesaikan masalah tertentu berdasarkan berbagai metrik evaluasi. Secara keseluruhan nilai akurasi yang didapatkan diatas 90%, namun algoritma Random Forest memiliki kinerja yang lebih baik. Dimana hasil tingkat akurasi yang didapatkan dari algoritma Random Forest sebesar 99,75%. Maka dari itu, model ataupun pola yang dihasilkan dari algoritma Random Forest nantinya dipergunakan untuk membantu dalam proses diagnosa penyakit gagal ginjal dan algoritma Random Forest merupakan algoritma yang memiliki kinerja lebih baik.

**Kata Kunci:** Perbandingan; Kinerja; Random Forest; K-NN; Data Mining

**Abstract**—Kidney failure is one of the most common chronic diseases worldwide. This condition occurs when the kidneys lose their ability to filter waste and excess fluid from the blood. Kidney failure is a serious condition that occurs when kidney function decreases significantly or stops altogether. Kidney failure has a wide impact on the physical, mental, and social health of patients. Therefore, early treatment and a holistic approach are needed to minimize its impact. In the health sector, technological advances have enabled more effective processing of medical data through the application of data mining. Data Mining is the process of exploring and analyzing large amounts of data to find patterns, relationships, or valuable information that was previously unknown. Classification in Data Mining is the process of grouping or categorizing data into certain classes or labels based on the attributes or features it has. In the classification itself, there are various algorithms in it such as the K-Nearest Neighbor (KNN) and Random Forest (RF) algorithms. The K-Nearest Neighbor (KNN) and Random Forest (RF) algorithms are two algorithms that are widely used in classification tasks. Therefore, this study will carry out a comparison process on the performance of the K-Nearest Neighbor algorithm and the Random Forest algorithm. Comparison of data mining algorithm performance to evaluate and determine which algorithm is the most effective and efficient in solving a particular problem based on various evaluation metrics. Overall, the accuracy value obtained is above 90%, but the Random Forest algorithm has better performance. Where the accuracy level results obtained from the Random Forest algorithm are 99.75%. Therefore, the model or pattern produced by the Random Forest algorithm will later be used to assist in the process of diagnosing kidney failure and the Random Forest algorithm is an algorithm that has better performance.

**Keywords:** Comparison; Performance; Random Forest; K-NN; Data Mining

## 1. PENDAHULUAN

Penyakit gagal ginjal merupakan salah satu penyakit kronis yang banyak dijumpai di seluruh dunia. Kondisi ini terjadi ketika ginjal kehilangan kemampuan untuk menyaring limbah dan cairan berlebih dari darah, yang dapat menyebabkan komplikasi serius seperti gangguan elektrolit, anemia, dan tekanan darah tinggi. Berdasarkan data dari Organisasi Kesehatan Dunia (WHO), prevalensi gagal ginjal kronis terus meningkat setiap tahunnya, terutama di negara-negara berkembang, termasuk Indonesia. Oleh karena itu, penting untuk melakukan diagnosis dini untuk mengurangi risiko komplikasi dan meningkatkan kualitas hidup pasien[1]–[3].

Penyakit gagal ginjal merupakan kondisi serius yang terjadi ketika fungsi ginjal menurun secara signifikan atau berhenti sama sekali. Ginjal memiliki peran penting dalam menyaring limbah metabolisme, menjaga keseimbangan cairan dan elektrolit, serta mengatur tekanan darah. Ketika ginjal gagal menjalankan fungsi-fungsi ini, berbagai dampak buruk pada kesehatan dapat terjadi, baik secara fisik maupun psikologi. Ginjal yang tidak berfungsi optimal menyebabkan penumpukan zat beracun seperti urea, kreatinin, dan produk metabolisme lainnya dalam darah[4]–[6].

Penyakit gagal ginjal memiliki dampak yang luas pada kesehatan fisik, mental, dan sosial pasien. Oleh karena itu, penanganan dini dan pendekatan holistik sangat diperlukan untuk meminimalkan dampaknya. Diagnosis dini, manajemen medis yang tepat, serta dukungan psikologis dan sosial dapat membantu pasien mempertahankan kualitas hidup yang lebih baik. Teknologi modern, seperti sistem pendukung keputusan berbasis data mining, juga dapat digunakan untuk mendukung diagnosis dan pengelolaan penyakit ini secara lebih efektif[7]–[9].

Dalam bidang kesehatan, kemajuan teknologi telah memungkinkan pengolahan data medis secara lebih efektif melalui penerapan data mining. Data mining adalah proses ekstraksi informasi berharga dari kumpulan data besar, yang dapat digunakan untuk mendukung pengambilan keputusan klinis, termasuk diagnosis penyakit. Salah satu teknik yang sering digunakan dalam data mining adalah algoritma klasifikasi, yang bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang ditemukan dalam dataset[10]–[12].

*Data Mining* adalah proses menggali dan menganalisis data dalam jumlah besar untuk menemukan pola, hubungan, atau informasi berharga yang sebelumnya tidak diketahui. Tujuannya adalah untuk membantu pengambilan keputusan, prediksi, atau pengelompokan data secara lebih efektif. Proses data mining melibatkan berbagai langkah, seperti pengumpulan data, pembersihan data, analisis, dan interpretasi hasil. Pada data mining terdapat banyak teknik didalamnya salah satu teknik yang dapat dilakukan adalah Teknik Klasifikasi untuk melakukan proses penyelesaian[13],[14].

Klasifikasi dalam *Data Mining* adalah proses mengelompokkan atau mengkategorikan data ke dalam kelas atau label tertentu berdasarkan atribut atau fitur yang dimilikinya. Tujuan utama klasifikasi adalah untuk membuat model atau fungsi yang dapat memetakan data masukan (*input*) ke dalam kategori yang telah ditentukan sebelumnya (*output*)[15]–[17]. Pada klasifikasi sendiri terdapat berbagai macam algoritma didalamnya seperti Algoritma *K-Nearest Neighbor* (KNN) dan *Random Forest* (RF).

Algoritma *K-Nearest Neighbor* (KNN) dan *Random Forest* (RF) merupakan dua algoritma yang banyak digunakan dalam tugas-tugas klasifikasi. KNN adalah metode berbasis instance yang mengklasifikasikan data berdasarkan kedekatan dengan data pelatihan terdekat. Algoritma ini dikenal sederhana dan intuitif, tetapi memiliki keterbatasan dalam menangani dataset yang besar dan berdimensi tinggi. Sebaliknya, *Random Forest* adalah algoritma berbasis ensemble yang menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang lebih akurat dan stabil. Algoritma ini unggul dalam menangani dataset kompleks, tetapi membutuhkan sumber daya komputasi yang lebih besar.

Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma sederhana namun efektif dalam data mining dan machine learning yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bekerja berdasarkan prinsip bahwa data serupa biasanya memiliki karakteristik yang sama. Dimana proses dari algoritma *K-Nearest Neighbor* berdasarkan dengan jarak dari tetangga terdekat[18]–[20]. Dalam menunjang pelaksanaan penelitian, dapat diketahui beberapa penelitian terdahulu berkaitan dengan algoritma *K-Nearest Neighbor*.

Penelitian pertama dilakukan oleh Sulistyowati Munawaroh, dkk pada tahun 2024 dengan judul penelitian yang dilakukan “Klasifikasi Tingkat Kecemasan Atlet Sebelum Bertanding Menggunakan Algoritma *K-Nearest Neighbor* (KNN) Berbasis Website” dimana pada penelitian hasil yang didapatkan bahwa proses pengujian dilakukan dengan membagi data menjadi 80:20 dengan nilai  $K=5$  serta hasil akurasi yang didapatkan sebesar 100% [21].

Penelitian Kedua yang dilakukan oleh Agfa Oktaviana, dkk juga pada tahun 2024 dengan judul penelitian “Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma *K-Nearest Neighbor* (K-NN)” dimana dari proses yang dilakukan hasil pengujian bahwa dengan menggunakan evaluasi akurasi, presisi, recall, dan f1-score, masing-masing memberikan hasil yaitu 88%, 83%, 87%, dan 85% [22].

Penelitian selanjutnya yang dilakukan oleh Uktupi Nijunnihayah, dkk pada tahun 2024 dengan judul penelitian “Implementasi Algoritma *K-Nearest Neighbor* untuk Prediksi Penjualan Alat Kesehatan pada Media Alkes” serta hasil pengujian penelitian bahwa proses prediksi yang dilakukan mendapatkan tingkat akurasi sebesar 95% [23].

Penelitian terakhir yang dilakukan oleh Sekar Ayu Fitria, dkk juga pada tahun 2024 dengan judul penelitian “Data Mining Classification Untuk Prediksi Jumlah Mahasiswa Aktif dan Cuti Angkatan 2020 Menggunakan Metode *K-Nearest Neighbor*” dari proses yang dilakukan didapatkan hasil bahwasannya tingkat akurasi yang didapatkan sebesar 89% [24].

*Random Forest* adalah algoritma *ensemble learning* yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan menggabungkan banyak *decision tree* yang dibuat dari subset data pelatihan. Hasil akhir diperoleh dengan cara voting (untuk klasifikasi) atau rata-rata (untuk regresi) dari prediksi semua pohon keputusan[25]–[27]. Maka dari itu, kiranya diperlukan penelitian terdahulu sebagai penunjang dalam proses pelaksanaan penelitian.

Penelitian pertama dilakukan oleh Krisna Giana Putra pada tahun 2024 dengan judul pada penelitian “Penentuan Penerima Bantuan Program Keluarga Harapan menggunakan Algoritma *Random Forest* di Desa Kebonrejo” dimana pada proses yang dilakukan didapatkan hasil penelitian bahwa evaluasi kinerja terhadap model dilakukan dengan menggunakan confusion matrix dengan nilai akurasi, presisi, recal dan F1. Hasil dari keseluruhan nilai mencapai 90% [28].

Penelitian berikutnya dilakukan oleh Laelatul Hidayah dan Muhammad Imron Rosadi pada tahun 2024 dimana penelitian yang dilakukan “Penerapan Algoritma *Random Forest* Untuk Memprediksi Jumlah Santri Baru” serta hasil yang didapatkan pada penelitian bahwa algoritma *Random Forest* memiliki tingkat akurasi yang tinggi sebesar 97,73% dalam memprediksi jumlah santri baru[29].

Penelitian ketigan juga telah dilakukan oleh Muhammad Rizky Ardiansyah Putra dan Rissa Nurfitriana Handayani pada tahun 2024 dengan penelitian yang dilakukan “Perbandingan Algoritma Decision Tree dan Random Forest Dalam Pengklasifikasian Penyakit Tiroid” dari proses yang dilakukan hasil yang didapatkan bahwa algoritma Random Forest memiliki kinerja lebih baik dengan nilai akurasi sebesar 94,81% [30].

Serta penelitian terakhir yang dilakukan oleh Duwi Cahya Putri Buani pada tahun 2024 dengan judul dari penelitian yang dilakukan “Deteksi Dini Penyakit Diabetes dengan Menggunakan Algoritma Random Forest”. Dari proses yang telah dilakukan pada penelitian didapatkan hasil bahwa nilai akurasi yang didapatkan berdasarkan proses pengukuran kinerja sebesar 98,78% [31].

Dalam konteks klasifikasi penyakit gagal ginjal, pemilihan algoritma yang tepat menjadi tantangan yang signifikan. Karakteristik dataset medis, seperti jumlah data yang terbatas, ketidakseimbangan kelas, dan fitur yang kompleks, memengaruhi kinerja algoritma klasifikasi. Oleh karena itu, perlu dilakukan analisis komparatif untuk menentukan algoritma yang paling sesuai digunakan dalam klasifikasi data medis terkait gagal ginjal.

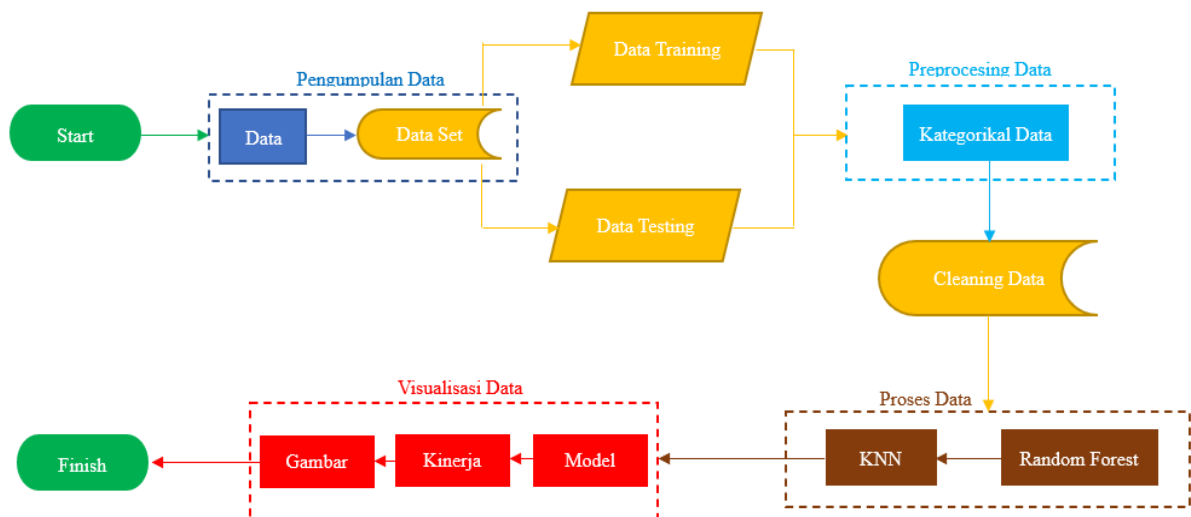
Maka dari itu, pada penelitian ini akan dilakukan proses perbandingan terhadap kinerja algoritma *K-Nearest Neighbor* dan algoritma *Random Forest*. Perbandingan kinerja algoritma data mining untuk mengevaluasi dan menentukan algoritma mana yang paling efektif dan efisien dalam menyelesaikan masalah tertentu berdasarkan berbagai metrik evaluasi. Dengan membandingkan kinerja beberapa algoritma, pengguna atau peneliti dapat memilih metode yang paling sesuai dengan karakteristik data dan tujuan analisis.

Penelitian ini bertujuan untuk membandingkan kinerja algoritma KNN dan *Random Forest* dalam klasifikasi data mining untuk penyakit gagal ginjal. Dengan menganalisis metrik kinerja seperti akurasi, presisi, recall, dan waktu pemrosesan, diharapkan penelitian ini dapat memberikan wawasan yang berguna dalam pengembangan sistem pendukung keputusan di bidang kesehatan. Selain itu, hasil penelitian ini juga diharapkan dapat membantu praktisi medis dan peneliti dalam memilih algoritma yang optimal untuk aplikasi diagnosis penyakit.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Pada tahapan penelitian dibawah ini merupakan alur dari tahapan penelitian serta proses pengumpulan data yang dilakukan, adalah sebagai berikut:



Gambar 1. Tahapan Penelitian

### 2.2 Data Mining

*Data mining* merujuk pada proses eksplorasi dan analisis data dalam jumlah besar untuk menemukan pola, tren, hubungan tersembunyi, atau informasi berharga yang dapat digunakan untuk membuat keputusan atau prediksi. Data mining melibatkan penggunaan berbagai teknik matematis, statistik, dan algoritma komputer untuk menggali wawasan dari dataset yang besar dan kompleks [14], [32].

*Data mining* adalah proses menemukan pola atau informasi yang berguna dalam jumlah besar data yang tidak terstruktur dengan menggunakan teknik statistik dan algoritma komputer. Mereka menjelaskan bahwa data mining tidak hanya berfokus pada pencarian pola tetapi juga pada penyaringan data yang relevan untuk membangun model prediksi atau klasifikasi [33], [34].

*Data mining* bertujuan untuk mengekstrak pengetahuan yang tidak terlihat atau tidak langsung tampak di permukaan data. Ini mencakup identifikasi pola yang berulang, hubungan antar variabel, dan informasi tersembunyi yang dapat mendukung pengambilan keputusan atau tindakan lebih lanjut. *Data mining* sebagai langkah penting dalam

proses pengetahuan untuk mengidentifikasi pola yang ada, yang dapat digunakan untuk membuat prediksi dan keputusan berbasis data. Data mining sering diterapkan pada volume data yang sangat besar (*big data*) dan kompleks, yang tidak dapat dianalisis secara manual atau dengan menggunakan metode analisis tradisional. Ini melibatkan pengolahan data dari berbagai sumber, termasuk basis data, *file log*, data sensor, dan data media sosial[35], [36].

Secara keseluruhan, *data mining* adalah proses yang sangat penting dalam dunia yang semakin didorong oleh data, karena dapat mengubah data mentah menjadi informasi yang bernilai tinggi dan membantu dalam perencanaan serta pengambilan keputusan yang lebih akurat dan berbasis bukti. Proses ini menggunakan teknik dan algoritma canggih untuk mengolah data yang kompleks dan tidak terstruktur[37], [38].

### 2.3 Algoritma K-Nearest Neighbor

*K-Nearest Neighbor* (K-NN) adalah algoritma pembelajaran mesin yang termasuk dalam kategori *supervised learning* dan digunakan untuk klasifikasi dan regresi. Pada dasarnya, algoritma ini mengklasifikasikan data berdasarkan kedekatannya dengan titik data lainnya. Dalam K-NN, "K" merujuk pada jumlah tetangga terdekat yang digunakan untuk menentukan kelas atau nilai suatu data[4], [22].

Algoritma K-NN mengasumsikan bahwa data dengan fitur yang mirip berada dalam kelas yang sama, sehingga untuk mengklasifikasikan suatu data baru, algoritma ini mencari "k" tetangga terdekat dari data tersebut dan memutuskan kelas berdasarkan mayoritas kelas di antara tetangga terdekat tersebut (untuk klasifikasi) atau nilai rata-rata (untuk regresi). Adapun tahapan proses pada algoritma *K-Nearest Neighbor* [9], [21], [23]:

1. Pisahkan dataset menjadi dua bagian: data pelatihan (*training set*) dan data uji (*test set*).
2. Pilih nilai "K" yang akan digunakan, yaitu jumlah tetangga terdekat yang akan dipertimbangkan. Nilai "K" biasanya dipilih berdasarkan eksperimen, dan nilai yang lebih kecil dapat menyebabkan model lebih sensitif terhadap *noise*, sementara nilai yang besar dapat menghasilkan keputusan yang lebih stabil tetapi kurang sensitif.
3. Menghitung jarak terdekat, Untuk menentukan tetangga terdekat, pertama-tama kita harus menghitung jarak antara data uji dan setiap data dalam data pelatihan dengan menggunakan rumus *Euclidean Distance*:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

4. Mencari tetangga terdekat, Setelah jarak dihitung, urutkan titik-titik pelatihan berdasarkan jaraknya ke titik uji yang ingin diklasifikasikan. Pilih "K" titik pelatihan dengan jarak terkecil. Ini adalah tetangga terdekat dari data uji.
5. Melakukan klasifikasi data, Lihat kelas mayoritas dari K tetangga terdekat. Kelas yang paling sering muncul di antara K tetangga tersebut adalah kelas yang diprediksi untuk data uji.

### 2.4 Algoritma Random Forest

*Random Forest* adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi, yang bekerja dengan menggabungkan beberapa pohon keputusan (*decision trees*) untuk menghasilkan model yang lebih akurat dan lebih stabil. Konsep dasarnya adalah membuat banyak pohon keputusan yang masing-masing dilatih dengan data yang sedikit berbeda, dan hasil akhirnya digabungkan untuk memberikan prediksi yang lebih baik[28], [31].

*Random Forest* adalah algoritma *ensemble learning*, yang berarti menggabungkan beberapa model pembelajaran (dalam hal ini, pohon keputusan) untuk menghasilkan prediksi yang lebih kuat. Setiap pohon keputusan dalam *Random Forest* dilatih menggunakan subset acak dari data pelatihan (dikenal sebagai *bootstrapping*) dan memilih subset acak dari fitur untuk menentukan *split* pada setiap *node*. Setelah pohon-pohon keputusan dilatih, untuk klasifikasi, prediksi akhir diambil berdasarkan hasil mayoritas voting dari semua pohon. Untuk regresi, prediksi akhir adalah rata-rata dari hasil prediksi semua pohon. Adapun tahapan dalam proses algoritma *Random Forest* sebagai berikut[26], [27]:

- a. Pisahkan data menjadi data pelatihan (*training set*) dan data uji (*test set*). Pastikan data terbersih atau lakukan preprocessing, seperti menangani nilai yang hilang atau mengubah fitur kategorikal menjadi numerik jika diperlukan.
- b. Proses pertama dalam membangun *Random Forest* adalah *bootstrapping*, yaitu mengambil subset acak dari data pelatihan dengan pengembalian (artinya, beberapa data bisa terpilih lebih dari sekali, beberapa mungkin tidak terpilih). Setiap pohon keputusan akan dilatih menggunakan subset acak ini, yang meningkatkan keragaman antar pohon dan mengurangi *overfitting*.
- c. Pada setiap pohon, pembelahan (*split*) data dilakukan berdasarkan subset acak dari fitur yang ada. Dengan cara ini, setiap pohon menggunakan subset acak dari data dan fitur, yang membuat pohon-pohon lebih beragam. Pemisahan (*Splitting*) dilakukan menggunakan kriteria seperti *Gini impurity* atau *Entropy* untuk klasifikasi, atau *Mean Squared Error* (MSE) untuk regresi.

#### 1. Gini Impurity

*Gini impurity* mengukur seberapa sering elemen yang dipilih secara acak dari dataset akan salah klasifikasi. Semakin kecil nilai Gini, semakin baik pemisahan yang dilakukan pada node tersebut. Rumus *Gini impurity* untuk *node t* adalah:

$$Gini(t) = 1 - \sum_{i=1}^k p_i^2 \quad (2)$$

## 2. Entropy

Entropy adalah ukuran ketidakpastian atau keacakan dalam sebuah *node*. Semakin kecil nilai *entropy*, semakin homogen data dalam *node* tersebut. Rumus *entropy* untuk *node t* adalah:

$$Entropy(t) = - \sum_{i=1}^k p_i \log_2(p_i) \quad (3)$$

## 3. Mean Squared Error (MSE)

Untuk regresi, *Random Forest* menggunakan MSE sebagai kriteria untuk pemisahan. MSE mengukur rata-rata kuadrat selisih antara nilai yang diprediksi oleh model dan nilai yang sebenarnya. Rumus MSE adalah:

$$MSE(t) = \frac{1}{n_t} \sum_{i=1}^{n_t} (y_i - \hat{y}_t)^2 \quad (4)$$

- d. Setelah semua pohon dilatih, untuk klasifikasi, prediksi akhir diambil berdasarkan mayoritas *voting* dari semua pohon. Setiap pohon memberikan prediksi kelas, dan kelas yang paling sering dipilih menjadi hasil prediksi model. Untuk regresi, hasil prediksi dihitung dengan cara mengambil rata-rata dari prediksi yang diberikan oleh masing-masing pohon.
- e. Evaluasi model dilakukan dengan menggunakan metrik seperti akurasi, *precision*, *recall*, *F1-score* untuk klasifikasi, atau *Mean Squared Error* (MSE) dan R2 untuk regresi.
- f. Proses *tuning* dilakukan untuk menemukan hyperparameter yang optimal, seperti:
  - a. Jumlah pohon (*n\_estimators*)
  - b. Maksimum kedalaman pohon (*max\_depth*).
  - c. Jumlah fitur yang dipertimbangkan untuk *split* (*max\_features*).
 Penggunaan *cross-validation* dapat membantu menentukan nilai parameter yang optimal untuk menghindari *overfitting* atau *underfitting*.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Analisa Masalah

Salah satu penyakit kronis yang sering dijumpai khususnya di Indonesia adalah penyakit Gagal Ginjal. Penyakit ini memiliki kondisi jika ginjal tidak lagi dapat menyaring terhadap cairan berbahaya yang masuk dalam tubuh melalui darah. Penyakit ini termasuk dalam jenis penyakit serius yang dapat menyebabkan terhadap komplikasi penyakit lainnya. Peran penting ginjal pada tubuh untuk melakukan penyaringan atau penetralan cairan yang masuk dalam tubuh, sehingga kiranya perlu dilakukan proses pendeteksi terhadap penyakit gagal ginjal tersebut. Secara umum penyakit tersebut memberikan dampak yang serius terhadap kesehatan fisik, mental ataupun sosial bagi penderita penyakit gagal ginjal tersebut. Maka dari itu, kiranya perlu dilakukan pendekatan untuk penyelesaian terhadap deteksi penyakit gagal ginjal. Sekarang ini, proses diagnosa dapat dibantu dengan menggunakan peranan teknologi informasi. Dengan adanya teknologi informasi akan mempermudah dalam proses diagnosa yang dilakukan. Salah satu tahapan atau cara yang berkaitan dengan penggunaan teknologi informasi adalah *data mining*. *Data mining* merupakan cara untuk melakukan pengolahan data yang sangat besar sehingga didapatkan sebuah informasi ataupun pola yang terkandung dalam kumpulan data tersebut. Pada *data mining* tersebut, terdapat beberapa teknik yang dilakukan salah satunya adalah teknik klasifikasi. Pada klasifikasi data mining terdapat algoritma yang dapat digunakan yaitu Algoritma *K-Nearest Neighbor* dan Algoritma *Random Forest*. Perbandingan dari kedua algoritma bertujuan untuk mencari algoritma yang memiliki kinerja lebih baik untuk dapat dipergunakan dalam proses diagnosa yang dilakukan. Sebelum dilakukan proses penelitian, terlebih dahulu diketahui terhadap sampel terhadap data yang akan digunakan. Pada dataset penyakit gagal ginjal tersebut terdapat 24 atribut dengan 2 nilai kelas dan memiliki 400 *record data*. Adapun untuk atribut pada dataset dapat dilihat berikut:

**Tabel 1.** Data Atribut dan Kelas Dataset

No	Nama Atribut / Kelas	Jenis	Keterangan
1	Age	Atribut	Umur
2	BP	Atribut	Blood Pressure
3	SG	Atribut	Specific Gravity
4	Al	Atribut	Albumin
5	Su	Atribut	Sugar
6	RBC	Atribut	Red Bloods Cell
7	PC	Atribut	Pus Cell
8	PCC	Atribut	Pus Cell Pumps
9	Ba	Atribut	Bacteria
10	BGR	Atribut	Blood Glucose Random
11	BU	Atribut	Blood Urea
12	SC	Atribut	Serum Creatinine
13	Sod	Atribut	Sodium



No	Nama Atribut / Kelas	Jenis	Keterangan
14	Pot	Atribut	Potassium
15	Hemo	Atribut	Hemoglobin
16	PCV	Atribut	Packed Cell Volume
17	WBCC	Atribut	White Blood Cell Count
18	RBCC	Atribut	Red Blood Cell Count
19	Htn	Atribut	Hypertension
20	DM	Atribut	Diabetes Mellitus
21	CAD	Atribut	Coronary Arteri Disease
22	Appet	Atribut	Appetite
23	PE	Atribut	Pedal Edema
24	Ane	Atribut	Anemia
25		Kelas	

Setelah diketahui jenias atribut seperti pada Tabel 1 di atas, maka selanjutnya untuk mengatahui sampel data pada dataset. Dimana sampel data dapat dilihat di bawah ini:

**Tabel 2.** Sampel Dataset Gagal Ginjal

No	age	bp	sg	Al	su	rbc	pc	pcc	ba
1	Ederly	Normal	Normal	High	Low	normal	normal	notpresent	notpresent
2	Children	Normal	Normal	Low	Low	normal	normal	notpresent	notpresent
3	Ederly	Normal	Normal	High	Low	normal	normal	notpresent	notpresent
4	Ederly	Normal	Normal	Low	Low	normal	abnormal	present	notpresent
5	Ederly	Normal	Normal	High	Low	normal	normal	notpresent	notpresent
6	Ederly	Normal	Normal	Low	Low	abnormal	normal	notpresent	notpresent
7	Ederly	Normal	Normal	High	Low	abnormal	normal	notpresent	notpresent
8	Tennager	Normal	Normal	High	Normal	normal	abnormal	notpresent	notpresent
9	Ederly	Normal	Normal	Low	Low	normal	abnormal	present	notpresent
10	Ederly	Normal	Normal	High	Low	abnormal	abnormal	present	notpresent
11	Ederly	Normal	Normal	High	Normal	abnormal	abnormal	present	notpresent
12	Ederly	Normal	Normal	Low	Low	abnormal	abnormal	present	notpresent
13	Ederly	Normal	Normal	Low	Low	abnormal	normal	present	notpresent
14	Ederly	Normal	Low	High	Low	abnormal	normal	notpresent	notpresent
15	Ederly	Normal	Normal	Low	Low	normal	abnormal	present	present
...	...	...	...	...	...	...	...	...	...
400	Ederly	Normal	Normal	High	Low	normal	normal	notpresent	notpresent

**Tabel 2.** Lanjutan Sampel Dataset Gagal Ginjal

No	bgr	bu	sc	sod	pot	Hemo	pcv	wc	rc
1	Normal	Normal	Normal	Low	Low	Normal	Normal	Normal	Normal
2	Normal	Normal	Normal	Low	Low	Low	Low	Normal	Low
3	High	High	High	Low	Low	Low	Low	Normal	Low
4	Normal	High	High	Low	Low	Low	Low	Normal	Low
5	Normal	Normal	High	Low	Low	Low	Low	Normal	Normal
6	Normal	Normal	Normal	Normal	Low	Low	Low	Normal	Normal
7	Normal	High	High	Low	Low	Low	Low	Low	Low
8	High	Normal	Normal	Low	Low	Low	Normal	Normal	Normal
9	Normal	High	High	Low	Low	Low	Low	Normal	Low
10	Normal	High	High	Low	Low	Low	Low	High	Low
11	High	High	High	Low	Low	Low	Low	Low	Low
12	High	High	High	Low	Low	Low	Low	Low	Low
13	High	High	High	Normal	Low	Low	Low	High	Low
14	Normal	High	High	Normal	Low	Low	Low	Low	Low
15	Normal	High	High	Low	Normal	Low	Low	Normal	Low
...	...	...	...	...	...	...	...	...	...
400	Normal	Normal	Normal	Normal	Low	Normal	Normal	Normal	High

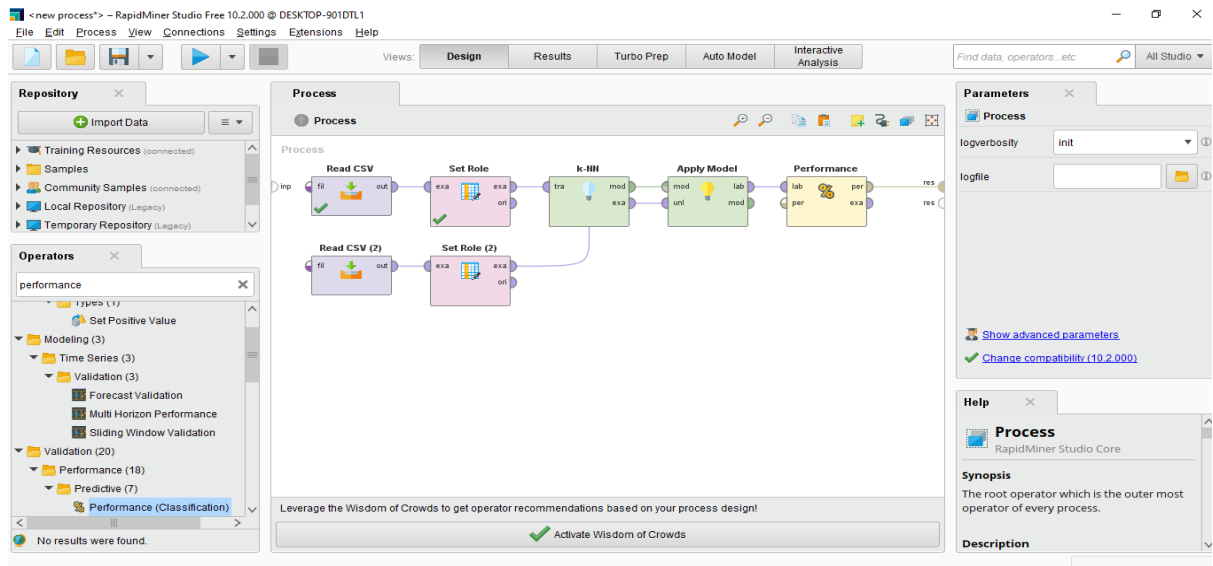
**Tabel 2.** Lanjutan Sampel Dataset Gagal Ginjal

No	htn	dm	cad	appet	pe	ane	classification
1	yes	yes	no	good	no	no	ckd
2	no	no	no	good	no	no	ckd

No	htn	dm	cad	appet	pe	ane	classification
3	no	yes	no	poor	no	yes	ckd
4	yes	no	no	poor	yes	yes	ckd
5	no	no	no	good	no	no	ckd
6	yes	yes	no	good	yes	no	ckd
7	no	no	no	good	no	no	ckd
8	no	yes	no	good	yes	no	ckd
9	yes	yes	no	good	no	yes	ckd
10	yes	yes	no	poor	no	yes	ckd
11	yes	yes	no	good	no	yes	ckd
12	yes	yes	no	poor	yes	no	ckd
13	yes	yes	yes	poor	yes	no	ckd
14	yes	yes	yes	poor	yes	no	ckd
15	yes	yes	yes	poor	yes	no	ckd
...	...	...	...	...	...	...	...
400	no	no	no	good	no	no	notckd

### 3.2 Penerapan Algoritma K-NN

Setelah diketahui terhadap sampel dataset penyakit ginjal, maka proses selanjutnya yang dilakukan adalah melakukan proses pengujian terhadap kinerja algoritma K-NN dan *Random Forest*. Proses pertama yang dilakukan adalah melakukan pengujian terhadap algoritma K-NN. Pengujian terhadap algoritma K-NN dilakukan dengan melakukan pengujian terhadap beberapa nilai K. Nilai K yang digunakan pada pengujian yaitu K=5, K=9, K=13, K= 17 dan K=21. Proses pengujian yang dilakukan dengan menggunakan *tools* Rapid Miner. Adapun proses tersebut dapat dilihat berikut:



Gambar 2. Operator Pengujian Algoritma K-NN Pada Rapidminer

Setelah dilakukan penyusunan operator pengujian pada rapidminer, maka selanjutnya dilakukan proses pengujian dengan pengukuran kinerja. Adapun proses – proses pengujian algoritma K-NN dengan nilai K=5, K=9, K=13, K= 17 dan K=21 dapat dilihat pada Gambar 3 berikut:

accuracy: 95.75%

	true ckd	true notckd	class precision
pred. ckd	227	7	97.01%
pred. notckd	10	156	93.98%
class recall	95.78%	95.71%	

Gambar 3. Hasil Pengujian Algoritma K-NN dengan nilai K=5

Pada Gambar 3 dapat dilihat terhadap hasil pengujian kinerja algoritma K-NN dengan nilai K=5. Pada gambar tersebut dapat terlihat bahwasannya nilai akurasi yang didapatkan sebesar **95,75%**.

accuracy: 95.00%

	true ckd	true notckd	class precision
pred. ckd	228	11	95.40%
pred. notckd	9	152	94.41%
class recall	96.20%	93.25%	

**Gambar 4.** Hasil Pengujian Algoritma K-NN dengan nilai K=9

Pada Gambar 4 dapat dilihat terhadap hasil pengujian kinerja algoritma K-NN dengan nilai K=9. Pada gambar tersebut dapat terlihat bahwasannya nilai akurasi yang didapatkan sebesar **95,00%**.

accuracy: 94.50%

	true ckd	true notckd	class precision
pred. ckd	228	13	94.61%
pred. notckd	9	150	94.34%
class recall	96.20%	92.02%	

**Gambar 5.** Hasil Pengujian Algoritma K-NN dengan nilai K=13

Pada Gambar 5 dapat dilihat terhadap hasil pengujian kinerja algoritma K-NN dengan nilai K=13. Pada gambar tersebut dapat terlihat bahwasannya nilai akurasi yang didapatkan sebesar **94,50%**.

accuracy: 93.25%

	true ckd	true notckd	class precision
pred. ckd	226	16	93.39%
pred. notckd	11	147	93.04%
class recall	95.36%	90.18%	

**Gambar 6.** Hasil Pengujian Algoritma K-NN dengan nilai K=17

Pada Gambar 6 dapat dilihat terhadap hasil pengujian kinerja algoritma K-NN dengan nilai K=17. Pada gambar tersebut dapat terlihat bahwasannya nilai akurasi yang didapatkan sebesar **93,25%**.

accuracy: 93.75%

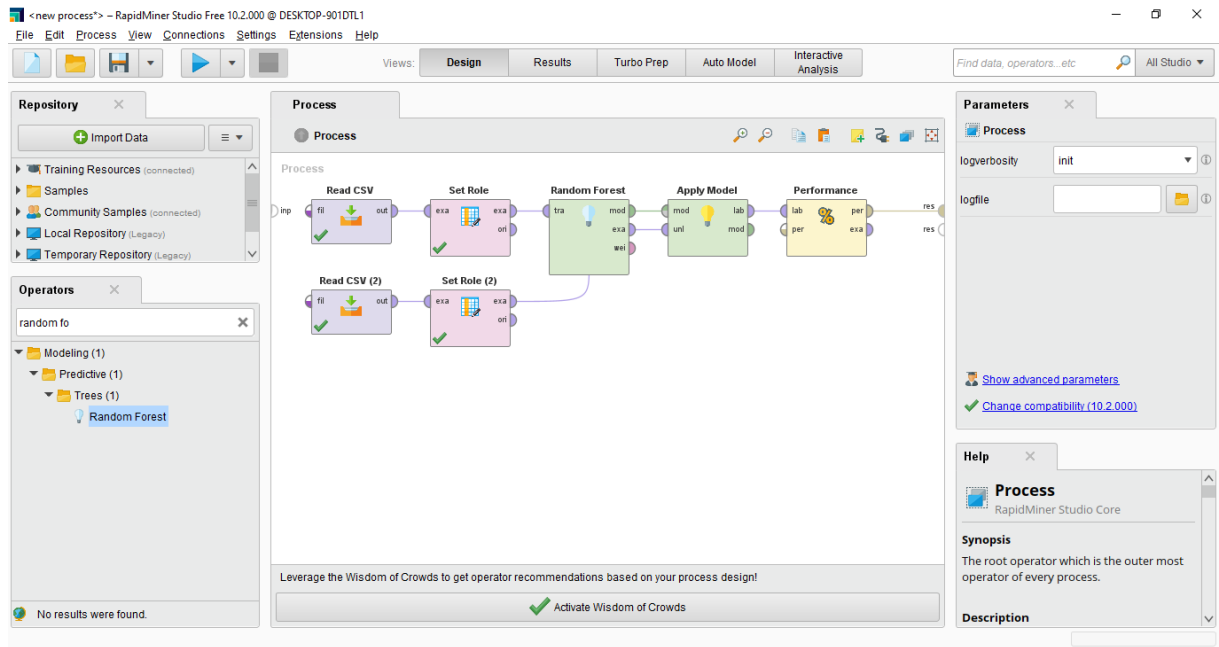
	true ckd	true notckd	class precision
pred. ckd	227	15	93.80%
pred. notckd	10	148	93.67%
class recall	95.78%	90.80%	

**Gambar 7.** Hasil Pengujian Algoritma K-NN dengan nilai K=21

Pada Gambar 7 dapat dilihat terhadap hasil pengujian kinerja algoritma K-NN dengan nilai K=21. Pada gambar tersebut dapat terlihat bahwasannya nilai akurasi yang didapatkan sebesar **93,75%**.

### 3.3 Penerapan Algoritma Random Forest

Selanjutnya, proses pengujian yang dilakukan setelah dilakukan pengujian dengan algoritma K-NN yaitu melakukan pengujian dengan menggunakan algoritma *Random Forest*. Proses pengujian yang dilakukan dengan menggunakan *tools* Rapid Miner. Adapun proses tersebut dapat dilihat berikut:



**Gambar 8.** Operator Pengujian Algoritma K-NN Pada Rapidminer

Setelah dilakukan penyusunan operator pengujian pada rapidminer, maka selanjutnya dilakukan proses pengujian dengan pengukuran kinerja. Adapun proses – proses pengujian algoritma random forest dapat dilihat pada gambar berikut:

**accuracy: 99.75%**

	true ckd	true notckd	class precision
pred. ckd	237	1	99.58%
pred. notckd	0	162	100.00%
class recall	100.00%	99.39%	

**Gambar 9.** Hasil Pengujian Algoritma Random Forest

Pada Gambar 9 dapat dilihat terhadap hasil pengujian kinerja algoritma Random. Pada gambar tersebut dapat terlihat bahwasannya nilai akurasi yang didapatkan sebesar **99,75%**.

### 3.4 Pembahasan

Setelah dilakukan serangkaian proses pengujian, maka selanjutnya dilakukan proses pembahasan. Pembahasan bertujuan untuk melihat hasil akhir dari proses perbandingan kinerja algoritma. Adapun hasil tersebut dapat dilihat pada Tabel 3 berikut:

**Tabel 3.** Hasil Pengujian Kinerja Algoritma

No	Algoritma	Akurasi
1	Algoritma K-NN K=5	95,75%
2	Algoritma K-NN K=9	95,00%
3	Algoritma K-NN K=13	94,50%
4	Algoritma K-NN K=17	93,25%
5	Algoritma K-NN K=21	93,75%
6	Algoritma Random Forest	99,75%

Dari Tabel 3 di atas dapat diliha secara keseluruhan hasil pengujian terhadap kinerja algoritma. Pada tabel tersebut secara keseluruhan nilai akurasi yang didapatkan diatas 90%, namun algoritma *Random Forest* memiliki kinerja yang lebih baik. Dimana hasil tingkat akurasi yang didapatkan dari algoritma *Random Forest* sebesar 99,75%. Maka dari itu, model ataupun pola yang dihasilkan dari algoritma *Random Forest* nantinya dipergunakan untuk membantu dalam proses diagnosa penyakit gagal ginjal dan algoritma *Random Forest* merupakan algoritma yang memiliki kinerja lebih baik.

#### 4. KESIMPULAN

Berdasarkan proses yang telah dilakukan untuk mempermudah dalam proses pemahaman hasil penelitian maka diperlukan penyusunan kesimpulan. Adapun kesimpulan yang didapatkan dari proses penelitian yaitu dengan adanya penggunaan teknologi informasi dan data mining membantu untuk mempermudah dalam proses pengolahan data yang besar pada dataset penyakit gagal ginjal. Dari proses yang telah dilakukan secara keseluruhan nilai akurasi yang didapatkan diatas 90%, namun algoritma *Random Forest* memiliki kinerja yang lebih baik. Dimana hasil tingkat akurasi yang didapatkan dari algoritma *Random Forest* sebesar 99,75%. Maka dari itu, model ataupun pola yang dihasilkan dari algoritma *Random Forest* nantinya dipergunakan untuk membantu dalam proses diagnosa penyakit gagal ginjal dan algoritma *Random Forest* merupakan algoritma yang memiliki kinerja lebih baik.

#### REFERENCES

- [1] W. A. S, R. A. Ramadhan, and R. S. I. S, “Komparasi Algoritma Machine Learning dalam Memprediksi Penyakit Gagal Ginjal,” *J. Penelit. dan Karya Ilm.*, vol. 1, no. Desember, pp. 363–374, 2023.
- [2] Safuan Safuan, “Deteksi Penyakit Gagal Ginjal Kronis Menggunakan Algoritma ID3,” *Elkom J. Elektron. dan Komput.*, vol. 13, no. 1, pp. 8–17, 2020, doi: 10.51903/elkom.v13i1.136.
- [3] I. Wisnuadji Gamadarena and I. Waspada, “IMPLEMENTASI DATA MINING UNTUK DETEKSI PENYAKIT GINJAL KRONIS (PGK) MENGGUNAKAN K-NEAREST NEIGHBOR (KNN) DENGAN BACKWARD ELIMINATION,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, pp. 417–426, 2020, doi: 10.25126/jtiik.202071896.
- [4] A. K. Wijaya, R. Syifa, I. N. Rahmadianto, and R. K. Hapsari, “Identifikasi Penyakit Ginjal Kronis Menggunakan Algoritma K-Nearest Neighbour (k-NN),” *Semin. Nas. Inform. Bela Negara*, vol. 4, pp. 105–109, 2024.
- [5] P. A. R. Devi, “KLASIFIKASI PENYAKIT GAGAL GINJAL KRONIS DENGAN METODE KNN (STUDI KASUS RS DI KAB GRESIK),” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1739–1748, 2024.
- [6] I. Yulianti, R. A. Saputra, M. S. Mardiyanto, and A. Rahmawati, “Optimasi Akurasi Algoritma C4.5 Berbasis Particle Swarm Optimization dengan Teknik Bagging pada Prediksi Penyakit Ginjal Kronis,” *Techno.Com*, vol. 19, no. 4, pp. 411–421, 2020, doi: 10.33633/tc.v19i4.3579.
- [7] B. A. Candra Permana and I. K. Dewi Patwari, “Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes,” *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63–69, 2021, doi: 10.29408/jit.v4i1.2994.
- [8] M. Rizal, M. Z. Syahaf, S. R. Priyambodo, and Y. Rhamdani, “Optimasi Algoritma Naïve Bayes Menggunakan Forward Selection Untuk Klasifikasi Penyakit Ginjal Kronis,” *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 5, no. 1, pp. 71–80, 2023, doi: 10.53580/naratif.v5i1.200.
- [9] N. Fatimah Indrianti, A. Kania Ningsih, and R. Ilyas, “Implementasi Data Mining Untuk Klasifikasi Penyakit Gagal Ginjal Kronis Menggunakan Metode K-Nearest Neighbor,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2255–2260, 2024, doi: 10.36040/jati.v8i2.9464.
- [10] M. Syahrizal, S. Aripin, D. P. Utomo, Mesran, Sarwandi, and N. A. Hasibuan, “The application of the K-NN imputation method for handling missing values in a dataset,” *AIP Conf. Proc.*, vol. 2048, no. 1, 2024.
- [11] Mesran, M. Syahrizal, Sarwandi, S. Aripin, D. P. Utomo, and A. Karim, “A comparison of the performance of data mining classification algorithms on medical datasets with the application of data normalization,” *AIP Conf. Proc.*, vol. 3048, no. 1, 2024.
- [12] D. P. Utomo, Mesran, Sarwandi, S. Aripin, M. Syahrizal, and Pristiwanto, “A comparative analysis of data normalization on data mining classification performance,” *AIP Conf. Proc.*, vol. 3048, no. 1, 2024.
- [13] A. D. W. M. Sidik, I. Himawan Kusumah, A. Suryana, Edwinanto, M. Artiyasa, and A. Pradiftha Junfithrana, “Gambaran Umum Metode Klasifikasi Data Mining,” *Fidel. J. Tek. Elektro*, vol. 2, no. 2, pp. 34–38, 2020, doi: 10.52005/fidelity.v2i2.111.
- [14] N. B. Putri and A. W. Wijayanto, “Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing,” *Komputika J. Sist. Komput.*, vol. 11, no. 1, pp. 59–66, 2022, doi: 10.34010/komputika.v11i1.4350.
- [15] Y. Azhar, A. K. Firdausy, and P. J. Amelia, “Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke,” *SINTECH (Science Inf. Technol. J.*, vol. 5, no. 2, pp. 191–197, 2022, doi: 10.31598/sintechjournal.v5i2.1222.
- [16] S. Rahayu and J. J. Purnama, “Klasifikasi Konsumsi Energi Industri Baja Menggunakan Teknik Data Mining,” *J. Teknoinfo*, vol. 16, no. 2, p. 395, 2022, doi: 10.33365/jti.v16i2.1984.
- [17] W. Irmayani, “Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi,” *J. Khatulistiwa Inform.*, vol. IX, no. 1, pp. 68–72, 2021.
- [18] I. S. Muallif, H. Budiman, and N. Ransi, “Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma K-Nearest Neighbor,” in *PROSIDING SEMINAR NASIONAL PEMANFAATAN SAINS DAN TEKNOLOGI INFORMASI*, 2023, vol. 1, no. 1, pp. 297–306.
- [19] D. Cahyanti, A. Rahmayani, and S. A. Husniar, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [20] M. A. Muzani, M. I. A. Sukri, S. N. Fauziah, A. Fatkhurohman, and D. Ariantanto, “Data Mining Untuk Klasifikasi Produk Menggunakan Algoritma K-Nearest Neighbor Pada Toko Online,” *Pros. SISFOTEK ISSN 2597-3584*, vol. 5, no. 1, pp. 141–145, 2021, [Online]. Available: <http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/273>.
- [21] S. Munawaroh, U. A. Rosyidah, and R. Yanuarti, “Klasifikasi Tingkat Kecemasan Atlet Sebelum Bertanding Menggunakan Algoritma K-Nearest Neighbor (KNN) Berbasis Website,” *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 5, no. 2, pp. 87–94, 2024.
- [22] A. Oktaviana, D. P. Wijaya, A. Pramuntadi, and D. Heksaputra, “Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (K-NN),” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 812–818, 2024, doi: 10.57152/malcom.v4i3.1268.
- [23] U. Nijunnihayah, S. S. Hilabi, F. Nurapriani, and E. Novalia, “Implementasi Algoritma K-Nearest Neighbor untuk Prediksi



- Penjualan Alat Kesehatan pada Media Alkes,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 695–701, 2024, doi: 10.57152/malcom.v4i2.1326.
- [24] S. A. Fitria, H. Oktafia, L. Wijaya, and D. Irawan, “Data Mining Classification Untuk Prediksi Jumlah Mahasiswa Aktif dan Cuti Angkatan 2020 Menggunakan Metode K-Nearest Neighbor,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 5, pp. 2637–2645, 2024, doi: 10.30865/klik.v4i5.1795.
- [25] G. A. Sandag, “Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest,” *CogITo Smart J.*, vol. 6, no. 2, pp. 167–178, 2020, doi: 10.31154/cogito.v6i2.270.167-178.
- [26] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, “Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah,” *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [27] L. Sari, A. Romadloni, and R. Listyaningrum, “Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest,” *Infotekmesin*, vol. 14, no. 1, pp. 155–162, 2023, doi: 10.35970/infotekmesin.v14i1.1751.
- [28] K. G. Putra, “Penentuan Penerima Bantuan Program Keluarga Harapan menggunakan Algoritma Random Forest di Desa Kebonrejo,” *Innov. J. Soc. Sci. Res. Vol.*, vol. 4, no. 5, pp. 7242–7257, 2024.
- [29] L. Hidayah and M. I. Rosadi, “Penerapan Algoritma Random Forest Untuk Memprediksi Jumlah Santri Baru,” *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 5, no. 1, pp. 49–54, 2024.
- [30] M. R. A. Putra and Rissa Nurfitriana Handayani, “Perbandingan Algoritma Decision Tree dan Random Forest Dalam Pengklasifikasian Penyakit Tiroid,” in *E-PROSIDING SISTEM INFORMASI*, 2024, vol. 5, no. 2, pp. 166–172.
- [31] D. C. P. Buani, “Deteksi Dini Penyakit Diabetes dengan Menggunakan Algoritma Random Forest,” *EVOLUSI J. Sains dan Manaj.*, vol. 12, no. 1, pp. 1–8, 2024, doi: 10.31294/evolusi.v12i1.21005.
- [32] S. Widaningsih and S. Yusuf, “Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2598–2611, 2022, doi: 10.35957/jatisi.v9i3.859.
- [33] S. Ucha Putri, E. Irawan, and F. Rizky, “Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5,” *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 2, no. 1, pp. 39–46, 2021.
- [34] H. P. Doloksaribu and Y. T. Samuel, “KOMPARASI ALGORITMA DATA MINING UNTUK ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI,” *J. Teknol. Inf.*, vol. 16, no. 1, pp. 1–11, 2022.
- [35] A. R. Wibowo and A. Jananto, “Implementasi Data Mining Metode Asosiasi Algoritma FP-Growth Pada Perusahaan Ritel,” *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 200, 2020, doi: 10.35585/inspir.v10i2.2585.
- [36] F. Y. Rahman, I. I. Purnomo, and N. Hijriana, “PENERAPAN ALGORITMA DATA MINING UNTUK KLASIFIKASI KUALITAS AIR,” *Technologia*, vol. 13, no. 3, pp. 228–232, 2022.
- [37] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTISI>.
- [38] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, “Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means,” *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.