



## OPTIMASI ALGORITMA GENETIKA UNTUK MEMPREDIKSI PENYAKIT HEPATITIS DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR (KNN)

Rahmi Nurul Arapah<sup>1</sup>, Fitriyani<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya

[rahminurul1933@gmail.com](mailto:rahminurul1933@gmail.com)<sup>1</sup>, [fitriyani@gmail.com](mailto:fitriyani@gmail.com)<sup>2</sup>

### Abstract

*Health is the most important part for humans, and every individual certainly wants to have a healthy body. The liver is one of the key organs in the human body, and one cannot live without it. Liver disease, or what is often known as hepatitis, has various types and is one of the most dangerous. If not detected early, hepatitis can develop into liver cancer. This study aims to evaluate the results of genetic algorithm optimization in predicting hepatitis using the K-Nearest Neighbor (KNN) method. The main objective of this study is to detect hepatitis in patients early so that treatment can be carried out to reduce the risk of chronic hepatitis. The data used in this study comes from secondary data obtained from Kaggle. Data testing was carried out using the Rapidminer application. Based on the results of the study, it was concluded that the optimization technique using a genetic algorithm succeeded in increasing the prediction accuracy of the K-Nearest Neighbor (KNN) algorithm in identifying hepatitis. Before optimization, the K-Nearest Neighbor (KNN) algorithm had an accuracy of 73.81%. However, after optimization with the genetic algorithm, the accuracy increased to 95.24%, representing an increase of 21.43% from the initial accuracy without the genetic algorithm.*

*Keywords: Genetic Algorithm, Hepatitis, K-Nearest Neighbor (KNN), Optimization*

### Abstrak

Kesehatan merupakan bagian terpenting bagi manusia, dan setiap individu tentu ingin memiliki tubuh yang sehat. Hati merupakan salah satu organ kunci dalam tubuh manusia, dan seseorang tidak dapat hidup tanpa hati. Penyakit hati, atau yang sering dikenal sebagai hepatitis, memiliki berbagai jenis dan termasuk salah satu yang berbahaya. Jika tidak terdeteksi secara dini, penyakit hepatitis dapat berkembang menjadi kanker hati. Penelitian ini bertujuan untuk mengevaluasi hasil optimasi algoritma genetika dalam memprediksi penyakit hepatitis dengan menggunakan metode K-Nearest Neighbor (KNN). Tujuan utama dari penelitian ini adalah untuk mendeteksi penyakit hepatitis pada pasien secara dini sehingga penanganan dapat dilakukan untuk mengurangi



risiko hepatitis kronis. Data yang digunakan dalam penelitian ini bersumber dari data sekunder yang diperoleh dari Kaggle. Pengujian data dilakukan menggunakan aplikasi Rapidminer. Berdasarkan hasil penelitian, diperoleh kesimpulan bahwa optimasi teknik dengan menggunakan algoritma genetika berhasil meningkatkan akurasi prediksi algoritma K-Nearest Neighbor (KNN) dalam mengidentifikasi penyakit hepatitis. Sebelum dioptimasi, algoritma K-Nearest Neighbor (KNN) memiliki akurasi sebesar 73.81%. Namun, setelah dioptimasi dengan algoritma genetika, akurasi meningkat menjadi 95.24%, menunjukkan peningkatan sebesar 21.43% dari akurasi awal tanpa algoritma genetika.

Kata kunci: Genetic Algorithm, Hepatitis, K-Nearest Neighbor (KNN), Optimization

## 1. PENDAHULUAN

Kesejahteraan manusia sangat tergantung pada kesehatannya. Setiap individu mendambakan keadaan tubuh yang prima. Mewujudkan hak untuk hidup sehat bagi semua orang juga merupakan elemen penting dari kesejahteraan penduduk [1].

Hati adalah salah satu organ yang sangat penting, penting dalam tubuh manusia. Organ vital bagian dalam badan ini juga disebut kelenjar eksokrin. Karena memiliki fungsi mengeluarkan empedu di usus. Namun semua fungsi dan manfaatnya, hati tidak bisa menghindari masalah dapat menyebabkan penyakit. Penyakit hati adalah peradangan yang disebabkan oleh infeksi virus, bakteri atau zat beracun, sehingga hati tidak menjalankan fungsinya dengan baik, dan tidak mudah ditemukan pada tahap awal dalam mendiagnosis penyakit hati. Penyakit hati atau yang biasanya disebut liver memiliki berbagai macam jenis, salah satunya yang berbahaya dan sering terjadi adalah hepatitis [2].

Hepatitis merupakan gangguan pada organ hati yang mengakibatkan peradangan pada sel-sel hati. Kelainan ini disebabkan oleh gangguan atau kerusakan pada membran hati. Terdapat dua faktor penyebab utama, yaitu faktor infeksi dan faktor non-infeksi. Faktor infeksi melibatkan virus hepatitis dan bakteri yang dapat ditemukan di dalam hati. Peradangan ini dapat diidentifikasi melalui peningkatan kadar enzim hati. Diagnosis awal penyakit ini

---

<sup>1</sup> Orin Nuraeni et al., "Sistem Pakar Diagnosa Kondisi Gigi Tiruan Menggunakan Metode Naïve Bayes Classifier" 7, no. 1 (2023): 79–88.

<sup>2</sup> Nugroho, "Optimasi Algoritma Genetika Dalam Memprediksi Penyakit Hepatitis Dengan Menggunakan Metode Decision Tree," 2019, <https://ecampus.pelitabangsa.ac.id/pb/AmbilLampiran?ref=24670&jurusan=&jenis=Item&usingId=false&download=false&clazz=ais.database.model.file.LampiranLain>.



biasanya ditegakkan setelah mengamati gejala yang muncul dan melalui tes fungsi hati yang sering disebut sebagai Liver Function Test (LFT). Beberapa atribut dari hasil pemeriksaan ini dapat digunakan secara efektif untuk menganalisis kondisi penyakit tersebut [3].

Pada studi sebelumnya yang telah dilaksanakan oleh [4] dengan menggunakan pendekatan data mining C4.5, hasilnya mencapai tingkat akurasi sekitar 89.29%. Sebaliknya, penelitian yang dilakukan oleh [5] dengan memanfaatkan metode Naïve Bayes mencapai tingkat akurasi sekitar 96.77% dalam memprediksi penyakit hepatitis. Selanjutnya, penelitian [6] menerapkan teknik Principal Component Analysis (PCA) dan Support Vector Machine (SVM) untuk memprediksi penyakit hepatitis dan mencapai tingkat akurasi sekitar 93.54%. Kemudian, penelitian yang dilakukan oleh [7] mencakup penggunaan algoritma K-Nearest Neighbor, Naïve Bayes, dan Neural Network dalam memprediksi penyakit hepatitis, dengan tingkat akurasi berturut-turut sekitar 93%, 72.92%, dan 82.97%. Oleh karena itu, studi ini akan melakukan penelitian dengan memanfaatkan pendekatan data mining menggunakan metode K-Nearest Neighbor yang dioptimalkan melalui Algoritma Genetika.

Oleh karena itu, dari penjelasan di atas, dibutuhkan tindakan yang spesifik untuk mengatasi penyakit hepatitis. Salah satu upaya yang dapat dilakukan adalah melakukan diagnosis awal terhadap penyakit hepatitis, dan salah satu metode untuk mendeteksinya lebih awal adalah dengan memanfaatkan Teknologi Informasi. Sejalan dengan kemajuan dalam ilmu pengetahuan dan teknologi informasi, cabang ilmu yang dikenal sebagai data mining telah menarik perhatian yang signifikan [8].

Data mining adalah perpaduan dari ilmu statistik, kecerdasan buatan (sistem pakar) dan penelitian dalam bidang database, untuk mencapai hal ini, perlu dilakukan proses penyaringan melalui berbagai dataset yang luas atau

---

<sup>3</sup> Lusy Octoria Sitanggung and Nurdin Bahtiar, "Aplikasi Data Mining Untuk Mendeteksi Penyakit Hepatitis Menggunakan Metode Support Vector Machine (SVM) Polynomial (Studi Kasus: Data Pasien Hati India)," *Jurnal Masyarakat Informatika* 10, no. 1 (2019): 20–27, <https://doi.org/10.14710/jmasif.10.1.31490>.

<sup>4</sup> Nugroho, "Optimasi Algoritma Genetika Dalam Memprediksi Penyakit Hepatitis Dengan Menggunakan Metode Decision Tree."

<sup>5</sup> Buani, "Prediksi Penyakit Hepatitis Menggunakan Algoritma Naïve Bayes Dengan Seleksi Fitur Algoritma Genetika," *Evolusi: Jurnal Sains Dan Manajemen* 6, no. 2 (2018): 1–5, <https://doi.org/10.31294/evolusi.v6i2.4381>.

<sup>6</sup> Triyana Fadila, *Prediksi Penyakit Hepatitis Menggunakan Metode Principal Component Analysis Dan Support Vector Machine (Pca-Svm)*, 2020.

<sup>7</sup> Sulastris Sulastris, Kristophorus Hadiono, and Muchamad Taufiq Anwar, "Analisis Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan Menggunakan Algoritma K-Nearest Neighbor, Naïve Bayes Dan Neural Network," *Dinamik* 24, no. 2 (2020): 82–91, <https://doi.org/10.35315/dinamik.v24i2.7867>.

<sup>8</sup> Buani, "Prediksi Penyakit Hepatitis Menggunakan Algoritma Naïve Bayes Dengan Seleksi Fitur Algoritma Genetika."

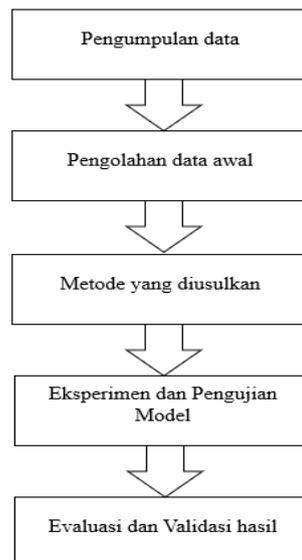


melakukan penelusuran yang cerdas untuk mengidentifikasi data yang memiliki nilai signifikan. Data mining terkait erat dengan beragam disiplin ilmu seperti sistem basis data, penggalian data, analisis statistik, pembelajaran mesin, pencarian informasi, dan komputasi tingkat lanjut [ ].

Tujuan utama dari penelitian ini adalah untuk menginvestigasi pengoptimalan algoritma genetika dalam memprognosis penyakit hepatitis dengan menggunakan pendekatan K-Nearest Neighbor (KNN) dan untuk menilai tingkat akurasi hasilnya.

## 2. METODOLOGI PENELITIAN

Dalam penelitian ini, sumber data yang digunakan berasal dari Kaggle.com. Informasi yang terkandung dalam dataset ini berhubungan dengan kondisi penyakit Hepatitis dan telah diproses menggunakan berbagai teknik data mining untuk menghasilkan metode yang presisi. Metode tersebut nantinya dapat digunakan sebagai panduan dalam melakukan diagnosis penyakit hepatitis dengan penerapan metode K-Nearest Neighbor yang telah



diptimalkan melalui Algoritma Genetika. Dalam rangka penelitian ini, beberapa langkah penelitian dilakukan, antara lain:

### Gambar 1. Tahapan Penelitian

#### 1. Pengumpulan Data

Pada tahap ini ditentukan data yang Akan diproses. Mencari data yang tersedia dan menjelaskan data yang dikumpulkan dan darimana data tersebut diperoleh sehingga dapat digunakan penelitian.

#### 2. Pengolahan data awal

Pada tahap ini dilakukan pemilihan data, data dibersihkan dan ditransformasikan menjadi bentuk yang diinginkan sehingga dapat dilakukan persiapan.

#### 3. Metode yang diusulkan





Teknik validasi yang digunakan menggunakan split validation. Dataset diambil salah satu bagian untuk dijadikan sebagai data *testing* dan untuk yang lainnya sebagai data *training*. Dataset yang digunakan adalah dataset penyakit hepatitis. Untuk perhitungan sebagai berikut:

1. Menentukan nilai k, untuk studi kasusnya nilai k adalah = 3
2. Hitung jarak antara data testing dan data training

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(30 - 32)^2 + (85 - 59)^2} = \sqrt{(-2)^2 + (26)^2} \\ &= \sqrt{4 + 676} = \sqrt{680} = 26,07 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(50 - 32)^2 + (135 - 59)^2} = \sqrt{(18)^2 + (76)^2} \\ &= \sqrt{324 + 5.776} = \sqrt{6.100} = 78,10 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(78 - 32)^2 + (96 - 59)^2} = \sqrt{(46)^2 + (37)^2} \\ &= \sqrt{2.116 + 1.369} = \sqrt{3485} = 59,03 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(34 - 32)^2 + (105 - 59)^2} = \sqrt{(2)^2 + (46)^2} \\ &= \sqrt{4 + 2.116} = \sqrt{2.120} = 46,04 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(34 - 32)^2 + (95 - 59)^2} = \sqrt{(2)^2 + (36)^2} \\ &= \sqrt{4 + 1.296} = \sqrt{1.300} = 36,05 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(51 - 32)^2 + (105 - 59)^2} = \sqrt{(19)^2 + (46)^2} \\ &= \sqrt{361 + 2.116} = \sqrt{2.477} = 49,76 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(23 - 32)^2 + (105 - 59)^2} = \sqrt{(-9)^2 + (46)^2} \\ &= \sqrt{81 + 2.116} = \sqrt{2.197} = 46,87 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(39 - 32)^2 + (105 - 59)^2} = \sqrt{(7)^2 + (46)^2} \\ &= \sqrt{49 + 2.116} = \sqrt{2.165} = 46,52 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(30 - 32)^2 + (105 - 59)^2} = \sqrt{(-2)^2 + (46)^2} \\ &= \sqrt{4 + 2.116} = \sqrt{2.120} = 46,04 \end{aligned}$$

$$\begin{aligned} \text{➤ } d &= \sqrt{(39 - 32)^2 + (78 - 59)^2} = \sqrt{(7)^2 + (19)^2} \\ &= \sqrt{49 + 361} = \sqrt{410} = 20,24 \end{aligned}$$

Berikut adalah hasil perhitungan K-Nearest Neighbor (KNN) yang ditunjukkan pada Tabel 1.

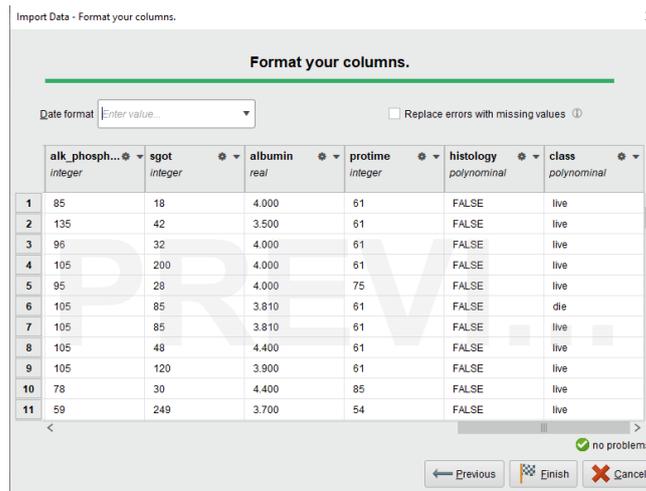


Age	Alk_phosphate	Jarak Euclidean	Urutan Jarak (kecil ke besar)
30	85	26,07	2
50	135	78,10	9
78	96	59,03	8
34	105	46,04	4
34	95	36,05	3
51	105	49,76	7
23	105	46,87	6
39	105	46,52	5
30	105	46,04	4
39	78	20,24	1

**Tabel 1. Hasil Perhitungan K-Nearest Neighbor (KNN)**

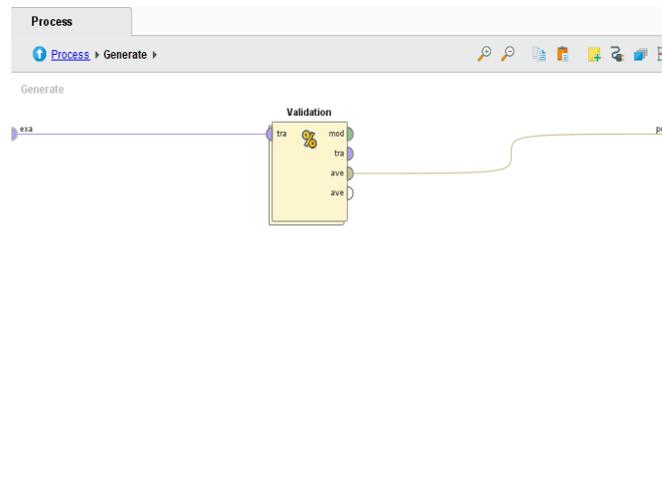
### 3.1.2. Pengujian K-Nearest Neighbor (KNN)

Pertama menggunakan operator Read CSV yang digunakan untuk melakukan import dataset dengan Cara pilih Import Configuration Wizard lalu pilih data yang Akan diolah dan dipilih diagnosa rolenya menjadi label



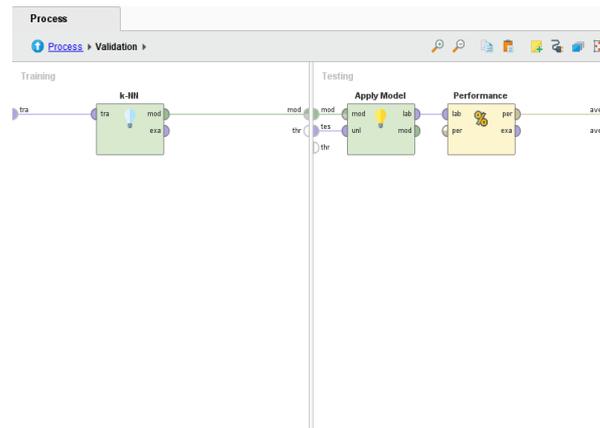
**Gambar 3. Import Dataset**

Pada langkah berikutnya, kita memanfaatkan metode validasi pembagian data (split validation). Validasi ini bertujuan untuk memisahkan dataset menjadi dua bagian sesuai dengan rasio pembagian yang telah ditetapkan sebelumnya. Rasio pembagian yang dipilih akan digunakan sebagai data pelatihan (training), sementara sisi lainnya akan digunakan sebagai data pengujian (testing).



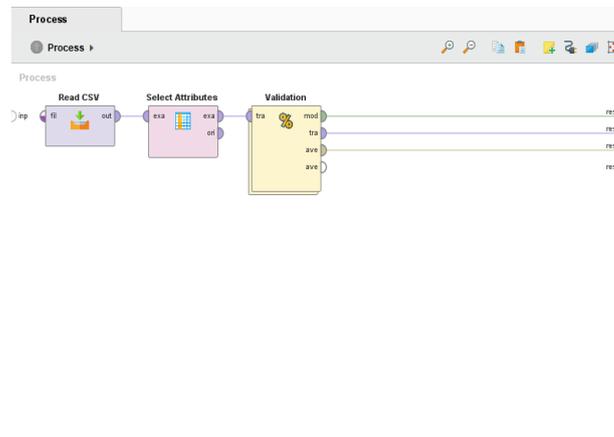
**Gambar 4. Split Validation**

Pada bagian training digunakan algoritma K-Nearest Neighbor (KNN) Dalam fase pengujian, digunakan operator apply model untuk melakukan klasifikasi model pada dataset pengujian, dan operator perform digunakan untuk menampilkan tingkat akurasi.



**Gambar 5. Model Pengujian K-Nearest Neighbor**

Dari proses yang dijalankan maka dapat dilihat hasilnya sebagai berikut:



**Gambar 6. Hasil Accuracy K-Nearest Neighbor**

### 3.1.3. Pengujian K-Nearest Neighbor (KNN) + Algoritma Genetika

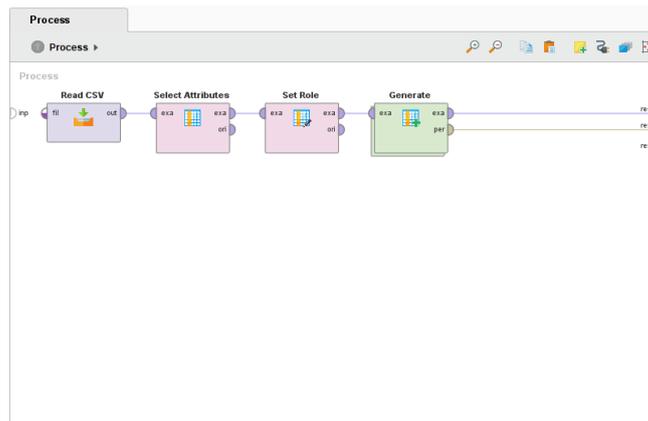
Model awal Pengujian K-Nearest Neighbor (KNN) + Algoritma Genetika sama seperti pengujian K-Nearest Neighbor (KNN), hanya saja pada pengujian K-Nearest Neighbor (KNN) + dalam proses algoritma, terdapat kebutuhan akan operator tambahan yang dikenal sebagai Set role dan Generate. Fungsi operator Set role digunakan untuk mengimplementasikan Algoritma Genetika, sementara operator Generate digunakan dalam konteks metode Algoritma Genetika. Dalam langkah Generate ini, terdapat penerapan Split Validation yang bertujuan untuk membagi data menjadi dua bagian, seperti yang terlihat dalam gambar berikut ini:



accuracy:73.81%

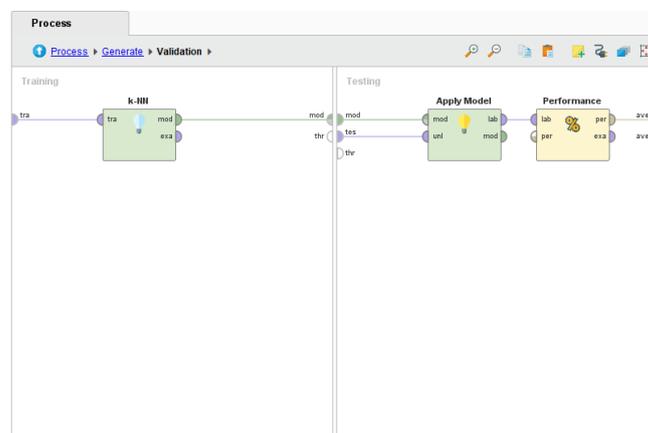
	true live	true die	class precision
pred. live	31	8	79.49%
pred. die	3	0	0.00%
class recall	91.18%	0.00%	

**Gambar 7. Konfirmasi Operator Generate**



**Gambar 8. Split Validation didalam Generate**

Pada Gambar 8 dan Gambar 9 dapat dilihat didalam operator Generate terdapat Split Validation yang sama dengan Split Validation pada pengujian K-Nearest Neighbor (KNN).





### Gambar 9. Model K-Nearest Neighbor + Algoritma Genetika

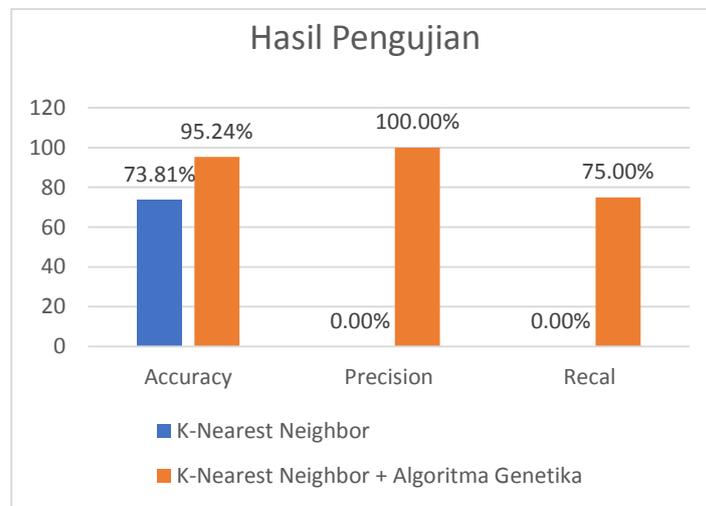
Dari proses yang dijalankan pada Gambar 9 maka dapat dilihat hasilnya sebagai berikut:

accuracy: 95.24%

	true live	true die	class precision
pred. live	34	2	94.44%
pred. die	0	6	100.00%
class recall	100.00%	75.00%	

### Gambar 10. Hasil Accuracy K-Nearest Neighbor + Algoritma Genetika

#### 3.2. Pembahasan



### Gambar 11. Grafik Hasil Pengujian

Pada Gambar 11 dapat dilihat nilai accuracy yang diperoleh dari pengujian K-Nearest Neighbor sebesar 73.81% sedangkan pada pengujian K-Nearest Neighbor + Algoritma Genetika 95.24%, yang artinya nilai accuracy meningkat 21.43%. Pada nilai Precision, hasil yang diperoleh dari pengujian K-Nearest Neighbor 0.00% sedangkan pengujian K-Nearest Neighbor + Algoritma Genetika 100.00%, nilainya meningkat 100.00%. Selanjutnya nilai recal, pada pengujian K-Nearest Neighbor 0.00% sedangkan pengujian K-Nearest Neighbor + Algoritma Genetika 75.00% yang artinya nilai recal meningkat 75.00%.

#### 4. KESIMPULAN

Berdasarkan beberapa hasil pengujian pada penelitian yang telah dilakukan maka dapat disimpulkan bahwa teknik optimasi menggunakan algoritma genetika berhasil meningkatkan akurasi dari algoritma K-Nearest Neighbor (KNN) dalam mendiagnosa penyakit Hepatitis. Hasil akurasi dari algoritma K-Nearest Neighbor (KNN) sebesar 73.81% dan setelah dioptimasi



menggunakan algoritma genetika hasil akurasiya meningkat sebesar 95.24% yang artinya akurasiya meningkat sebesar 21.43% dari algoritma K-Nearest Neighbor (KNN) tanpa Algoritma genetika.

#### DAFTAR PUSTAKA

- [1]. Buani. "Prediksi Penyakit Hepatitis Menggunakan Algoritma Naïve Bayes Dengan Seleksi Fitur Algoritma Genetika." *Evolusi : Jurnal Sains Dan Manajemen* 6, no. 2 (2018): 1–5. <https://doi.org/10.31294/evolusi.v6i2.4381>.
- [2]. Fadila, Triyana. *Prediksi Penyakit Hepatitis Menggunakan Metode Principal Component Analysis Dan Support Vector Machine (Pca-Svm)*, 2020.
- [3]. Nugroho. "Optimasi Algoritma Genetika Dalam Memprediksi Penyakit Hepatitis Dengan Menggunakan Metode Decision Tree," 2019. <https://ecampus.pelitabangsa.ac.id/pb/AmbilLampiran?ref=24670&jurusan=&jenis=Item&usingId=false&download=false&clazz=ais.database.model.file.LampiranLain>.
- [4]. Nuraeni, Orin, Program Studi, Sistem Informasi, Universitas Adhirajasa, Reswara Sanjaya, and Kota Bandung. "Sistem Pakar Diagnosa Kondisi Gigi Tiruan Menggunakan Metode Naïve Bayes Classifier" 7, no. 1 (2023): 79–88.
- [5]. Sitanggang, Lusy Octoria, and Nurdin Bahtiar. "Aplikasi Data Mining Untuk Mendeteksi Penyakit Hepatitis Menggunakan Metode Support Vector Machine (SVM) Polynomial (Studi Kasus: Data Pasien Hati India)." *Jurnal Masyarakat Informatika* 10, no. 1 (2019): 20–27. <https://doi.org/10.14710/jmasif.10.1.31490>.
- [6]. Sulastri, Sulastri, Kristophorus Hadiono, and Muchamad Taufiq Anwar. "Analisis Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan Menggunakan Algoritma K-Nearest Neighbor, Naïve Bayes Dan Neural Network." *Dinamik* 24, no. 2 (2020): 82–91. <https://doi.org/10.35315/dinamik.v24i2.7867>.