# OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN NORMALISASI DAN SELEKSI FITUR UNTUK KLASIFIKASI PENYAKIT LIVER

**Siti Zulaikhah HR, Abdul Aziz, Wahyudi Harianto**
Program Studi Teknik Informatika S1, Fakultas Sains dan Teknologi
Universitas PGRI Kanjuruhan Malang, Jalan S.Supriadi No.48 Malang, Indonesia
*yanti.rukkmana@gmail.com*

**ABSTRAK**

Sulitnya mengenali penyakit *liver* sejak dini menjadi permasalahan yang sering terjadi. Dengan berkembangnya teknologi, saat ini diagnosis penyakit *liver* dapat menggunakan metode *data mining*. Penelitian ini bertujuan untuk mengetahui penerapan seleksi fitur dan normalisasi data untuk mencari model optimasi untuk klasifikasi penyakit *liver* dengan algoritma *K-Nearest Neighbor*. Metode *K-Nearest Neighbor* dipilih karena memiliki prinsip sederhana dan mudah digunakan, tetapi pada beberapa penelitiannya memiliki akurasi relatif rendah. Proses klasifikasi pada penelitian ini dilakukan dengan melakukan pengisian data yang kosong setelah itu dilakukan pembobotan atribut dan normalisasi pada data yang terpilih. Pada penelitian ini nilai akurasi yang paling optimal didapat ketika menggunakan normalisasi *min-max*, dengan seleksi fitur *Information Gain* dan *Gain Ratio* digunakan nilai rata-rata untuk mengisi kekosongan data dengan menggunakan nilai k = 10. Sedangkan pada seleksi fitur *Symmetrical Uncertainty* dapat menggunakan nilai 0 untuk kekosongan data dan nilai *k* = 5.

***Kata kunci:*** *K-Nearest Neighbor (KNN), Optimasi, Normalisasi, Seleksi Fitur*

## 1. PENDAHULUAN

Masalah yang ditimbulkan oleh penyakit *liver* adalah susah mengenali penyakit *liver* sejak dini, bahkan ketika penyakit tersebut sudah menyebar. Diagnosa penyakit *liver* yang lebih awal dapat meningkatkan tindak kelangsungan hidup pasien. Dengan perkembangan teknologi saat ini diagnosa penyakit dapat menggunakan metode *data mining*. Salah satu pengembangan dari *data mining* adalah klasifikasi. Metode klasifikasi banyak digunakan untuk menentukan keputusan sesuai pengetahuan baru yang didapat dari pengolahan data lampau menggunakan perhitungan suatu algoritma [1].

Metode yang digunakan untuk membangun model klasifikasi dalam penyakit liver yaitu *K-Nearest Neighbor* (KNN). Karena pada penelitian ini sebelumnya yang dilakukan oleh [2] yang membandingkan kinerja algoritma *Naïve Bayes*, *Decision Tree* dan *K-Nearest Neighbor* (KNN) menunjukkan bahwa kinerja KNN memiliki nilai akurasi terendah dengan rata-rata 56,7%. Pada penelitian [3] mengusulkan untuk melakukan seleksi fitur dengan cara membuang atribut yang kurang berpengaruh terhadap data yang diujikan sebelum dilakukan klasifikasi dengan metode KNN.

Sedangkan dalam penelitian [4] menggunakan *gain ratio* sebagai dasar dalam pembobotan setiap atribut pada algoritma KNN. Dari penelitian tersebut, algoritma KNN dengan menggunakan seleksi fitur *gain ratio* dinilai lebih intuitif dan mudah dipahami. Hasil yang diperoleh dari penelitian tersebut mendapatkan peningkatan akurasi sebesar 5%.

Penelitian lain juga dilakukan dengan melakukan proses *preprocessing* dengan normalisasi dan mengisi *missing value*. Normalisasi dilakukan pada data yang telah dipilih atributnya berdasarkan proses seleksi fitur dan *missing value* dilakukan dilakukan karena adanya data yang kosong atau tidak memiliki nilai atau informasi pada beberapa atributnya. Oleh karena itu pada penelitian ini berfokus pada peningkatan akurasi pada klasifikasi penyakit liver dengan menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan menggunakan normalisasi pada data yang telah dilakukan seleksi fitur.

## *2.* TINJAUAN PUSTAKA

### 2.1. Klasifikasi

Menurut [5] klasifikasi data mining merupakan penempatan suatu objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi merupakan suatu tipe data yang dianalisis dan dapat membantu untuk menentukan kelas dari sampel data yang ingin diklasifikasikan dan menemukan hubungan antar atribut.

### 2.2. *K-Nearest Neighbor* (KNN)

Perhitungan algoritma KNN menggunakan jarak antara seluruh seluruh data *training* dan data *testing*, dan kemudian data terdekat atau data yang paling banyak memiliki kemiripin akan diambil untuk klasifikasi [6]. Perhitungan algortima KNN menggunakan jarak antara seluruh data *training* dan data *testing*, dan kemudian data terdekat atau data yang paling banyak memiliki kemiripan akan diambil untuk klasifikasi.

### 2.3. Seleksi Fitur

Seleksi fitur merupakan proses pemilihan atribut yang sesuai dari sekumpulan atribut yang banyak, Algoritma memroses data lebih cepat dengan mengurangi atribut atau fitur yang tidak relevan. Oleh karena itu, pemilihan seleksi fitur yang tepat akan menghasilkan hasil klasifikasi yang lebih cepat [7].

### 2.4. *Information Gain*

*Information gain* merupakan salah satu algortima dalam *machine learning* yang populer dan memiliki keunggulan dalam membatasi pentingnya suatu atribut, karena tidak semua atribut dapat digunakan dalam proses klasifikasi sehingga akan membuat kinerja algoritma menjadi tidak efisien.

### 2.5. *Gain Ratio*

Peningkatan dari *information gain* dalam mengoptimalkan nilai yang normalisasi untuk fitur didalam klasifikasi, pada *gain ratio* memerlukan hasil perhitungan *information gain*. Sehingga *gain ratio* memerlukan hasil modifikasi dari perhitungan yang sama seperti *information gain*.

### 2.6. *Symmetrical Uncertainty*

*Symmetrical Uncertainty* merupakan suatu pendekatan untuk mengukur nilai suatu fitur dengan mengukur ketidakpastian simetris yang terkait dengan kelas dan mengkompensasi penyimpangan dalam metode seleksi fitur *information gain* [8].

### 2.7. Normalisasi

Proses normalisasi data dilakukan untuk menyamarkan nilai data apabila terdapat perbedaan yang signifikan pada *range* nilai, sehingga memudahkan dalam penerapannya [9].

### 2.8. *Missing Value*

*Missing value* dapat terjadi karena informasi tentang objek sulit dicari, tidak diberikan atau informasi tersebut memang tidak ada. *Missing value* adalah masalah yang terjadi dalam penelitian. Adanya *missing* pada data yang dianalisis dapat menurunkan keakuratan dan kualitas data pada saat diolah [10].

### 2.9. *Confusion Matrix*

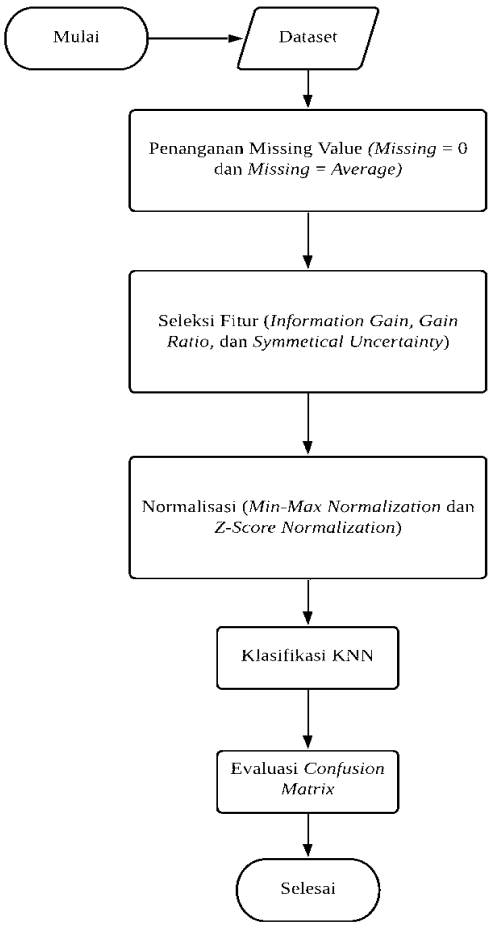
*Confusion matrix* merupakan alat untuk mengevaluasi model klasifikasi untuk memperkirakan objek mana benar atau salah [11]. *Matrix* prediksi akan dibandingkan dengan kelas asli dari data input, yaitu nilai aktual dan prediksi klasifikasi.

### 2.10. 10-*Fold Cross Validation*

Proses ini diulangi hingga 10 kali hingga semua *record* data menjadi bagian dari data uji. Proses ini juga dikenal sebagai 10-*fold cross validation*. Validasi ini telah terbukti memberikan kinerja algoritmik yang lebih stabil, oleh karena itu banyak digunakan oleh para peneliti [11].

## 3. METODE PENELITIAN

Dalam mencari model optimasi yang diusulkan pada penelitian ini untuk meningkatkan akurasi pada *K-Nearest Neighbor* dengan menggunakan penanganan pada *missing value*, normalisasi data dan dilakukan seleksi fitur, maka dilakukan pengujian penelitian dengan tahapan penelitiannnya ditunjukkan pada gambar 1.




Gambar 1. Flowchart Penelitian

### 3.1. Pengambilan Data

Pada pengumpulan data diperoleh dari Website UCI *Machine Learning Repository* (https://archive.ics.uci.edu/ml/index.php) yaitu berupa dataset ILPD (*Indian Liver Patient Dataset*) tahun 2012. Dataset tersebut diolah menggunakan algoritma K-Nearest Neighbor. Dan dataset ini terdapat beberapa atribut, yaitu Age, Gender, Total pada Bilirubin, Direct Bilirubin, ALK (*Alkaline Phosphotase*), SGPT (*Alamine Aminoteransferase*), SGOT (*Aspartate Aminotransferase*), TP (*Total Protein*), ALB (*Albumin)*, A/G (*Ratio Albumin and Globulin Ratio*), Selector (*Class*). Dengan jumlah data sebanyak 583.

### 3.2. Penanganan *Missing Value*

Penanganan *missing value* perlu dilakukan pada data yang akan diteliti, karena adanya missing value dapat menurunkan hasil akurasi pada proses klasifikasi data. Pada penelitian ini dilakukan dua sekenario penanganan *missing*. Pertama, dilakukan dengan mengisi data yang *missing* (kosong) dengan 0 dan sekenario selanjutnya dengan mengganti missing value berdasarkan nilai rata-rata dari seluruh data pada atribut yang terdapat kekosongan data.

### 3.3. Seleksi Fitur

Tahapan selanjutnya yang akan dilakukan adalah seleksi fitur, seleksi fitur dilakukan agar algoritma memroses data lebih cepat. Metode seleksi fitur *Information Gain*, *Gain Ratio*, dan *Symmetrical Uncertainty* akan dilakukan pada penelitian ini dengan tujuan untuk mengurangi atribut yang tidak relevan pada data.

### 3.4. Normalisasi Data

Proses normalisasi data dilakukan untuk menyeimbangkan nilai data. Data yang akan dinormalisasi adalah data yang telah dilakukan proses penanganan missing value. Proses ini dilakukan untuk menyeimbangkan nilai data apabila data memiliki rentang nilai yang jauh dan memudahkan untuk proses klasifikasinya.

### 3.5. Klasifikasi KNN

Tahap klasifikasi KNN dilakukan pada data yang sudah diseleksi fitur dan normalisasi terlebih dahulu yang sudah diproses sebelumnya. Untuk penentuan nilai k yang nantinya akan dipilih sebagai model klasifikasi yang paling optimal dilakukan dengan menguji variasi nilai k menggunakan angka 1 sampai 10. Pengujian ini untuk melihat pengaruh perubahan nilai k terhadap nilai *accuracy*. Serta dilakukan juga pengujian *cross validation* untuk melihat efektifitas kinerja model.

### 3.6. Evaluasi dengan *Confusion Matrix*

Tahap evaluasi dilakukan diakhir proses penelitian. Tahap ini bermanfaat untuk melakukan pengujian model dan nilai akurasi yang dihasilkan. Dalam penelitian ini evaluasi dilakukan dengan *confusion matrix*.

## 4. HASIL DAN PEMBAHASAN

Dalam meningkatkan akurasi algortima *K-Nearest Neighbor* (KNN) dengan normalisasi dan seleksi fitur untuk klasifikasi penyakit *liver* memiliki beberapa tahapan.

### 4.1. Penanganan *Missing Value*

Penanganan *missing value* dilakukan dengan mengisi data yang kosong dengan nilai 0 yang dapat dilihat pada tabel dibawah

Tabel 1. Penanganan *Missing Value* = 0

| *Age* | *Gender* | TB | … | A/G | *Diagnosis* |
|---|---|---|---|---|---|
| 45 | *Female* | 0.9 | … | 0 | *Liver* |
| 51 | *Male* | 0.8 | … | 0 | *Liver* |
| 35 | *Female* | 0.6 | … | 0 | *Non-Liver* |
| 27 | *Male* | 1.3 | … | 0 | *Non-Liver* |

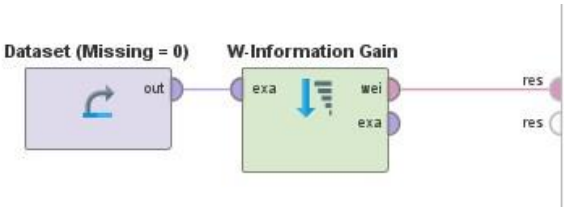
Kemudian dilakukan penangan *missing value* lain dilakukan dengan mengisi nilai rata-rata dari seluruh atribut yang memiliki kekosongan data yang dapat dilihat pada tabel 2 berikut.

Tabel 2. Penanganan *Missing Value = Average*

| *Age* | *Gender* | TB | … | A/G | *Diagnosis* |
|---|---|---|---|---|---|
| 45 | *Female* | 0.9 | … | 0.95 | *Liver* |
| 51 | *Male* | 0.8 | … | 0.95 | *Liver* |
| 35 | *Female* | 0.6 | … | 0.95 | *Non-Liver* |
| 27 | *Male* | 1.3 | … | 0.95 | *Non-Liver* |

### 4.2. Seleksi Fitur *Information Gain*

Seleksi fitur dilakukan setelah melalui proses penangan *missing value*. Dalam proses perhitungan bobot atribut dilakukan dengan menggunakan *RapidMiner Studio* dengan skema pengujian sebagai berikut




Gambar 2. Skema Seleksi Fitur Information Gain

Dari hasil pengujian dengan skema diatas diketahui bahwa atribut DB (*Direct Bilirubin*) memiliki bobot paling besar dengan nilai 0.086 yang berarti bahwa atribut tersebut memiliki pengaruh yang paling besar terhadap dataset, dan begitu seterusnya hingga atribut TP (*Total Protiens*) yang memiliki pengaruh paling kecil pada data set dengan nilai atribut 0.005 yang secara lengkap bisa dilihat pada tabel dibawah ini

Tabel 3. Hasil *Information Gain*

| No | Atribut | *Information Gain* |
|---|---|---|
| 1 | DB | 0.086 |
| 2 | TB | 0.085 |
| 3 | *Sgot Asparte Aminotransferase* | 0.067 |
| 4 | *Alkhos Alkanine Phosphotase* | 0.066 |
| 5 | *Sgpt Alamine Aminotransferase* | 0.060 |
| 6 | A/G | 0.026 |
| 7 | *Age* | 0.021 |
| 8 | ALB | 0.020 |
| 9 | *Gender* | 0.005 |
| 10 | TP | 0.005 |

Pemilihan atribut dilakukan dengan menggunakan persamaan

$$\log_2 10 = 3.32$$

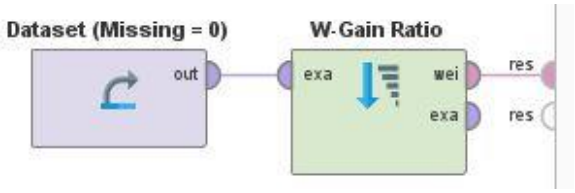
Sehingga dataset *dataset Indian Liver Patient Datasets* memiliki 10 atribut yang selanjutnya diilih 3 atribut tertinggi yang dianggap sebagai atribut paling berpengaruh berdasarkan nilai pembobotan tertinggi untuk dilakukan proses klasifikasi yang dapat dilihat pada tabel 2

Tabel 4. Atribut terilih pada *Information Gain*

| No | Atribut | *Information Gain* |
|---|---|---|
| 1 | DB | 0.086 |
| 2 | TB | 0.085 |
| 3 | *Sgot Asparte Aminotransferase* | 0.067 |

### 4.3. Seleksi Fitur *Gain Ratio*

Pada proses perhitungan bobot atribut dengan seleksi fitur *gain* ratio dilakukan dengan menggunakan *RapidMiner Studio* dengan skema pengujian ditunjukkan pada gambar 3




Gambar 3. Skema seleksi fitur Gain Ratio

Proses pembacaan data menggunakan operator "*Retrieve*" dilakukan dengan cara *drag and drop* pada data yang sudah di import. Setelah itu ditambahkan operator "*Weight by Information Gain Ratio*", operator ini digunakan untuk memberikan bobot pada setiap atribut.

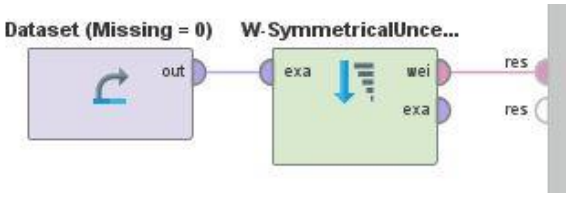
Tabel 5. Hasil seleksi fitur *Gain Ratio*

| No | Atribut | *Gain Ratio* |
|---|---|---|
| 1 | *Sgpt Alamine Aminotransferase* | 0.227 |
| 2 | *Age* | 0.201 |
| 3 | *Sgot Asparte Aminotransferase* | 0.170 |
| 4 | DB | 0.115 |
| 5 | TB | 0.113 |
| 6 | *Alkhos Alkanine Phosphotase* | 0.066 |
| 7 | TP | 0.061 |
| 8 | A/G | 0.054 |
| 9 | ALB | 0.054 |
| 10 | *Gender* | 0.006 |

Atribut terpilih pada seleksi fitur *gain ratio* pada dataset *Indian Liver Patient Dataset* adalah *Sgpt Alamine Aminotransferase*, *Age* dan *Sgot Asparte Aminotransferase* yang dapat dilihat pada Tabel 6

Tabel 6. Atribut terpilih pada *Gain Ratio*

| No | Atribut | *Gain Ratio* |
|---|---|---|
| 1 | *Sgpt Alamine Aminotransferase* | 0.227 |
| 2 | *Age* | 0.201 |
| 3 | *Sgot Asparte Aminotransferase* | 0.170 |




Gambar 4. Skema seleksi fitur symmetrical uncertainty

### 4.4. Seleksi Fitur *Symmetrical Uncertainty*

Pada proses perhitungan bobot atribut dengan seleksi fitur *symmetrical uncertainty* dilakukan dengan menggunakan RapidMiner Studio dengan skema pengujian seperti Gambar 3 diatas didapat hasil seperti tabel 7 berikut

Tabel 7. Hasil *seleksi* fitur *Symmetrical Uncertainty*

| No | Atribut | *Symmetrical Uncertainty* |
|---|---|---|
| 1 | TB | 0.069 |
| 2 | DB | 0.066 |
| 3 | *Sgpt Alamine Aminotransferase* | 0.045 |
| 4 | *Alkhos Alkanine Phosphotase* | 0.044 |
| 5 | *Sgot Asparte Aminotransferase* | 0.035 |
| 6 | A/G | 0.029 |
| 7 | *Age* | 0.024 |
| 8 | ALB | 0.016 |
| 9 | *Gender* | 0.006 |
| 10 | TP | 0.004 |

Dari hasil tersebut dapat dilihat bahwa atribut TB (*Total Bilirubin*) memiliki bobot paling besar dengan nilai 0.069 yang berarti bahwa atribut tersebut memiliki pengaruh yang paling besar terhadap dataset, dan begitu seterusnya hingga atribut TP (*Total Protiens*) yang memiliki pengaruh paling kecil pada data set dengan nilai atribut 0.004.

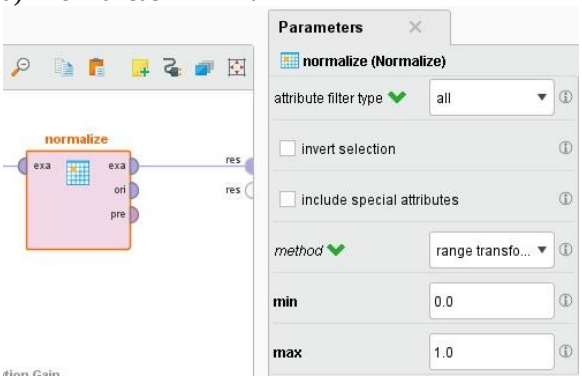
Tabel 8. Hasil seleksi fitur *symmetrical uncertainty*

| No | Atribut | *Information Gain* |
|---|---|---|
| 1 | DB | 0.069 |
| 2 | TB | 0.066 |
| 3 | *Sgpt Alamine Aminotransferase* | 0.045 |

### 4.5. Normalisasi

Normalisasi dilakukan pada data yang ada untuk memetakan data dalam rentang tertentu. Data yang dinormalisasi adalah data berdasarkan atribut yang dipilih pada tahap pemilihan data menggunakan seleksi fitur.

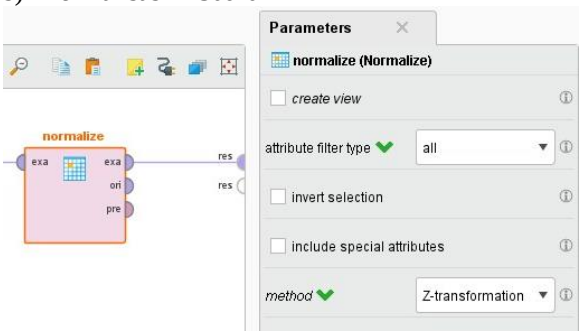
a) Normalisasi *Min-Max*




Gambar 5. Proses min-max normalization

Proses normalisasi data dilakukan dengan bantuan *RapidMiner Studio* dengan menggunakan operator "*Retrieve*" yang dilakukan dengan cara *drag and drop* data yang sudah di *import* sebelumya. Setelah data sudah terbaca, tambahkan operator "*Select Attributes*" dan "*Normalize*" dengan parameter yang sudah terpapar pada gambar 5.

b) Normalisasi *Z-Score*



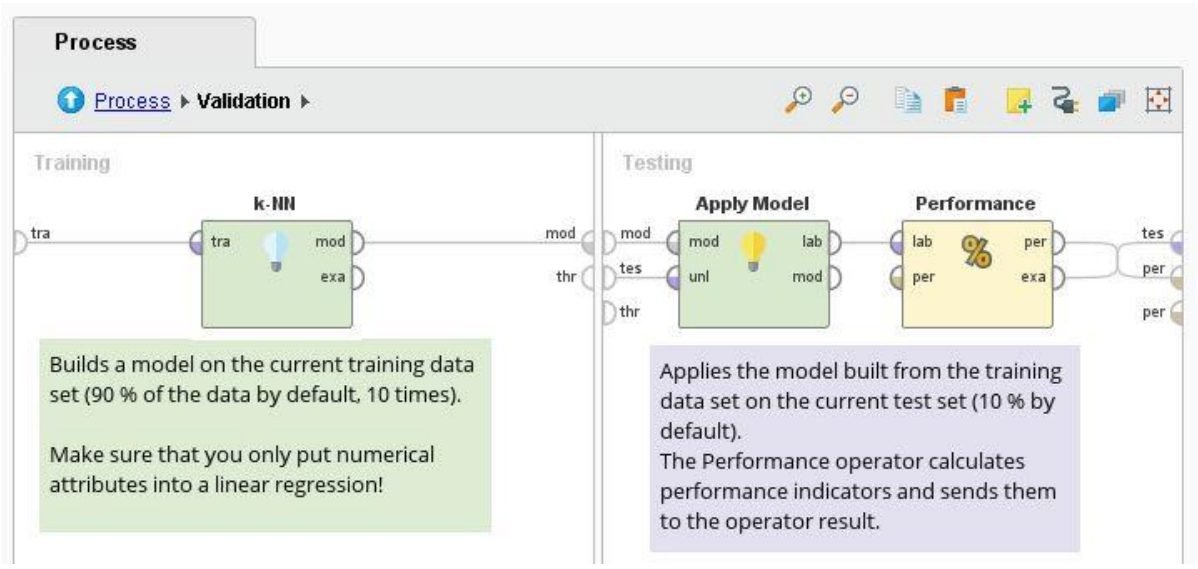

Gambar 6. Proses z-score normalization

Tahap normalisasi dilakukan pada data yang dipilih pada proses seleksi fitur. Proses normalisasi data menggunakan bantuan *RapidMiner Studio*. Pada proses *Z-Score Normalization* menggunakan *RapidMiner* ditunjukkan pada Gambar 6.

### 4.6. Klasifikasi *K-Nearest Neighbor* (KNN)

Pada proses klasifikasi dari *Indian Liver Patient Dataset* menggunakan algoritma *K-Nearest Neighbor* dengan menggunakan 10-*Fold Cross Validation*. Pada operator "*Cross Validation*" terdapat parameter *number of folds'* yang digunakan untuk mengatur berapa jumlah *folds* yang dibutuhkan, pada penelitian ini akan menggunakan 10 nilai *folds.* 10-*Fold Cross Validation* artinya data akan dibagi menjadi 90% data menjadi data latih dan 10%-nya menjadi data uji. Dengan *sampling type* menggunakan *shuffled sampling.*

Didalam operator "*Cross Validation*" menggunakan operator "*k-NN*" dengan parameter nilai *k* disesuaikan dengan kebutuhan dan pada parameter pengukuran jarak dipilih e*uclidean distance.* Sehingga pada operator *measure types* dipilih *NumericalMeasure* dan pada parameter *numerical measure* pilih pada pilihan *EuclideanDistance.* Operator KNN pada *RapidMiner* dan parameternya ditampilkan pada Gambar 7.

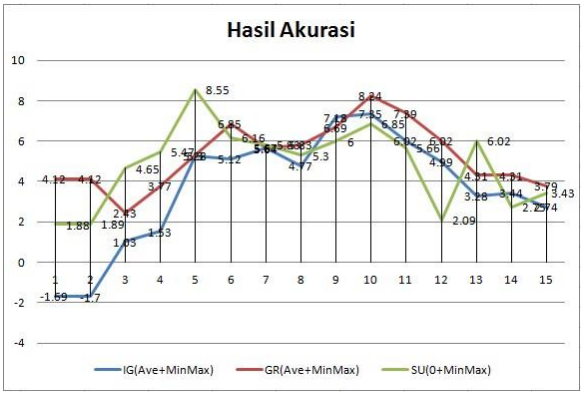



Gambar 7. Operator *KNN pada* RapidMiner Studio

Pada penelitian ini melakukan penangana pada data yang mengalami kekosongan atau disebut *missing value*, normalisasi *min-max* dan *Z-Score*, serta dilakukan seleksi fitur dengan *information gain, gain ratio*, dan *symmetrical uncertainty* untuk mengoptimalkan kinerja algoritma *K-Nearest Neighbor* dengan mengujikan nilai *k* dari 1 sampai 10 sehingga bisa menghasilkan nilai akurasi yang lebih baik dengan tanpa dilakukan seleksi fitur. Data set yang digunakan dalam penelitian ini adalah *Indian Liver Patient Datasets* (ILPD) yang diperoleh dari *UCI Machine Learning Repository*. Data ini adalah data sekunder yang merupakan data yang digunakan dalam penelitiannya didapatkan secara tidak langsung melainkan melalui perantara. Pada data ILPD terdapat 10 atribut dan 1 atribut *class*.

Penelitian ini dilakukan untuk mencari model klasifikasi yang paling optimal untuk mengklasifikasi penyakit *liver*, dengan cara membandingkan hasil akurasi algoritma KNN setelah dilakukan penanganan *missing value*, normalisasi dan seleksi fitur dengan algoritma KNN yang tanpa menggunakan seleksi fitur untuk mengetahui pengaruh seleksi fitur pada tingkat akurasi algoritma KNN.

*Missing* pada data dikarenakan adanya data yang kosong pada atributnya, *missing value* pada data ini dilakukan dengan mengganti nilai yang *missing* dengan 0 dan nilai rata-rata dari atribut yang memiliki kekosongan data. Selanjutnya dilakukan normalisasi pada data, metode normalisasi menggunakan *Min-Max* dan *Z-Score* yang dapat memberikan peningkatan akurasi dengan merepresentasikan data asli dalam bentuk data baru dengan rentang nilai yang hamper seragam. Karena hal tersebut dapat menyederhanakan data sehingga proses yang dilakukan lebih cepat dan memberikan hasil optimasi yang lebih meningkat pada algoritma KNN. Selain itu dilakukan juga proses seleksi fitur, seleksi fitur yang digunakan pada penelitian ini menggunakan pembobotan *Information Gain* (IG), *Gain Ratio* (GR), dan *Symmetrical Uncertainty* (SU). Proses klasifikasi ini dilakukan dengan menggunakan nilai *k* antara 1 sampai 10.




Gambar 8. Hasil Akurasi

Penelitian ini mendapatkan hasil akurasi sesuai model yang paling optimal dapat dilihat seperti gambar 8, pada *information gain* dan gain ratio didapatkan nilai paling optimal pada *missing value* yang dilakukan penanganan dengan nilai rata-rata pada data yang memiliki nilai kekosongan dan dilakukan normalisasi *min-max* juga nilai *k* = 10. Sedangkan pada *symmetrical uncertainty* didapatkan nilai optimal pada nilai *k* = 5 dan menggunakan penanganan *missing value* dengan 0 dan normalisasi *min- max* yang dapat dilihat pada gambar berikut

## 5. KESIMPUL8AN DAN SARAN

Dari hasil penelitian terkalit implementasi normalisasi dan seleksi fitur untuk meningkatkan akurasi ada algoritma *K-Nearest Neighbor* (KNN) pada *Indian Liver Patient* Dataset dapat ditarik kesimpulan sebagai berikut: Pada penelitian ini penerapan seleksi fitur dengan menggunakan pembobotan *Information Gain, Gain Ratio*, dan *Symmetrical Uncertainty* pada *Indian Liver Patient Dataset* menggunakan lagoritma KNN tidak selalu meningkatkan hasil akurasi setelah proses klasifikasinya. Implementasi normalisasi data pada seleksi fitur juga dapat mempengaruhi hasil akurasi, normalisasi data yang tepat saat menggunakan seleksi fitur pada data ini adalah menggunakan *Min-Max Normalization.* Model klasifikasi yang didapatkan pada penelitian ini berupa model optimasi pada seleksi fitur *Information Gain* dan *Gain Ratio* dapat meningkatkan hasil akurasi apabila dilakukan pengisian data yang kosong dengan nilai rata-rata pada data yang memiliki kekosongan data dan normalisasi *Min-Max* serta menggunakan nilai k = 10, apa bila dengan seleksi fitur *Symmetrical Uncertainty* dapat menggunakan penanganan data kosong dengan 0 dan normalisasi data *Min-Max* serta menggunakan nilai k = 5. Adapun saran untuk penelitian yang ingin mengembangkan labih lanjut adalah pada penelitian selanjutnya diharapkan dapat mengembangkan penelitian ini dengan menggunakan algoritma lain yang dapat meningkatkan hasil akurasi pada klasifikasi penyakit *liver* serta pada penelitian selanjutnya diharapkan mampu meningkatkan kinerja algortima KNN pada data lain yang lebih besar.

## DAFTAR PUSTAKA

[1] D. &. K. M. A. Sugianti, "Pengingkatan Akurasi Algoritma KNN dengan Seleksi Fitur Gain Ratio Untuk Klasifikasi Penyakit Diabetes Melitus," *IC-Tech,* vol. 12, no. 2, 6, 2017.

[2] P. I. Ashari, "Performance Comparising Between Naive Bayes, Decission Tree and K-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 4, no. 11, pp. 33-39, 2018.

[3] M. E. S. &. S. R. W. Daniel, "The Analysis of Attribution Reduction of K-Nearest Neighbor (KNN) Algorithm by Using Chi-Square," *In Journal of Physics: Conference Series,* vol. 1424, no. 1, 2019.

[4] A. Duneja and T. Puyalnithi, "Enhancing Classification Accuracy of K-Nearest Neighbours Algorithm Using Gain Ratio," *International Research Journal of Engineering and Technology (IRJET),* vol. 4, no. 9, pp. 1385-1388, 2017.

[5] S. &. S. D. Susanto, Pengantar Data Mining: Menggali Pengetahuan Dari Bongkahan Data., Yogyakarta: Andi Offset, 2018.

[6] d. I. Binabar S. W., "Optimasi Parameter K Pada Algoritma KNN Untuk Deteksi Penyakit Kanker Payudara," *IC-Tech,* vol. 2, pp. 11-18, 2017.

[7] L. H. R, "An Intellegent Model for Liver Disease Diagnosis Artificial Intellegence in Medicine," 2019.

[8] A. K. B. Ginting, M. S. Lydia and M. Z. Elviwaty, "Peningkatan Akurasi Metode K-Nearest Neighbor dengan Seleksi Fitur," *JURNAL MEDIA INFORMATIKA BUDIDARMA,* vol. 5, no. 4, pp. 1714-1719, 2021.

[9] F. Yunita, "Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru," *Sistemasi,* vol. 7, no. 3, p. 238, 2018.

[10] T. Hendrawati, "KAJIAN METODE IMPUTASI DALAM MENANGANI MISSING DATA.," in *Pros. Semin. Nas. Mat. dan Pendidik. Mat*, 2018.

[11] S. D. Indrayanti, "Peningkatan Akurasi Algoritma KNN Dengan Seleksi Fitur Gain Ratio Untuk Klasifikasi Penyakit Diabetes Mellitus.," *IC-Tech,* vol. 12, no. 2, 2017.