

Penerapan Algoritma *K-Nearest Neighbord* Untuk Prediksi Kematian Akibat Penyakit Gagal Jantung

Dede Andri Muhammad Reza
Universitas Buana Perjuangan
Karawang, Indonesia
if17.dedereza@mhs.ubpkarawang.ac.id

Amril Mutoi Siregar
Universitas Buana Perjuangan
Karawang, Indonesia
amrilmutoi@ubpkarawang.ac.id

Rahmat
Universitas Buana Perjuangan
Karawang, Indonesia
rahmat@ubpkarawang.ac.id

Abstrak—

Penyakit kardiovaskular (PKV) menurut definisi dari *World Health Organization (WHO)* dalam pusdatin Kemenkes RI 2014 yaitu penyakit yang disebabkan oleh gangguan fungsi jantung dan pembuluh darah, seperti penyakit jantung koroner, hipertensi, stroke dan penyakit gagal jantung. Data *WHO* mengatakan, lebih dari 17 juta manusia di dunia meninggal disebabkan penyakit jantung dan pembuluh darah. Gagal jantung merupakan suatu keadaan dimana jantung tidak dapat memompa darah yang mencukupi untuk kebutuhan tubuh. Mengingat berharganya organ vital seperti jantung, memprediksi gagal jantung telah menjadi prioritas bagi dokter dan tenaga medis, tetapi hingga saat ini prediksi kejadian terkait gagal jantung dalam praktik klinis biasanya gagal mencapai akurasi yang tinggi. Dengan diterapkannya *data mining* diharapkan dapat menjadi suatu informasi untuk meminimalisir penyakit gagal jantung di Indonesia ataupun dunia. Penelitian ini menggunakan algoritma *K-Nearest Neighbord* yaitu merupakan algoritma klasifikasi berdasarkan kedekatan jarak suatu data dengan data yang lain. Hasil yang didapat pada penelitian ini yaitu menghitung manual menggunakan *Microsoft excell* dengan menggunakan nilai $k=7$ yang menghasilkan kategori peristiwa kematian data testing adalah **Tidak**. Pengujian pada rapidminer dilakukan dengan menggunakan pergantian nilai k , akurasi tertinggi didapat pada nilai $k=7$ dengan nilai akurasi 94,92%. Kemudian pengujian pada bahasa pemrograman *python* menghasilkan nilai akurasi 68%.

Kata kunci— gagal jantung, *data mining*, prediksi, algoritma *K-Nearest Neighbord*

I. PENDAHULUAN

Penyakit kardiovaskular (PKV) menurut definisi dari World Health Organization (WHO) dalam pusdatin Kemenkes RI 2014 yaitu penyakit yang disebabkan oleh gangguan fungsi jantung dan pembuluh darah, seperti penyakit jantung koroner, hipertensi, stroke dan penyakit gagal jantung. Setiap tahunnya lebih dari 36 juta orang meninggal karena Penyakit Tidak Menular (PTM) 63% dari semua kematian. Secara global PTM menjadi pemicu kematian nomor satu setiap tahunnya. Kematian dini yang disebabkan oleh penyakit jantung terjadi sebesar 4% di negara berpendapatan tinggi dan 42% terjadi di negara berpendapatan rendah [1]. Data WHO mengatakan, lebih dari 17 juta manusia di dunia meninggal disebabkan penyakit jantung dan pembuluh darah [2].

Penyakit gagal jantung masuk ke kategori penyakit kardiovaskular. Gagal jantung adalah suatu kondisi ketika jantung tidak bisa mendapatkan cukup darah untuk memenuhi kebutuhan tubuh. Penderita gagal jantung merupakan pasien paling banyak di rumah sakit. Prevalensi gagal jantung di Amerika dan Eropa sekitar 1-2%. Di Indonesia belum ada data epidemiologi untuk gagal jantung, tetapi pada survei kesehatan nasional 2003 ditemukan bahwa penyakit sistem sirkulasi merupakan penyebab kematian utama di Indonesia (26,4%) dan pada profil Kesehatan Indonesia tahun 2003 disebutkan bahwa penyakit jantung berada pada urutan kedelapan [3]. Mengingat berharganya organ vital seperti jantung, memprediksi gagal jantung telah menjadi prioritas bagi dokter dan tenaga medis, namun sampai saat ini prediksi perihail gagal jantung dalam praktik klinis biasanya gagal mendapatkan akurasi yang tinggi [4].

Pendekatan *data mining* menjadi sangat krusial dalam aspek Kesehatan untuk mengambil keputusan berlandaskan data klinis yang besa [5]. *Data mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan memakai teknik atau metode tertentu [6]. Diterapkannya *data mining* diharapkan dapat menjadi suatu informasi untuk mencegah atau mengurangi penderita penyakit gagal jantung di Indonesia dan dunia, sehingga angka penderita penyakit gagal jantung dapat menurun.

Algoritma *K-Nearest Neighbor (KNN)* yakni algoritma klasifikasi berlandaskan kedekatan jarak suatu data dengan data yang lain. Pada algoritma *K-Nearest Neighbor (KNN)*, data berdimensi q , jarak dari data tersebut ke data yang lain dapat dihitung. Nilai jarak inilah yang ditetapkan sebagai nilai kedekatan atau kemiripan antara data uji dengan data latih. Nilai K pada *K-Nearest Neighbor (KNN)* berarti K -data terdekat dari data uji [7]

Telah banyak penelitian tentang prediksi dengan memakai algoritma *K-Nearest Neighbor (KNN)* salah satunya penelitian yang di lakukan oleh Gultom dan Simanjuntak (2020) [8], yaitu prediksi tingkat kelancaran pembayaran kredit bank. Algoritma yang digunakan yaitu naïve bayes dan *K-Nearest Neighbor (KNN)*, dataset yang digunakan mengandung 363 data dan 14 atribut. Hasil yang di peroleh menyatakan bahwa pengujian yang dilakukan dengan menggabungkan algoritma *Naïve Bayes* dan *K-Nearest Neighbor (KNN)* mampu melakukan prediksi dengan baik dengan akurasi 80%. Sehingga gabungan kedua algoritma tersebut cocok digunakan untuk prediksi kelancaran pembayaran kredit pada bank, dan dapat membantu petugas bank dalam menentukan nasabahnya.

Selain itu ada juga penelitian Lestari (2014) [7], penelitian ini meneliti tentang mendeteksi penyakit jantung. Algoritma yang digunakan dalam penelitian ini yaitu *K-Nearest Neighbor (KNN)*, dataset yang digunakan mengandung 100 data pasien penyakit jantung dan 13 atribut. Hasil dari penelitian ini yaitu, algoritma *K-Nearest Neighbor (KNN)* dengan memakai confusion matrix dan kurva ROC, diperoleh nilai akurasi 70% dan termasuk klasifikasi baik sebab memiliki nilai AUC 0.875.

Teknik-teknik penelitian yang telah dilakukan Gultom dengan Simanjuntak (2020) dan Lestari (2014) Beserta penjelasan yang sudah di paparkan, maka pada penelitian ini akan mengambil judul “Penerapan Algoritma K-Nearest Neighbour (KNN) Untuk Prediksi Kematian Akibat Penyakit Gagal Jantung”.

II. DATA DAN METODE

A. Bahan dan Peralatan Penelitian

Penelitian ini menggunakan dataset penderita penyakit gagal jantung. Data berisi tiga belas atribut dengan 299 orang penderita penyakit gagal jantung yang di peroleh dari <https://www.kaggle.com/andrewmvd/heart-failure-clinical-data>.

Tabel 1 Data Penelitian

| No | U | A | KF | D | FE | TDT | T | KS | NS | JK | M | W | PK |
|-----|-----|-----|------|-----|-----|-----|--------|-----|-----|-----|-----|-----|-----|
| 1 | 75 | 0 | 582 | 0 | 20 | 1 | 265000 | 1,9 | 130 | 1 | 0 | 4 | 1 |
| 2 | 55 | 0 | 7861 | 0 | 38 | 0 | 263358 | 1,1 | 136 | 1 | 0 | 6 | 1 |
| 3 | 65 | 0 | 146 | 0 | 20 | 0 | 162000 | 1,3 | 129 | 1 | 1 | 7 | 1 |
| 4 | 50 | 1 | 111 | 0 | 20 | 0 | 210000 | 1,9 | 137 | 1 | 0 | 7 | 1 |
| 5 | 65 | 1 | 160 | 1 | 20 | 0 | 327000 | 2,7 | 116 | 0 | 0 | 8 | 1 |
| 6 | 90 | 1 | 47 | 0 | 40 | 1 | 204000 | 2,1 | 132 | 1 | 1 | 8 | 1 |
| 7 | 75 | 1 | 246 | 0 | 15 | 0 | 127000 | 1,2 | 137 | 1 | 0 | 10 | 1 |
| 8 | 60 | 1 | 315 | 1 | 60 | 0 | 454000 | 1,1 | 131 | 1 | 1 | 10 | 1 |
| 9 | 65 | 0 | 157 | 0 | 65 | 0 | 263358 | 1,5 | 138 | 0 | 0 | 10 | 1 |
| 10 | 80 | 1 | 123 | 0 | 35 | 1 | 388000 | 9,4 | 133 | 1 | 1 | 10 | 1 |
| 11 | 75 | 1 | 81 | 0 | 38 | 1 | 368000 | 4 | 131 | 1 | 1 | 10 | 1 |
| 12 | 62 | 0 | 231 | 0 | 25 | 1 | 253000 | 0,9 | 140 | 1 | 1 | 10 | 1 |
| 13 | 45 | 1 | 981 | 0 | 30 | 0 | 136000 | 1,1 | 137 | 1 | 0 | 11 | 1 |
| 14 | 50 | 1 | 168 | 0 | 38 | 1 | 276000 | 1,1 | 137 | 1 | 0 | 11 | 1 |
| 15 | 49 | 1 | 80 | 0 | 30 | 1 | 427000 | 1 | 138 | 0 | 0 | 12 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 299 | 50 | 0 | 196 | 0 | 45 | 0 | 395000 | 1,6 | 136 | 1 | 1 | 285 | 0 |

Kemudian, untuk memenuhi kebutuhan penelitian membutuhkan alat seperti perangkat keras dan perangkat lunak sebagai berikut :

I. Perangkat Keras

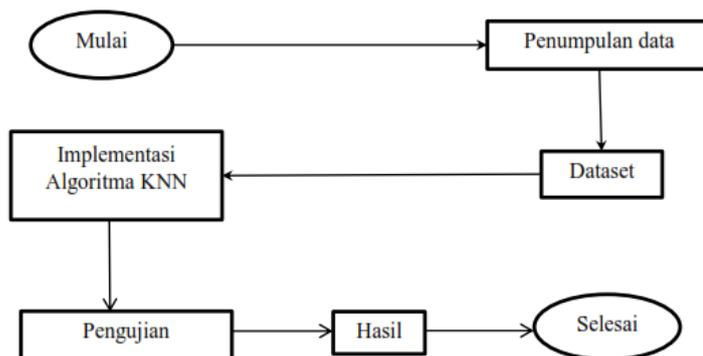
- Leptop
- Storage 500GB
- Memori 4GB DDR4

II. Perangkat Lunak

- Sistem operasi *windows 10 pro*
- *Microsoft office word*
- *Microsoft office excel*
- Aplikasi *rapidminer*
- *Google colaboratory*

B. Prosedur Penelitian

Prosedur penelitian ini sebagai berikut :



Gambar 1 Prosedur Penelitian

C. Pengumpulan Data

Pada tahap ini berisi penjelasan data penelitian yang digunakan, tahapan awal yang dilakukan dalam penelitian ini yaitu pengumpulan data yang diperoleh dari website kaggle. Data yang diperoleh merupakan dataset pasien penderita penyakit gagal jantung yang dikumpulkan selama masa tindak lanjut, dataset berjumlah 299 data dan 13 atribut.

D. Implementasi Algoritma *K-Nearest Neighbord (KNN)*

Pada tahapan ini yang dilakukan yaitu, mengimplementasikan algoritma *K-Nearest Neighbord (KNN)*. Algoritma *K-Nearest Neighbord* dihitung dengan menggunakan rumus *euclidean distance* sebagai berikut [9].

$$d_{Euclidean}(x, y) = \sqrt{\sum_i^p (x_1 - x_2)^2}$$

Keterangan:

X1 = Sampel data

X2 = Data uji atau data testing

i = variabel data

d = jarak

p = Dimensi data

Nilai yang dihasilkan dalam implementasi algoritma KNN berupa tingkat akurasi dari algoritma KNN dalam memprediksi kematian akibat penyakit gagal jantung. Implementasi penelitian ini dilakukan beberapa proses yaitu:

- 1) Perhitungan secara manual.
- 2) Pengujian menggunakan *tool RapidMiner*.
- 3) Implementasi algoritma KNN dengan menggunakan bahasa pemrograman *python*.

E. Pengujian

Pengujian dilakukan untuk menentukan tingkat akurasi algoritma *KNN*. Evaluasi dan validasi hasil hitung menggunakan *confision matrix* dengan rumus akurasi, *precision*, *recall* [10].

Tabel 2 *Confusion matrix*

| <i>Correct Classification</i> | <i>Classification</i> | |
|-------------------------------|-----------------------|---------|
| | Positif | Negatif |
| Positif | TP | FN |
| Negatif | FP | TN |

Evaluasi dan validasi dihitung memakai rumus akurasi, *precision*, *recall* dan *f-measure* berikut ini:

a) Akurasi

Perhitungan akurasi dilakukan dengan membagi jumlah data yang diklasifikasi secara benar dengan total sampel data yang diuji. Berikut dibawah ini merupakan rumus akurasi.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN}$$

b) *Precision*

Menghitung nilai dengan membagi jumlah data benar yang bernilai *True Positive* (TP) dibagi dengan jumlah data benar *True Positive* (TP) dan ditambah data yang bernilai *False Positive* (FP).

$$Precision = \frac{TP}{TP + FP}$$

c) *Recall*

Recall dihitung dengan membagi data yang benar yang bernilai *True Positive* (TP) dengan hasil penjumlahan dari data benar yang bernilai *True Positive* (TP) dan data yang salah bernilai *False Positive* (FP)

$$Recall = \frac{TP}{TP + FN}$$

d) *F-Measure*

Menghitung nilai *F-Measure* didapat dari perhitungan pembagian hasil perkalian *precision* dan *recall* dengan hasil penjumlahannya dikalikan dua.

$$F\text{-Measure} = 2 \times \frac{precision \times recall}{precision + recall}$$

III. HASIL DAN PEMBAHASAN

A. Seleksi Data

Bagian ini merupakan tahapan selanjutnya dari pengumpulan data. Pada penelitian ini terdapat 299 data pasien penderita penyakit jantung, tetapi setelah di seleksi ditemukan beberapa data di dalam atribut (Usia) yang nilainya tidak masuk akal dan sangat jauh berbeda dengan nilai data lainnya. Sedangkan untuk atributnya, dari 13 atribut dalam data gagal jantung ada atribut yang dihapus yaitu (Trambosit) karena menurut, dr. Amadeo Drian Basfiansa sejauh ini belum ditemukan kaitan yang spesifik antara kondisi gagal jantung kongestif atau CHF, dengan trombotosis. Setelah proses seleksi data maka, data yang digunakan yaitu berjumlah 297 data dan 12 atribut. Berikut ini contoh dataset yang sudah di seleksi.

Tabel 3 setelah tahap seleksi

| No | U | A | KF | D | FE | TDT | KS | NS | JK | M | W | PK |
|------|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1. | 75 | 0 | 582 | 0 | 20 | 1 | 1,9 | 130 | 1 | 0 | 4 | 1 |
| 2. | 55 | 0 | 7861 | 0 | 38 | 0 | 1,1 | 136 | 1 | 0 | 6 | 1 |
| 3. | 65 | 0 | 146 | 0 | 20 | 0 | 1,3 | 129 | 1 | 1 | 7 | 1 |
| 4. | 50 | 1 | 111 | 0 | 20 | 0 | 1,9 | 137 | 1 | 0 | 7 | 1 |
| 5. | 65 | 1 | 160 | 1 | 20 | 0 | 2,7 | 116 | 0 | 0 | 8 | 1 |
| 6. | 90 | 1 | 47 | 0 | 40 | 1 | 2,1 | 132 | 1 | 1 | 8 | 1 |
| 7. | 75 | 1 | 246 | 0 | 15 | 0 | 1,2 | 137 | 1 | 0 | 10 | 1 |
| 8. | 60 | 1 | 315 | 1 | 60 | 0 | 1,1 | 131 | 1 | 1 | 10 | 1 |
| 9. | 65 | 0 | 157 | 0 | 65 | 0 | 1,5 | 138 | 0 | 0 | 10 | 1 |
| 10. | 80 | 1 | 123 | 0 | 35 | 1 | 9,4 | 133 | 1 | 1 | 10 | 1 |
| 11. | 75 | 1 | 81 | 0 | 38 | 1 | 4 | 131 | 1 | 1 | 10 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 297. | 50 | 0 | 196 | 0 | 45 | 0 | 1,6 | 136 | 1 | 1 | 285 | 0 |

B. Hasil Perhitungan Manual

Pada tahap ini dilakukan perhitungan manual dengan menggunakan algoritma *KNN*. Sebelum melakukan perhitungan manual *KNN*, perlu disiapakan data kelas atau label, data atribut, data lama (*training*) dan data baru (*testing*).

- a) Kategori peristiwa kematian akibat penyakit gagal jantung.

Tabel 4 kategori peristiwa kematian

| Atribut | Nilai | Kategori |
|--------------------|-------|-----------|
| Peristiwa Kematian | 1 | Meninggal |
| | 0 | Tidak |

- b) Pengujian dilakukan dengan membutuhkan data uji (*testing*), data testing mempunyai atribut yang serupa dengan data (*training*), namun kelas atau kategorinya belum diketahui

Tabel 5 data testing

| U | A | KF | D | FE | TDT | KS | NS | JK | M | W | PK |
|----|---|-----|---|----|-----|-----|-----|----|---|-----|----|
| 59 | 1 | 350 | 0 | 45 | 1 | 1,5 | 125 | 0 | 1 | 105 | ? |

C. Tahapan perhitungan *K-Nearest Neighbord (KNN)*

Pada perhitungan *KNN* ada sejumlah tahapan yang harus dikerjakan, berikut adalah tahapan-tahapannya.

- 1) Menentukan parameter *K*, yaitu jumlah banyaknya tetangga terdekat yaitu *K-7*
- 2) Menghitung jarak antara data baru dan semua data *training*
- 3) Mengurutkan jarak tersebut dan menentukan tetangga mana yang terdekat berdasarkan jarak minimum ke-*k*
- 4) Menentukan kategori dari tetangga terdekat
- 5) Menggunakan kategori mayoritas sederhana dari tetangga yang terdekat tersebut sebagai nilai prediksi dari data yang baru.

D. Perhitungan jarak

Perhitungan jarak dikerjakan dengan rumus algoritma *K-Nearest Neighbord* yang sudah ada, berikut dibawah ini merupakan contoh perhitungan jarak data pasien penderita penyakit gagal jantung ke 1 dan ke 2 terhadap data testing.

$$d(x, y) = \sqrt{\sum_t^p (x_1 - x_2)^2}$$

$$\begin{aligned}
 d1 &= \sqrt{(75 - 59)^2 + (0 - 1)^2 + (582 - 350)^2 + (0 - 0)^2 + (20 - 45)^2} \\
 &\quad \sqrt{+(1 - 1)^2 + (1,9 - 1,5)^2 + (130 - 125)^2 + (1 - 0)^2 + (0 - 1)^2 +} \\
 &\quad \sqrt{(4 - 105)^2} \\
 &= \sqrt{(256) + (1) + 53824 + (0) + (625) + (0) + (0,16) + (25) +} \\
 &\quad \sqrt{(1) + (1) + (10201)} \\
 &= \sqrt{64934,16} \\
 &= 254,8218201
 \end{aligned}$$

$$\begin{aligned}
 d2 &= \sqrt{(55 - 59)^2 + (0 - 1)^2 + (7861 - 350)^2 + (0 - 0)^2 +} \\
 &\quad \sqrt{(38 - 45)^2 + (0 - 1)^2 + (1,1 - 1,5)^2 + (136 - 125)^2 + (1 - 0)^2} \\
 &\quad \sqrt{+(0 - 1)^2 + (6 - 105)^2} \\
 &= \sqrt{(16) + (1) + (56415121) + (0) + (49) + (1) + (0,16) +} \\
 &\quad \sqrt{(121) + (1) + (1) + (9801)}
 \end{aligned}$$

$$= \sqrt{56.425.112,16}$$

$$= 7511.665072$$

Tabel 6 perhitungan jarak

| No | U | A | KF | D | FE | T D T | KS | NS | JK | M | W | PK | Rank | Eucliden Distance |
|------|-----|-----|------|-----|-----|-------------|------|-----|-----|-----|------|-----------|------|----------------------|
| 1. | 75 | 0 | 582 | 0 | 20 | 1 | 1,9 | 130 | 1 | 0 | 4 | Meninggal | 158 | 254,8218201 |
| 2. | 55 | 0 | 7861 | 0 | 38 | 0 | 1,1 | 136 | 1 | 0 | 6 | Meninggal | 297 | 7511,665072 |
| 3. | 65 | 0 | 146 | 0 | 20 | 0 | 1,3 | 129 | 1 | 1 | 7 | Meninggal | 89 | 227,8158028 |
| 4. | 50 | 1 | 111 | 0 | 20 | 0 | 1,9 | 137 | 1 | 0 | 7 | Meninggal | 166 | 259,9579966 |
| 5. | 65 | 1 | 160 | 1 | 20 | 0 | 2,7 | 116 | 0 | 0 | 8 | Meninggal | 77 | 215,070779 |
| 6. | 90 | 1 | 47 | 0 | 40 | 1 | 2,1 | 132 | 1 | 1 | 8 | Meninggal | 233 | 319,7723565 |
| 7. | 75 | 1 | 246 | 0 | 15 | 0 | 1,2 | 137 | 1 | 0 | 10 | Meninggal | 36 | 145,4100753 |
| 8. | 60 | 1 | 315 | 1 | 60 | 0 | 1,1 | 131 | 1 | 1 | 10 | Meninggal | 19 | 102,5434542 |
| 9. | 65 | 0 | 157 | 0 | 65 | 0 | 1,5 | 138 | 0 | 0 | 10 | Meninggal | 80 | 216,5225162 |
| 10. | 80 | 1 | 123 | 0 | 35 | 1 | 9,4 | 133 | 1 | 1 | 10 | Meninggal | 139 | 247,4316269 |
| 11. | 75 | 1 | 81 | 0 | 38 | 1 | 4 | 131 | 1 | 1 | 10 | Meninggal | 198 | 285,8920251 |
| 12. | 62 | 0 | 231 | 0 | 25 | 1 | 0,9 | 140 | 1 | 1 | 10 | Meninggal | 42 | 154,3449384 |
| 13. | 45 | 1 | 981 | 0 | 30 | 0 | 1,1 | 137 | 1 | 0 | 11 | Meninggal | 261 | 638,408302 |
| 14. | 50 | 1 | 168 | 0 | 38 | 1 | 1,1 | 137 | 1 | 0 | 11 | Meninggal | 71 | 205,5143791 |
| 15. | 49 | 1 | 80 | 0 | 30 | 1 | 1 | 138 | 0 | 0 | 12 | Tidak | 200 | 286,4336747 |
| 16. | 82 | 1 | 379 | 0 | 50 | 0 | 1,3 | 136 | 1 | 0 | 13 | Meninggal | 18 | 99,91516401 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | | ... | |
| ... | ... | ... | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | | | ... | |
| 196. | 45 | 0 | 2413 | 0 | 38 | 0 | 1,4 | 140 | 1 | 1 | 280 | Tidak | 285 | 2070,523366 |
| 297. | 50 | 0 | 196 | 0 | 45 | 0 | 1,6 | 136 | 1 | 1 | 285 | Tidak | 114 | 237,3204795 |

Selanjutnya setelah menghitung jarak langkah selanjutnya yaitu, mengurutkan jarak tersebut dan menentukan tetangga mana yang terdekat berdasarkan jarak minimum K=7

Tabel 7 tetangga terdekat

| No | U | A | KF | D | FE | TDT | KS | NS | JK | M | PK | PK | Distence |
|----|----|---|-----|---|----|-----|-----|-----|----|---|-----|-----------|-------------|
| 1. | 43 | 1 | 358 | 0 | 50 | 0 | 1,3 | 135 | 0 | 0 | 97 | Tidak | 22,60619384 |
| 2. | 65 | 1 | 335 | 0 | 35 | 1 | 0,8 | 136 | 0 | 0 | 120 | Tidak | 26,61747546 |
| 3. | 50 | 0 | 369 | 1 | 25 | 0 | 1,6 | 136 | 1 | 0 | 90 | Tidak | 34,53997684 |
| 4. | 72 | 1 | 328 | 0 | 30 | 1 | 1,7 | 138 | 0 | 1 | 88 | Meninggal | 36,55188094 |
| 5. | 55 | 0 | 336 | 0 | 45 | 1 | 0,9 | 140 | 0 | 0 | 74 | Tidak | 37,42138426 |
| 6. | 58 | 1 | 400 | 0 | 40 | 0 | 1 | 139 | 0 | 0 | 91 | Tidak | 54,03933752 |
| 7. | 65 | 1 | 305 | 0 | 25 | 0 | 1,1 | 141 | 1 | 0 | 87 | Tidak | 55,17390688 |

Berdasarkan hasil akhir dari perhitungan k-7 maka kategori kelas mayoritas adalah **Tidak** Karena 1 diantaranya **Meninggal** dan 6 **Tidak**, selanjutnya menentukan klasifikasi baru.

Tabel 8 hasil prediksi

| U | A | KF | D | FE | TDT | KS | NS | JK | M | W | PK |
|----|---|-----|---|----|-----|-----|-----|----|---|-----|-------|
| 59 | 1 | 350 | 0 | 45 | 1 | 1,5 | 125 | 0 | 1 | 105 | Tidak |

E. Hasil Pengujian RapidMiner dan Python

Pengujian pada rapidminer dilakukan untuk mengetahui nilai akurasi algoritma *K-Nearest Neighbord (KNN)*, dataset penyakit gagal jantung dibagi dua dengan presentase 80% digunakan sebagai data latih dan 20% data uji. Pengujian dilakukan dengan menggunakan pergantian nilai K, nilai K yang digunakan merupakan nilai K terbaik yaitu, terdapat pada K=7. Berikut ini merupakan hasil perhitungan *confision matrix* pada *rapidminer*.

accuracy: 94.92%

| | true Meninggal | true Tidak | class precision |
|-----------------|----------------|------------|-----------------|
| pred. Meninggal | 0 | 0 | 0.00% |
| pred. Tidak | 3 | 56 | 94.92% |
| class recall | 0.00% | 100.00% | |

Gambar 2 *confision matrix* pada *rapidminer*

Perhitungan *confision matrix* pada *rapidminer* menghasilkan akurasi sebesar 94,92% dengan nilai *precision* meninggal 0,00% *precision* tidak 94,92% dan *recall* meninggal 0,00% *recall* Tidak 100%

Tabel 9 perbandingan akurasi pada *rapidminer*

| K | K=3 | K=5 | K=7 |
|---------|--------|--------|--------|
| Akurasi | 88.14% | 91.53% | 94.92% |

F. Pengujian KNN pada python

Setelah melakukan pengujian algoritma *KNN* pada *rapidminer* langkah selanjutnya yaitu, melakukan pengujian pada bahasa pemrograman python menggunakan *google colaboratory*.

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Meninggal | 0.75 | 0.14 | 0.24 | 21 |
| Tidak | 0.68 | 0.97 | 0.80 | 39 |
| accuracy | | | 0.68 | 60 |
| macro avg | 0.71 | 0.56 | 0.52 | 60 |
| weighted avg | 0.70 | 0.68 | 0.60 | 60 |

Gambar 3 *confision matrix* pada python

Hasil akurasi yang dihitung menggunakan *confision matrix* pada pemrograman python, menghasilkan akurasi sebesar 68% dengan *precision* meninggal 75% *precision* tidak 68% dan *recall* meninggal 14% *recall* meninggal 97%.

G. Hasil

Hasil dari penelitian ini yaitu, implementasi algoritma *KNN* yang dihitung secara manual menggunakan *Microsoft excel*, menghasilkan prediksi dari data testing adalah Tidak. Kemudian pengujian akurasi dilakukan menggunakan tool RapidMiner dan bahasa pemrograman python. Berikut dibawah ini merupakan hasil pengujian tool RapidMiner dan bahasa pemrograman python.

Tabel 7 hasil pengujian rapidminer dan python

| Hasil pengujian | Rapidminer | Python |
|-----------------|------------|--------|
| Akurasi | 94,92% | 68% |

| | | |
|----------------------------|--------|-----|
| <i>Error</i> | 5,98% | - |
| <i>Reccall</i> meninggal | 0,00% | 14% |
| <i>Reccall</i> tidak | 100% | 97% |
| <i>Precision</i> meninggal | 0,00% | 75% |
| <i>Precision</i> tidak | 94,92% | 68% |

Berdasarkan tabel diatas maka, hasil dari nilai akurasi tertinggi terdapat pada pengujian menggunakan rapidminer dengan akurasi sebesar 94,92%. Hasil pengujian pada bahasa pemrograman python memiliki hasil akurasi yang berbeda karena, data testing yang digunakan pada pengujian bahasa pemrograman python yaitu random, sedangkan pengujian pada rapidminer tidak. Kemudian proses pemodelan yang dilakukan pada bahasa pemrograman python dilakukan dengan pemodelan yang sederhana, sehingga cenderung memiliki nilai akurasi yang lebih kecil yaitu 68%.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan tahapan penelitian yang telah dilakukan pada penerapan algoritma *K-Nearest Neighbord* untuk memprediksi kematian akibat penyakit gagal jantung, kesimpulan yang diperoleh dari penelitian ini, yaitu, penerapan algoritma *KNN* terhadap data penyakit gagal jantung dilakukan dengan menghitung secara manual menggunakan Microsoft excell, kemudian dilakukan perhitungan jarak dari data testing terhadap data training dengan menggunakan tetangga terdekat $k=7$ yang menghasilkan kategori peristiwa kematian data testing, yaitu **Tidak**. Kemudian pengujian akurasi dilakukan dengan menggunakan aplikasi rapidminer dan bahasa pemrograman python. Pengujian pada rapidminer dilakukan dengan menggunakan pergantian nilai k , akurasi tertinggi didapat pada nilai $k=7$ dengan nilai akurasi 94,92%. Kemudian pengujian pada bahasa pemrograman python menghasilkan nilai akurasi 68%.

B. Saran

Pada penelitian selanjutnya disarankan untuk menggugurkan algoritma yang berbeda atau mengkomparasinya sehingga diketahui akurasi yang tertinggi. Kemudian jika penelitian ini diterapkan oleh praktisi kesehatan, disarankan untuk memakai jumlah nilai $k=7$ agar mendapatkan nilai akurasi yang baik.

PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Dede Andri Muhammad Reza Dengan Judul Penerapan Algoritma *K-Nearest Neighbord* Untuk Prediksi Kematian Akibat Gagal Jantung yang dibimbing Amril Mutoi Siregar, M.Kom dan Rahmat, M.Pd.

DAFTAR PUSTAKA

- [1] Kementerian Kesehatan RI. (2014). *Situasi Kesehatan Jantung*.
- [2] Kementerian Kesehatan RI. (2019). *Hari Jantung Sedunia*.
- [3] Sari, P. D., Yonata, A., Haryadi, & Swadharma, B. (2016). *Penatalaksanaan Gagal Jantung NYHA II disertai Pleurapneumonia pada Laki-laki Usia 38 Tahun*. *J Medula Unila*, 1-6.
- [4] Chicco, D., & Jurman, G. (2020). *Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone*. *BMC Medical Informatics and Decision Making*, 20(1), 1-16. <https://doi.org/10.1186/s12911-020-1023-5>
- [5] Fiandra, Y. A., Defit, S., & Yuhandri, Y. (2017). *Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10)*. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 1(2), 82. <https://doi.org/10.29207/resti.v1i2.48>
- [6] Mardi, Y. (2017). *Data Mining : Klasifikasi Menggunakan Algoritma C4.5*. *Jurnal Edik Informatika*, 2(2), 213-219.
- [7] Lestari, M. E. I. (2014). *Penerapan Algoritma Klasifikasi Nearest Neighbord (K-NN) Untuk Mendeteksi Penyakit Jantung*. 7(September 2010), 366-371.
- [8] Gultom, F., & Simanjuntak, T. (2020). *Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Dengan Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor*. *METHOMIKA: Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 4(1), 98-102. <https://doi.org/10.46880/jmika.v4i2.158>
- [9] Reza Noviansyah, M., Rismawan, T., & Marisa Midyanti, D. (2018). *Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya)*. *Jurnal Coding, Sistem Komputer Untan*, 06(2), 48-56. Retrieved from <http://jurnal.untan.ac.id/index.php/jcskommipa/article/view/26672>
- [10] Rosandy, T. (2016). *Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree Untuk Menganalisa Kelancaran Pembiayaan*. *Jurnal TIM Darmajaya*, 02(01), 52-62.