

## Penerapan Data Mining Terhadap Klasifikasi Pasien Penderita Penyakit Liver Menggunakan Metode *K-Nearest Neighbor*

Sabilla Aldana<sup>1\*</sup>, Jati Sasongko Wibowo<sup>2</sup>

Teknik Informatika, Universitas Stikubank (Unisbank) Semarang, Semarang, Indonesia

\*e-mail *Corresponding Author*: sabillaa804@gmail.com

### Abstract

*Liver sufferers are increasing from year to year. Liver disease is caused by an unhealthy lifestyle that can damage the liver. Liver disease is considered a silent killer because of the possibility of symptoms arising. Therefore, knowing the symptoms of liver disease early on is very necessary, so that sufferers can take appropriate treatment. This study implements the K-Nearest Neighbor algorithm in predicting liver disease in patients. The data used is the Indian Liver Patient Dataset (ILPD) taken from the UCI Machine Learning Repository. There are several stages of the classification process that will be carried out, including data separation, dividing test data and training data, KNN modeling, then analyzed using a confusion matrix and also an accuracy score. In this study, the results were obtained from the level of accuracy of the data, namely the value of accuracy, precision, recall, and also the f1-score with an accuracy value of 70%, a precision of 66.5%, a recall of 59.5%, and an f1-score of 59.5% of the highest K = 7 value. So the K-Nearest Neighbor algorithm is quite accurate for classifying patient data with liver disease because the data accuracy rate is above 50%.*

**Keywords:** *Classification; Liver Disease; K-Nearest Neighbor; Confusion Matrix*

### Abstrak

Penderita liver meningkat dari tahun ke tahun. Penyakit liver disebabkan oleh gaya hidup yang tidak sehat sehingga dapat merusak hati. Penyakit liver dianggap sebagai pembunuh diam-diam karena adanya kemungkinan timbul gejala karena itu mengetahui adanya gejala penyakit liver sejak dini sangat diperlukan, agar penderita dapat melakukan pengobatan dengan tepat. Penelitian ini mengimplementasikan algoritma *K-Nearest Neighbor* dalam memprediksi penyakit liver yang diderita oleh pasien. Data yang digunakan adalah *Indian Liver Patient Dataset* (ILPD) yang diambil dari *UCI Machine Learning Repository*. Terdapat beberapa tahapan proses klasifikasi yang akan dilakukan, antara lain pemisahan data, membagi data uji dan data latih, permodelan KNN, kemudian dianalisa menggunakan *confusion matrix* dan juga *accuracy score*. Pada penelitian ini didapatkan hasil dari tingkat keakuratan data yaitu nilai akurasi, presisi, recall, dan juga f1-score dengan nilai akurasi sebesar 70%, presisi sebesar 66,5%, recall sebesar 59,5%, dan f1-score sebesar 59,5% dari nilai K = 7 yang paling tertinggi. Jadi algoritma *K-Nearest Neighbor* cukup akurat untuk mengklasifikasi data pasien penderita penyakit liver dikarenakan tingkat keakuratan data diatas 50%.

**Kata kunci:** *Klasifikasi; Penyakit Liver; K-Nearest Neighbor; Confusion Matrix*

### 1. Pendahuluan

Penderita liver (hati) meningkat dari tahun ke tahun. Hal ini dikarenakan oleh gaya hidup yang tidak sehat misalnya pengkonsumsian alkohol, serta karena keterlambatan diagnosa penyakit [1]. Liver atau hati adalah organ yang vital bagi manusia. Organ ini terletak di dalam rongga perut sebelah kanan, tepatnya di bawah diafragma [2]. Liver bertanggung jawab untuk fungsi-fungsi kritis dalam tubuh, dimana hilangnya fungsi-fungsi tersebut dapat menyebabkan kerusakan yang signifikan pada tubuh. Liver adalah satu-satunya organ dalam tubuh yang dapat dengan mudah mengganti sel-sel yang rusak, tetapi jika sel-sel itu hilang, maka liver tidak mungkin dapat memenuhi kebutuhan tubuh. Penyakit liver sering disebut sebagai pembunuh

diam-diam karena kemungkinan tidak timbulnya gejala [3]. Permasalahan yang biasanya terjadi adalah sulitnya mengenali penyakit liver sejak dini, bahkan ketika penyakit tersebut sudah menyebar. Padahal mengetahui adanya gejala penyakit liver sejak dini ini sangat diperlukan, agar penderita dapat melakukan pengobatan dengan tepat. Dengan diagnosa adanya penyakit liver lebih awal dapat meningkatkan tingkat kelangsungan hidup pasien [4-6].

Dengan perkembangan teknologi saat ini diagnosa penyakit dapat menggunakan data mining. Data mining merupakan langkah penting dalam penemuan pengetahuan dari serangkaian dataset berukuran besar [7]. Salah satu pengembangan dari data mining adalah klasifikasi. Metode klasifikasi dapat melakukan pembelajaran dengan memetakan suatu item dan kedalam kelas berdasarkan kelas data yang telah didefinisikan sebelumnya. Untuk memecahkan masalah tersebut, pada penelitian ini menggunakan metode yang digunakan untuk membangun model klasifikasi dalam penyakit liver yaitu algoritma *K-Nearest Neighbor* (KNN). *K-Nearest Neighbor* pada umumnya sudah banyak digunakan peneliti karena mempunyai kelebihan, yaitu nilai akurasi yang tinggi dan tidak ada asumsi pada data [8-10].

Tujuan penelitian ini adalah menerapkan data mining terhadap klasifikasi pasien penderita penyakit liver menggunakan metode *K-Nearest Neighbor*. Kemudian Manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah mengetahui kinerja terbaik berdasarkan performa tingkat akurasi dari metode *K-Nearest Neighbor* dalam melakukan klasifikasi pasien penderita penyakit liver.

## 2. Tinjauan Pustaka

Data mining merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer secara otomatis untuk menganalisa dan mengekstraksi pengetahuan [11]. Penelitian [12] menerapkan algoritma *K-Nearest Neighbor* untuk pengujian dengan nilai K tertentu pada dataset iris bunga yang berasal dari *UCI Machine Learning* yang berjumlah sebanyak 135 data. Pengukuran jarak pada KNN dalam penelitian ini menggunakan *euclidean distance* dengan nilai ketetanggan 3,4,5,6,7,8,9. Hasil pengujian menunjukkan metode *K-Nearest Neighbor* dalam klasifikasi data memiliki akurasi persentase yang baik ketika menggunakan data random. Persentase variasi nilai K 3,4,5,6,7,8,9 memiliki persentase 100 %.

Selanjutnya menerapkan metode klasifikasi *K-Nearest Neighbor* pada data penderita penyakit diabetes. Penelitian ini bertujuan untuk mengukur performa metode klasifikasi dalam mengelola dataset penderita diabetes. Metode yang digunakan yaitu algoritma *K-Nearest Neighbor*, dimana merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Pada hasil akhir penelitian ini, telah dihitung akurasi tertinggi 39% pada K=3, presisi tertinggi 65% pada K=3 dan K=5, *recall* tertinggi 36% pada K=3, dan *F-Measure* tertinggi 46% pada K=3 [13].

Penelitian mengenai klasifikasi pasien penderita penyakit liver dengan pendekatan *machine learning*. Pada penelitian ini diterapkan metode untuk klasifikasi penyakit liver menggunakan *machine learning* dan dibandingkan hasilnya dengan metode klasik. Data yang digunakan adalah *Indian Liver Patient Dataset (ILPD)* yang diambil dari *UCI Machine Learning*. Peneliti melakukan analisis menggunakan metode regresi logistik, *decision tree*, *naïve bayes*, *k-nearest neighbor*, dan *support vector machine*. Berdasarkan nilai akurasi dan presisi metode SVM memberikan hasil yang terbaik, tapi berdasarkan *recall* metode *K-Nearest Neighbor* memberikan hasil terbaik [14].

Dari beberapa penelitian terkait sebelumnya dapat disimpulkan bahwa penelitian tentang klasifikasi menggunakan algoritma *K-Nearest Neighbor* sudah pernah dilakukan, tetapi objek dan metode penyelesaian yang digunakan berbeda-beda untuk setiap studi kasusnya. Pada dasarnya metode *K-Nearest Neighbor* telah banyak digunakan dengan menggunakan nilai k yang bervariasi dan lebih banyak menganalisa data numerik. Penelitian ini merujuk pada penelitian sebelumnya dengan menggunakan metode *K-Nearest Neighbor*. Diharapkan dengan metode ini mendapatkan hasil akurasi yang baik dalam melakukan klasifikasi pada penyakit liver dengan menggunakan beberapa nilai ketetanggan atau nilai k.

## 3. Metodologi

### 3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari *Website UCI Machine Learning Repository* 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 data dan 11 atribut. Pada penelitian ini menggunakan metode *K-Nearest Neighbor*, dimana nantinya pada dataset dilakukan splitting data dengan perbandingan 8:2 yang artinya 80% untuk data latih dan 20% untuk data uji dan menggunakan nilai *k* 3,5,7, dan 9.

### 3.2. Alur Penelitian

Untuk alur penelitian terdapat sebuah gambaran umum tentang tahapan proses dari penelitian tersebut. Berikut ini tahapan proses klasifikasi data menggunakan algoritma KNN yaitu:

#### 1) Penginputan Dataset

Pada tahap penginputan data langkah pertama yang dilakukan yaitu mengimport library utama yang digunakan pada klasifikasi KNN kemudian setelah itu bisa menginput dataset *ILPD (Indian Liver Patient Dataset)*.

#### 2) Preprocessing Data

##### 1. Pemisahan Data

Pada tahap pemisahan data ini dilakukan untuk memisahkan features dan label kedalam variabel *x* dan *y*. Variabel *x* adalah variabel yang memuat atribut dan variabel *y* adalah variabel yang memuat label kelas.

##### 2. Membagi data *training* (data latih) dan data *testing* (data uji)

Pada tahap pembagian data dilakukan untuk membagi data menjadi dua bagian yaitu data latih dan data uji dengan perbandingan 8:2 yang artinya 80% untuk data latih dan 20% untuk data uji.

##### 3. *Scaling* Data (Standarisasi Data)

Pada tahap *scaling* data digunakan agar jarak antar data tidak terlalu jauh. Untuk data yang perlu *discaling* adalah *x\_train* dan *x\_test*, untuk *y\_train* dan *y\_test* tidak perlu *discaling* karena isi dari datanya sudah berupa label.

#### 3) Melakukan Pengujian (*Training*) Menggunakan Algoritma KNN

Pada tahap *Training* KNN digunakan untuk uji klasifikasi pada data mining yang digunakan menggunakan algoritma *K-Nearest Neighbor*. Pada penelitian ini penulis melakukan *training* dengan menentukan nilai *K* yaitu 3, 5, 7, dan 9.

Algoritma *K-Nearest Neighbor* merupakan algoritma klasifikasi yang bekerja dengan mengambil sejumlah *K* data terdekat (tetangganya) sebagai acuan untuk menentukan kelas data baru. Untuk melakukan perhitungan metrik jarak pada algoritma KNN umumnya menggunakan rumus *Euclidean Distance*, berikut ini adalah cara untuk mencari jarak menggunakan rumus *Euclidean* [15]:

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \quad \dots \dots \dots (1)$$

Formula di atas dapat digunakan jika jumlah *independent variable* hanya ada satu variabel. Jika lebih dari satu *variable* maka dapat menggunakan rumus berikut ini.

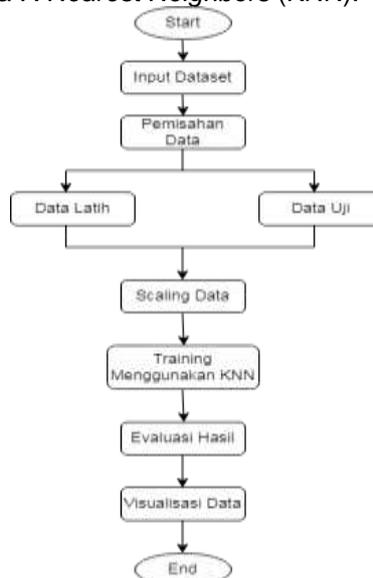
$$dis = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots} \quad \dots \dots \dots (2)$$

#### 4) Evaluasi dan Validasi

Pada tahap evaluasi dan validasi bertujuan untuk mengetahui hasil dari proses klasifikasi yang telah dilakukan. Pada proses evaluasi dan validasi kita akan mengetahui seberapa akurat hasil data mining yang digunakan pada penelitian ini dengan menggunakan *confusion matrix* dan juga *classification report* atau *accuracy score*.

**5) Visualisasi Data (Grafik)**

Pada tahap visualisasi data bertujuan untuk melihat grafik dari persebaran data menggunakan algoritma *K-Nearest Neighbors* (KNN).



Gambar 1. Tahapan Proses Klasifikasi Menggunakan KNN

**4. Hasil dan Pembahasan**

**4.1. Implementasi *K-Nearest Neighbor* Menggunakan *Python***

Pengujian data pada penelitian ini menggunakan pemrograman *Python* untuk permodelan algoritma *K-Nearest Neighbor* langkah pertama adalah import library utama yang akan digunakan kemudian menginput dataset *ILPD (Indian Liver Patient Dataset)*.

```

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

dataset = pd.read_csv('Indian Liver Patient Dataset (ILPD).csv')
  
```

Gambar 2. Import Dataset

Setelah menginput dataset maka dapat dicetak sampel data atau 5 data teratas pada dataset *ILPD (Indian Liver Patient Dataset)* yang dapat dilihat pada gambar 3 dibawah ini.

```

[4] dataset.head()
  
```

	age	gender	total_bilirubin	direct_bilirubin	alkphos	sgpt	sgot	total_protein	albumin	albumin_and_globulin_ratio	class
0	65	0	0.7	0.1	187	16	18	6.8	3.3	0.90	1
1	62	1	10.9	5.5	699	64	100	7.5	3.2	0.74	1
2	62	1	7.3	4.1	490	60	68	7.0	3.3	0.89	1
3	58	1	1.0	0.4	182	14	20	6.8	3.4	1.00	1
4	72	1	3.9	2.0	195	27	59	7.3	2.4	0.40	1

Gambar 3. Sampel Dataset ILPD (Indian Liver Patient Dataset)

Selanjutnya proses pemisahan data, pada tahap ini dilakukan untuk memisahkan *features* dan label kedalam variabel *x* dan *y*. Variabel *x* yang memuat atribut dan variabel *y* yang memuat label kelas.

```

x = dataset.iloc[:, [0,2]].values
y = dataset.iloc[:, -1].values
  
```

Gambar 4. Pemisahan Data

Pada gambar 4 pada variabel x yaitu atribut yang akan digunakan dalam penelitian ini adalah kolom *age* (umur) dan total bilirubin yaitu yang berada pada index 0 dan index 2. Kemudian pada variabel y yang akan digunakan adalah pada kolom terakhir yaitu *class*.

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.2, random_state=0)
```

Gambar 5. Pembagian Data

Gambar 5 merupakan proses untuk membagi data menjadi dua bagian yaitu data latih dan data uji dengan perbandingan 8:2 yang artinya 80% untuk data latih dan 20% untuk data uji. Dari proses pembagian data dihasilkan 466 data untuk data latih dan data uji berjumlah 177 data dari semua jumlah dataset yang berjumlah 583 data. Langkah selanjutnya yaitu melakukan permodelan algoritma *K-Nearest Neighbor* dengan rumus *Euclidean Distance* menggunakan fungsi *K-Neighbor Classifier*. Untuk nilai k yang akan digunakan pada penelitian ini adalah 3,5,7, dan 9 dimana prosesnya dapat dilihat pada gambar 6.

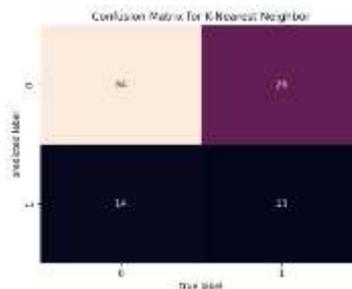
```
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=3, metric='euclidean', p=2)
classifier.fit(x_train, y_train)
y_pred = classifier.predict(x_test)
```

Gambar 6. Pemodelan *K-Nearest Neighbor*

## 4.2. Evaluasi hasil

Setelah melakukan pemodelan KNN langkah selanjutnya yaitu evaluasi dan validasi yang bertujuan untuk mengetahui hasil dari proses klasifikasi yang telah dilakukan. Pada proses evaluasi ini bisa dilakukan dengan menggunakan *confusion matrix* dan juga *classification report* (*accuracy score*).

### 1) Evaluasi Hasil Berdasarkan Nilai K = 3



Gambar 7. Hasil *Confusion Matrix* K=3

Dapat dilihat pada gambar 7 adalah hasil dari *confusion matrix* dengan nilai k = 3. Hasilnya adalah terdapat 64 data yang diprediksi benar (positif) kemudian 13 data yang diprediksi salah (negatif) sedangkan 26 data dan 14 data adalah yang diprediksi kurang tepat.

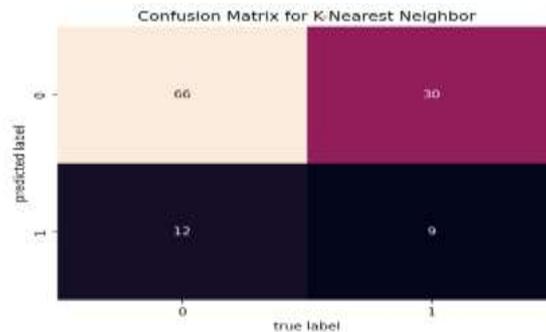
	precision	recall	f1-score	support
1	0.71	0.82	0.76	78
2	0.48	0.33	0.39	39
accuracy			0.66	117
macro avg	0.60	0.58	0.58	117
weighted avg	0.63	0.66	0.64	117

Gambar 8. Hasil dari *Classification Report* K=3

Kemudian untuk hasil *accuracy score* pada nilai k = 3 dapat dilihat pada gambar 8. Berdasarkan hasil yang didapatkan bahwa kelas penyakit liver positif mendapatkan nilai presisi

71%, *recall* 82%, dan *f1-score* 76%. Sedangkan kelas penyakit liver negatif mendapatkan nilai presisi 48%, *recall* 33%, dan *f1-score* 39%. Kemudian untuk nilai akurasi didapatkan 66%.

2) Evaluasi Hasil Berdasarkan Nilai K = 5



Gambar 9. Hasil *Confusion Matrix* K=5

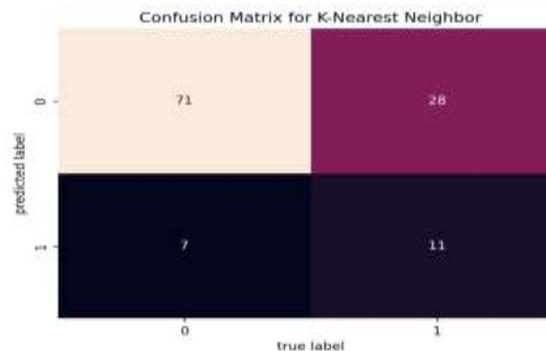
Dapat dilihat pada gambar 9 adalah hasil dari *confusion matrix* dengan nilai k = 3. Hasilnya adalah terdapat 66 data yang diprediksi benar (positif) kemudian 9 data yang diprediksi salah (negatif) sedangkan 30 data dan 12 data adalah yang diprediksi kurang tepat.

	precision	recall	f1-score	support
1	0.69	0.85	0.76	78
2	0.43	0.23	0.30	39
accuracy			0.64	117
macro avg	0.56	0.54	0.53	117
weighted avg	0.60	0.64	0.61	117

Gambar 10. Hasil dari *Classification Report* K=5

Dapat dilihat pada gambar 10 itu adalah hasil *accuracy score* pada nilai k = 5. Berdasarkan hasil yang didapatkan bahwa kelas penyakit liver positif mendapatkan nilai presisi 69%, *recall* 85%, dan *f1-score* 76%. Sedangkan kelas penyakit liver negatif mendapatkan nilai presisi 43%, *recall* 23%, dan *f1-score* 30%. Kemudian untuk nilai akurasi didapatkan 64%.

3) Evaluasi Hasil Berdasarkan Nilai K = 7



Gambar 11. Hasil *Confusion Matrix* K=7

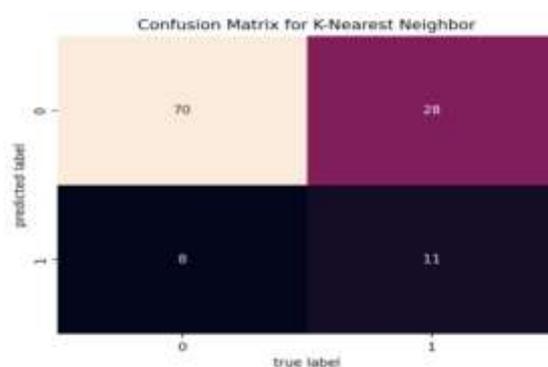
Dapat dilihat pada gambar 11 adalah hasil dari *confusion matrix* dengan nilai k = 7. Hasilnya adalah terdapat 71 data yang diprediksi benar (positif) kemudian 11 data yang diprediksi salah (negatif) sedangkan 28 data dan 7 data adalah yang diprediksi kurang tepat.

	precision	recall	f1-score	support
1	0.72	0.91	0.80	78
2	0.61	0.28	0.39	39
accuracy			0.70	117
macro avg	0.66	0.60	0.59	117
weighted avg	0.68	0.70	0.66	117

Gambar 12. Hasil dari *Classification Report* K=7

Pada gambar 12 merupakan hasil *accuracy score* pada nilai  $k = 7$ . Berdasarkan hasil yang didapatkan bahwa kelas penyakit liver positif mendapatkan nilai presisi 72%, *recall* 91%, dan *f1-score* 80%. Sedangkan kelas penyakit liver negatif mendapatkan nilai presisi 61%, *recall* 28%, dan *f1-score* 39%. Kemudian untuk nilai akurasi didapatkan 70%.

#### 4) Evaluasi Hasil Berdasarkan Nilai K = 9



Gambar 13. Hasil *Confusion Matrix* K=9

Dapat dilihat pada gambar 13 adalah hasil dari *confusion matrix* dengan nilai  $k = 9$ . Hasilnya adalah terdapat 70 data yang diprediksi benar (positif) kemudian 11 data yang diprediksi salah (negatif) sedangkan 28 data dan 8 data adalah yang diprediksi kurang tepat.

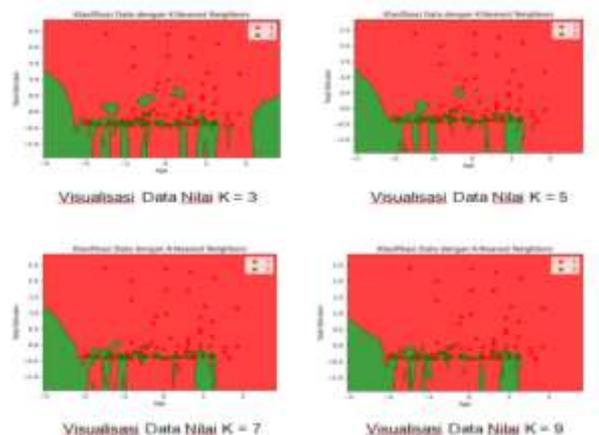
	precision	recall	f1-score	support
1	0.71	0.90	0.80	78
2	0.58	0.28	0.38	39
accuracy			0.69	117
macro avg	0.65	0.59	0.59	117
weighted avg	0.67	0.69	0.66	117

Gambar 14. Hasil dari *Classification Report* K=9

Pada gambar 14 merupakan hasil *accuracy score* pada nilai  $k = 9$ . Berdasarkan hasil yang didapatkan bahwa kelas penyakit liver positif mendapatkan nilai presisi 71%, *recall* 90%, dan *f1-score* 80%. Sedangkan kelas penyakit liver negatif mendapatkan nilai presisi 58%, *recall* 28%, dan *f1-score* 38%. Kemudian untuk nilai akurasi didapatkan 69%.

### 4.3. Visualisasi Data

Setelah tahapan selesai dilakukan semua, langkah selanjutnya adalah melakukan visualisasi data. Visualisasi Data bertujuan untuk melihat grafik dari persebaran data menggunakan algoritma *K-Nearest Neighbors* (KNN). Dan yang digunakan untuk visualisasi data pada penelitian ini adalah data testing (data uji) yang berjumlah 117 data dari 583 data pada dataset ILPD (*Indian Liver Patient Dataset*).



Gambar 15. Hasil Visualisasi Data

Pada gambar 15 adalah Hasil dari visualisasi data atau grafik menggunakan algoritma *K-Nearest Neighbor* berdasarkan nilai K yang sudah ditentukan yaitu 3,5,7, dan 9. Dapat dilihat pada gambar tersebut bahwa ada gambar daerah yang berwarna merah dan juga daerah yang berwarna hijau, itu artinya pada daerah yang berwarna merah adalah kelas pasien yang dinyatakan positif menderita penyakit liver dan daerah yang berwarna hijau adalah kelas pasien yang dinyatakan negatif atau tidak menderita penyakit liver berdasarkan kolom *age* (umur) dan total pada bilirubin. Apabila terdapat titik yang berwarna merah ada di daerah yang berwarna hijau itu menandakan bahwa data yang prediksinya kurang tepat karena nilai akurasi tidak sampai dengan 100%.

**4.4. Pembahasan**

Berdasarkan hasil evaluasi pada klasifikasi pasien penderita penyakit liver, didapat bahwa nilai akurasi tertinggi diperoleh oleh nilai K=7 sebesar **70%**, dilanjutkan oleh K=9 sebesar **69%**, K=3 sebesar **65%**, K=5 sebesar **64%**. Kemudian untuk nilai presisi tertinggi diperoleh oleh nilai K=7 sebesar **66,5%**, dilanjutkan oleh K=9 sebesar **64,5%**, K=3 sebesar **59,5%**, K=5 sebesar **56%**. Sedangkan nilai recall tertinggi diperoleh oleh nilai K=7 sebesar **59,5%**, dilanjutkan oleh K=9 sebesar **59%**, K=3 sebesar **57,5%**, K=5 sebesar **54%**. Dan terakhir untuk nilai f1-score tertinggi diperoleh oleh nilai K=7 sebesar **59,5%**, dilanjutkan oleh K=9 sebesar **59%**, K=3 sebesar **57,5%**, K=5 sebesar **53%**. Hasil pengujian keakuratan data tersebut menunjukkan bahwa kinerja pada nilai ketetapan atau K=7 lebih tinggi dari nilai K yang lainnya, dengan demikian dapat dikatakan bahwa kinerja algoritma *K-Nearest Neighbor* pada nilai K=7 adalah yang paling baik dari nilai K yang lainnya. Temuan ini sama dengan beberapa temuan terdahulu yang juga melakukan klasifikasi penyakit liver tetapi pada hasil yang didapatkan berbeda karena pada temuan tersebut kinerja algoritma *K-Nearest Neighbor* yang terbaik diperoleh pada nilai K=9 [2] dan nilai K=10 [16].

**5. Simpulan**

Berdasarkan hasil penelitian mengenai klasifikasi pasien penderita penyakit liver menggunakan metode *K-Nearest Neighbor* (KNN) didapatkan beberapa kesimpulan yaitu setelah melakukan 4 kali percobaan menggunakan metode *K-Nearest Neighbor* berdasarkan nilai K yaitu 3,5,7, dan 9 mendapatkan hasil keakuratan data yang terbaik adalah pada nilai K=7 yaitu nilai akurasi sebesar 70%, presisi 66,5%, recall 59,5%, dan f1-score 59,5%. Jadi algoritma *K-Nearest Neighbor* dinilai cukup akurat untuk mengklasifikasi data pasien penderita penyakit liver karena mendapat nilai akurasi diatas 50%. Saran untuk penelitian berikutnya bisa menggunakan algoritma yang lebih baru dan dapat dikembangkan dalam bentuk implementasi sebuah sistem berupa *website* atau aplikasi yang bertujuan untuk mempermudah masyarakat umum.

**Daftar Referensi**

[1] E. Pusporani, S. Qomariyah, and Irhamah, "Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan *Machine Learning*," *Inferensi.*, vol. 2, No. 1, pp. 25-32Maret 2019. [online]. Tersedia: <https://iptek.its.ac.id/index.php/inferensi/article/view/6810>

- [2] E.Y.Habibie, H.Oktavianto, and Q.A'yun, "Penerapan Algoritma *K-Nearest Neighbor* Untuk Klasifikasi Penyakit Liver," *Repository UM Jember*, Maret 2021. [online]. Tersedia: <http://repository.unmuhjember.ac.id/8680/>
- [3] M.R.Akhmad and T.A.Y.Siswa, "Implementasi *K-Nearest Neighbor* Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi," *Progresif J. Ilm. Komput.*, vol. 18, No. 2, hal. 185-192, Agustus 2022.
- [4] P.Putra, A.M.H.Pardede, and S.Syahputra, "Analisis Metode *K-Nearest Neighbor* (KNN) Dalam Klasifikasi Data Iris Bunga," *J. Tekn. Inform. Kaput.*, vol. 6, No. 1, pp. 297-305, 2022.
- [5] D.Hernandi and A.Jananto, "Penerapan Metode *K-Means* Untuk Memprediksi Pola Kedatangan Penduduk di Kota Jakarta," *Progresif J. Ilm. Komput.*, vol. 19, No. 1, hal. 34-48, Februari 2023.
- [6] E. Pusporani, S. Qomariyah, & I. Irhamah. "Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning. *Inferensi*, vol. 2, no. 1 pp. 25-32, 2019.
- [7] A.P. Ayudhitama, & U. Pujiyanto, "Analisa 4 Algoritma Dalam klasifikasi liver menggunakan RAPIDMINER". *Jurnal Informatika Polinema*, vol. 6, no. 2, pp. 1-9, 2020
- [8] A.M.Argina, "Penerapan Metode Klasifikasi *K-Nearest Neighbor* Pada Dataset Penderita Penyakit Diabetes," *Ind. J. of. Dat. Scien*, vol. 1, No. 2, pp. 29-23 Juli 2022.
- [9] H. Said, N. Matondang, and H.N. Irmanda "Penerapan Algoritma *K-Nearest Neighbor* Untuk Memprediksi Kualitas Air yang Dapat Dikonsumsi," *Techno.COM*, vol. 21, No. 2, pp. 256-267, Mei 2022.
- [10] A. Ali, M.A. Alrubei, L.F.M. Hassan, M.A. Al-Ja'afari, & S.H. Abdulwahed, Diabetes Diagnosis based on KNN. *IJUM Engineering Journal*, vo. 21, no. 1, pp. 175-181, 2020.
- [11] R.W. Putri, A. Ristyawan, & M.N. Muzaki, "Comparison Performance of K-NN and NBC Algorithm for Classification of Heart Disease". *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem dan Komputer*, vol. 2, no. 2, pp. 143-154, 2022.
- [12] Yahya, W.P.Hidayanti, "Penerapan Algoritma *K-Nearest Neighbor* Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada Lombok Vape On," *Infotek. J. Inform. Tekn*, vol. 3, No. 2, pp. 104-114 Juli 2020.
- [13] R.Fitriani and Dr.Rb.F.Hakim, "Analisis dan Implementasi Enam Metode *Machine Learning* Pada Dataset *Liver Disorders*," [online]. Tersedia: [https://www.academia.edu/37102691/ANALISIS\\_DAN\\_IMPLEMENTASI\\_ENAM\\_METODE\\_MACHINE\\_LEARNING\\_PADA\\_DATASET\\_LIVER\\_DISORDERS](https://www.academia.edu/37102691/ANALISIS_DAN_IMPLEMENTASI_ENAM_METODE_MACHINE_LEARNING_PADA_DATASET_LIVER_DISORDERS)
- [14] I.W.A.Purnawibawa, I.N.Purnama, and I.N.Y.A.Wijaya, "Komparasi Algoritme *K-Nearest Neighbor* dan *Support Vector Machines* Dalam Prediksi Layanan Produk *ICONNET*," *Progresif J. Ilm. Komput.*, vol. 18, No. 2, hal. 271-282, Agustus 2022.
- [15] S. Soegiarto, & B. Bahar, "Model Aplikasi Penentuan Jenis Beasiswa Berbasis Algoritma K-NN Termodifikasi". *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 5, no. 3, pp. 1181-1186, 2017.
- [16] S. Z. H. Rukmana, A. Aziz, & W. Harianto, "Optimasi Algoritma *K-Nearest Neighbor* (Knn) Dengan Normalisasi Dan Seleksi Fitur Untuk Klasifikasi Penyakit Liver". *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 439-445, 2022