

**EFEKTIVITAS *ELBOW METHOD* TERHADAP NILAI K PADA  
ALGORITMA *K-NEAREST NEIGHBOR* (K-NN) UNTUK  
KLASIFIKASI PENYAKIT *LIVER***

**SKRIPSI**



**OLEH:**

**IHYA' NASHIRUDIN ABRAR**  
**NIM. 181220082**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN ILMU KOMPUTER  
UNIVERSITAS MUHAMMADIYAH PONTIANAK  
2022**

**PERNYATAAN KEASLIAN**

**EFEKTIVITAS *ELBOW METHOD* TERHADAP NILAI *K* PADA  
ALGORITMA *K-NEAREST NEIGHBOR (K-NN)* UNTUK  
KLASIFIKASI PENYAKIT *LIVER***

**SKRIPSI**

Saya menyatakan bahwa Skripsi ini merupakan tulisan hasil kerja saya sendiri dan bukan orang lain, kecuali kutipan dan ringkasan yang sudah dicantumkan sumbernya.

Pontianak, 23 Desember 2022



Nashirudin Abrar  
NIM. 181220082

## LEMBAR PERSETUJUAN

Yang bertanda tangan dibawah ini, Dosen Pembimbing Skripsi,  
menerangkan bahwa:

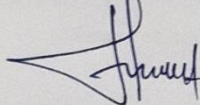
Nama : Ihya' Nashirudin Abrar

NIM : 181220082

Judul : Efektivitas *Elbow Method* Terhadap Nilai K Pada Algoritma *K-Nearest Neighbor* (K-NN) Untuk Klasifikasi Penyakit *Liver*.

## DIPERIKSA DAN DISETUJUI

Dosen Pembimbing I



Asrul Abdullah, S.Kom, M.Cs  
NIDN. 1128059002

Dosen Pembimbing II

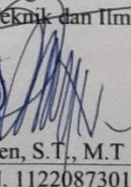


Sucipto, M.Kom  
NIDN. 1130038301

Mengetahui

Dekan Fakultas Teknik dan Ilmu Komputer



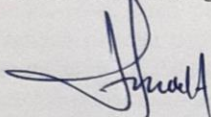
  
Puizen, S.T., M.T  
NIDN. 1122087301

## LEMBAR PENGESAHAN

Skripsi ini telah disidangkan dan dipertahankan di depan tim penguji pada hari Senin, tanggal 23 bulan Desember tahun 2021 dan diterima sebagai salah satu syarat akhir studi pada Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer Universitas Muhammadiyah Pontianak.

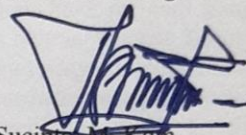
### TIM PEMBIMBING

Dosen Pembimbing I



Asrul Abdullah, S.Kom, M.Cs  
NIDN. 1128059002

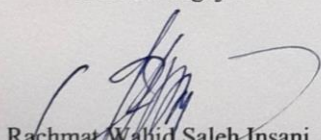
Dosen Pembimbing II



Sucipto, M. Kom  
NIDN. 1130038301

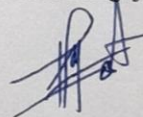
### TIM PENGUJI

Dosen Penguji I



Rachmat Walid Saleh Insani,  
S.Kom., M.Cs  
NIDN. 1120079001


Dosen Penguji II



Syarifah Putri Agustini, S.T.,  
M.Kom  
NIDN. 1111088803

Mengetahui  
Dekan Fakultas Teknik dan Ilmu Komputer



  
Dekan, S.T., M.T  
NIDN. 1122087301

## ABSTRAK

Dalam dunia kesehatan mendiagnosis penyakit hati tidaklah mudah, namun ada beberapa rekam medis yang mencatat gejala penyakit pasien dan diagnosis nya ini sangat bermanfaat bagi peneliti karena rekam medis yang ada dapat digunakan untuk membantu mendiagnosa penyakit hati, dengan melakukan metode *data mining* yang mana *data mining* dapat digunakan untuk memeriksa nilai tambahan dari kumpulan informasi sebagai informasi yang tidak dapat dilihat secara fisik dengan menggunakan perhitungan *K-Nearest Neighbor* (K-NN). Metode ini bekerja lebih efektif dibandingkan dengan metode lain, karena metode ini mencoba untuk mengelompokkan informasi baru dari kelas yang tidak jelas dengan memilih jumlah *k* yang paling dekat dengan informasi baru. Sebagian besar dari kelas *k* yang menyertainya dipilih sebagai kelas harapan untuk informasi baru dan menggunakan *Elbow Method* dalam menentukan *error rate* dalam menentukan nilai *k* optimal, tujuannya adalah untuk mendapatkan tingkat akurasi terbaik dari metode *K-Nearest Neighbor* (K-NN), hasil pengujian menunjukkan pengujian ke-3 menjadi yang terbaik dengan akurasi *training* 80,5% *testing* 78,9% dan akurasi setelah *Tuning* pada data *training* 82,2% dan *testing* 77,1% tanpa melakukan penyeimbangan data karena setelah melakukan penyeimbangan data menjadi *overfitting* bisa jadi karena data *resampling* tidak sesuai dengan data sesungguhnya sehingga model terlalu kompleks dalam melatih jumlah *noise* yang mengakibatkan *overfitting*.

**Kata Kunci:** *Classification, Machine Learning, Elbow Method, KNN, Liver.*

## **ABSTRACT**

*In the world of health diagnosing the liver disease is not easy, but several medical records record the patient's symptoms and, the diagnosis is very useful for researchers because existing medical records can be used to help diagnose liver disease, by carrying out data mining methods where data mining can be used to examine the additional value of a collection of information as information that cannot be seen physically by using the K-Nearest Neighbor (K-NN) calculation. This method works more effectively than the other methods because this method tries to classify new information from unclear classes by choosing the k amount that is closest to the new information. Most of the k classes that accompany it are selected as the hope class for new information and use the Elbow Method in determining the error rate in determining the optimal k value, the aim is to get the best level of accuracy from the K-Nearest Neighbor (K-NN) method, the test results show the 3rd test is the best with a training accuracy of 80.5% testing 78.9% and accuracy after tuning on the training data 82.2% and testing 77.1% without balancing the data because after balancing the data it becomes overfitting it could be because the resampling data does not match the actual, data so the model is too complex in training the amount of noise which results in overfitting.*

**Keywords:** *Classification, Machine Learning, Elbow Method, KNN, Liver.*

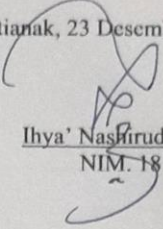
## KATA PENGANTAR

Alhamdulillah segala puji dan syukur kepada Allah SWT, penulis dapat menyelesaikan Skripsi ini yang berjudul “EFEKTIVITAS *ELBOW METHOD* TERHADAP NILAI *K* PADA ALGORITMA *K-NEAREST NEIGHBOR* (K-NN) UNTUK KLASIFIKASI PENYAKIT *LIVER*” atas motivasi yang telah diberikan kepada penulis, oleh karena itu penulis banyak mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua Orang Tua tercinta yang menjadi dasar motivasi dalam menyelesaikan Skripsi ini. Banyak sekali dukungan yang telah diberikan kepada penulis baik secara moral maupun materi.
2. Bapak Asrul Abdullah, S., Kom. M., CS. sebagai pembimbing utama yang telah memberikan bimbingan dan arahan yang sangat berguna dalam menyelesaikan Skripsi ini.
3. Bapak Sucipto, M., Kom. selaku pembimbing kedua dan memberikan bimbingan serta semangat kepada penulis dalam menyelesaikan Skripsi ini.
4. Seluruh Dosen dan tenaga Dosen yang pernah mengajar di Program Studi Teknik Informatika Fakultas Teknik yang sudah memberikan ilmu dari awal perkuliahan hingga sekarang.
5. Kepada kekasih, sahabat, dan teman teman kelas Teknik Informatika, banyak suka duka yang telah dilalui semasa perkuliahan hingga sampai penulis menyelesaikan skripsi ini.
6. Kepala dan staf perpustakaan Universitas Muhammadiyah Pontianak yang telah membantu penulis dalam mencari referensi.

Semoga skripsi ini bermanfaat bagi universitas dan pengembangan teknologi informasi dimasa depan.

Pontianak, 23 Desember 2022

  
Ihyar Nasfirudin Abrar  
NIM. 181220082



## DAFTAR ISI

PERNYATAAN KEASLIAN.....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN .....	iii
ABSTRAK .....	iv
KATA PENGANTAR .....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian .....	4
1.7 Sistematika Penulisan.....	4
BAB II LANDASAN TEORI .....	6
2.1 Tinjauan Pustaka .....	6
2.2 <i>Machine Learning</i> .....	11
2.3 <i>Data Mining</i> .....	11
2.3.1 Operasi <i>Data Mining</i> .....	11
2.3.2 Teknik <i>Data Mining</i> .....	12
2.4 <i>Knowledge Discovery in Database</i> .....	12
2.5 <i>K-Nearest Neighbor (K-NN)</i> .....	14
2.6 Klasifikasi.....	15
2.7 <i>Confusion Matrix</i> .....	15
2.8 AUC ( <i>Area Under Curve</i> ) ROC ( <i>Receiver Operating Characteristic</i> )..	17
2.9 Hati .....	18

2.10 Python.....	19
2.11 <i>Elbow Method</i> .....	19
2.12 CSS ( <i>Cascading Style Sheet</i> ).....	20
2.13 HTML.....	20
BAB III METODE PENELITIAN.....	21
3.1 Studi Pustaka .....	22
3.2 Mendefinisikan Masalah .....	22
3.3 Pengumpulan Data.....	22
3.4 Proses <i>Data Mining</i> .....	22
3.5 Integrasi Model dan <i>Website</i> .....	23
3.6 Kesimpulan.....	23
BAB IV ANALISIS DAN PERANCANGAN SISTEM.....	24
4.1 Analisis Kebutuhan Non-Fungsional .....	24
4.2 Analisis Kebutuhan Fungsional.....	25
4.3 Analisis Kebutuhan Data.....	25
4.4 Perancangan Proses Diagram Aliran .....	27
4.4.1 Perancangan Proses <i>Explore Data Analysis (EDA)</i> .....	28
4.4.2 Perancangan Proses <i>Preprocessing</i> .....	29
4.4.3 Perancangan Proses Membangun Model .....	30
4.4.4 Perancangan Proses Integrasi Model dan <i>Website</i> .....	30
4.5 Perancangan Antarmuka.....	31
4.5.1 Perancangan Antarmuka Beranda .....	31
4.5.2 Perancangan Antarmuka Prediksi .....	31
4.5.3 Perancangan Antarmuka Tentang .....	32
4.6 Perancangan Pengujian.....	33
BAB V HASIL DAN PENGUJIAN .....	34
5.1 Hasil <i>Exploratory Data Analysis (EDA)</i> .....	34
5.1.1 <i>Univariate Analysis</i> .....	34

5.1.2	<i>Bivariate Analysis</i> .....	36
5.1.3	<i>Multivariate Analysis</i> .....	37
5.2	Hasil <i>Preprocessing</i> atau <i>Data Cleaning</i> .....	39
5.2.1	<i>Missing value</i> .....	39
5.2.2	<i>Duplicate Data</i> .....	40
5.2.3	<i>Outlier</i> .....	41
5.2.4	<i>Imbalanced Data</i> .....	43
5.2.5	<i>Label Encoding</i> .....	45
5.2.6	Normalisasi .....	45
5.3	Hasil Membangun dan Melatih Model.....	47
5.3.1	<i>Elbow method</i> .....	47
5.3.2	Pemodelan <i>K-Nearest Neighbor (K-NN)</i> .....	48
5.3.3	<i>Hyperparameter Tuning</i> dan <i>GridSearchCV</i> .....	49
5.4	Hasil <i>Integrasi Model dan Web</i> .....	52
5.4.1	Hasil <i>Integrasi</i> .....	52
5.4.2	Hasil Halaman Beranda.....	52
5.4.3	Hasil Halaman Prediksi.....	54
5.4.4	Hasil Halaman Tentang.....	57
5.5	Pengujian .....	58
5.5.1	Evaluasi Hasil Pengujian.....	70
BAB VI KESIMPULAN DAN SARAN .....		72
6.1	Kesimpulan.....	72
6.2	Saran.....	72
DAFTAR PUSTAKA .....		74

## DAFTAR GAMBAR

Gambar 2.1 Tahapan Dalam KDD.....	13
Gambar 2.2 Confusion Matrix .....	16
Gambar 3.1 Diagram Alir Penelitian .....	21
Gambar 4.1 Perancangan Proses Diagram Aliran.....	27
Gambar 4.2 Perancangan Proses Explore Data Analysis (EDA).....	28
Gambar 4.3 Perancangan Proses Preprocessing.....	29
Gambar 4.4 Perancangan Proses Membangun Model .....	30
Gambar 4.5 Perancangan Proses Integrasi Model dan Website.....	30
Gambar 4.6 Perancangan Halaman Beranda.....	31
Gambar 4.7 Perancangan Halaman Prediksi .....	32
Gambar 4.8 Perancangan Halaman Tentang .....	32
Gambar 5.1 Univariate Analysis Persebaran Usia .....	34
Gambar 5.2 Univariate Analysis Gender .....	35
Gambar 5.3 Univariate Analysis Dataset .....	35
Gambar 5.4 Univariate Analysis Persebaran Data Semua Fitur .....	36
Gambar 5.5 Bivariate Analysis Histogram Age with Disease .....	37
Gambar 5.6 Multivariates Analysis Data Describe .....	37
Gambar 5.7 Multivariates Analysis Korelasi Setiap Fitur .....	38
Gambar 5.8 Memeriksa Missing Value .....	39
Gambar 5.9 Memeriksa data Duplikat .....	40
Gambar 5.10 Memeriksa Outlier Pada Direct Bilirubin .....	41
Gambar 5.11 Memeriksa Outlier Pada Setiap Fitur .....	41
Gambar 5.12 Menangani Outlier Pada Direct Bilirubin .....	42
Gambar 5.13 Menangani Outlier Pada Setiap Fitur .....	42
Gambar 5.14 Mengecek Data Tidak Seimbang .....	43
Gambar 5.15 Menangani Data Tidak Seimbang .....	44
Gambar 5.16 Mengecek Distribusi Tipe Data.....	45
Gambar 5.17 Hasil Normalisasi Data Menggunakan MinMaxScaler.....	46
Gambar 5.18 Hasil Elbow Method Error Rate VS K Value .....	47
Gambar 5.19 Hasil Pencarian K Terbaik Menggunakan Elbow Method.....	48
Gambar 5.20 Hasil Pengecekan Akurasi Data Training dan Testing.....	49
Gambar 5.21 Hasil <i>Grid Search Best Params</i> .....	50
Gambar 5.22 Hasil Akurasi Data Training dan Testing Setelah Tuning .....	51

Gambar 5.23 Halaman Beranda .....	53
Gambar 5.24 Halaman Beranda Informasi Tentang Penyakit Liver .....	53
Gambar 5.25 Halaman Prediksi Form Inputan .....	54
Gambar 5.26 Halaman Prediksi Bantuan .....	54
Gambar 5.27 Halaman Prediksi Tampilan Terkena Liver .....	56
Gambar 5.28 Halaman Prediksi Tampilan Tidak Terkena Liver .....	57
Gambar 5.29 Halaman Tentang Profil Peneliti .....	57
Gambar 5.30 Pengujian Pertama Training .....	58
Gambar 5.31 Pengujian Pertama Testing .....	59
Gambar 5.32 Pengujian Pertama Training Setelah Tuning .....	60
Gambar 5.33 Pengujian Pertama Testing Setelah Tuning .....	60
Gambar 5.34 Pengujian Kedua Training .....	61
Gambar 5.35 Pengujian Kedua Testing .....	62
Gambar 5.36 Pengujian Kedua Training Setelah Tuning .....	63
Gambar 5.37 Pengujian Kedua Testing Setelah Tuning .....	63
Gambar 5.38 Pengujian Ketiga Training .....	64
Gambar 5.39 Pengujian Ketiga Testing .....	65
Gambar 5.40 Pengujian Ketiga Training Setelah Tuning .....	66
Gambar 5.41 Pengujian Ketiga Testing Setelah Tuning .....	67
Gambar 5.42 Pengujian Keempat Training .....	68
Gambar 5.44 Pengujian Keempat Training Setelah Tuning .....	69
Gambar 5.45 Pengujian Keempat Testing Setelah Tuning .....	70

## DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka .....	6
Tabel 2.2 Nilai AUC (Area Under Curve) .....	18
Tabel 4.1 Analisis Kebutuhan Fungsional .....	25
Tabel 4.2 Fitur dan Keterangan.....	26
Tabel 4.3 Perancangan Tabel Pengujian Akurasi .....	33
Tabel 5.1 Deskripsi Form.....	55
Tabel 5.2 Hasil Pengujian Akurasi.....	71

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Fungsi organ hati bagi tubuh manusia berperan sangatlah penting, yaitu menghilangkan racun dalam darah hingga berfungsi untuk membantu proses pencernaan, jika organ hati ini tidak dapat berfungsi dengan baik dapat menyebabkan masalah serius pada tubuh, Penyakit *liver* ini menjadi salah satu penyakit yang mematikan di dunia. Organisasi Kesehatan Dunia (WHO) menyebut bahwa 1,34 juta kematian yang terjadi pada 2015 disebabkan penyakit hati. oleh karena itu prediksi akurat penyakit sangat penting dalam industri kesehatan dan medis. Untuk mengambil keputusan penting yang efektif dalam menganalisis dan memprediksi penyakit yang diderita [1].

Dalam dunia perawatan kesehatan saat ini, mendiagnosis penyakit hati sangatlah sulit. Namun, ada beberapa rekam medis yang mencatat gejala penyakit pasien dan diagnosisnya sebagai penyakit hati. Hal-hal seperti ini tentunya sangat bermanfaat bagi peneliti. Rekam medis yang ada dapat digunakan untuk membantu mendiagnosa penyakit pasien [2], Hasil tes fungsi hati dapat dijadikan acuan untuk mendiagnosis ada tidaknya penyakit hati. Tes-tes ini termasuk serum transaminase, alkaline phosphatase, bilirubin total, bilirubin terkonjugasi, protein total, albumin, dan rasio albumin terhadap globulin.

Salah satu strategi yang dapat digunakan untuk mengantisipasi penyakit hati adalah dengan memanfaatkan *data mining*. *Data mining* adalah perkembangan siklus yang digunakan untuk memeriksa nilai tambahan dari kumpulan informasi sebagai informasi yang tidak dapat dilihat secara fisik [3], Pada skripsi ini, penulis menggunakan perhitungan *K-Nearest Neighbor* (K-NN). Strategi urutan informasi bekerja agak efektif dibandingkan dengan teknik karakterisasi informasi yang berbeda. Perhitungan ini mencoba untuk mengelompokkan informasi baru dari kelas yang tidak jelas memilih k yang paling dekat dengan informasi baru.

Sebagian besar dari kelas informasi  $k$  yang menyertainya dipilih sebagai kelas harapan untuk informasi baru.  $k$  biasanya ditentukan sebagai angka ganjil untuk mencoba tidak menunjukkan jumlah jarak yang sama dalam sistem pengelompokan untuk informasi baru dan menggunakan *Elbow Method* dalam menentukan *error rate* dalam menentukan nilai  $k$  optimal, tujuannya adalah untuk mendapatkan tingkat akurasi terbaik dari metode *K-Nearest Neighbor* (K-NN).

Keakuratan metode *K-Nearest Neighbor* (K-NN) dalam mendiagnosa sebuah penyakit telah teruji dari penelitian terdahulu yang menggunakan metode ini dalam memprediksi atau mengklasifikasi sebuah penyakit. Seperti pada hasil penelitian yang telah dilakukan oleh M. Syukri Mustafa, dan I Wayan Simpen dalam memprediksi pasien yang terjangkit penyakit diabetes dan mendapatkan hasil akurasi terbaik sebesar 93,33% dengan error rate 6,67% [4].

Berdasarkan permasalahan yang dapat ditemukan diatas, penulis tertarik dalam membangun sebuah aplikasi menggunakan metode *K-Nearest Neighbor* (K-NN) untuk pemodelan klasifikasi penyakit *liver* dengan menggunakan data acuan test fungsi hati dari data sekunder yang diambil dari UCI Machine Learning Repository. Data yang digunakan yaitu ILDP (*Indian Liver Patient Dataset*).

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka perumusan masalah yang diangkat dalam penelitian ini adalah:

- a. Bagaimana cara mengatasi sulitnya mendiagnosa penyakit hati agar dapat didiagnosis meski tanpa ahli dalam bidang medis atau dokter?
- b. Bagaimana cara mengetahui  $k$  optimal menggunakan *elbow method* dalam metode *K-Nearest Neighbor* (K-NN)?
- c. Berapakah tingkat akurasi metode *K-Nearest Neighbor* (K-NN) dalam klasifikasi penyakit *liver*?



### 1.3 Batasan Masalah

Adapun Batasan masalah dari penelitian ini yaitu:

- a. Metode yang digunakan dalam aplikasi pada penelitian ini adalah *K-Nearest Neighbor* (K-NN).
- b. Aplikasi yang dibangun menggunakan bahasa pemrograman Python, dan bahasa markup HTML untuk membuat kerangka website.
- c. *Feature* yang digunakan adalah hasil tes fungsi hati: umur, jenis kelamin, total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase, total proteins, albumin dan globulin ratio.
- d. Data yang digunakan adalah data dari situs UCI Machine Learning Repository yaitu data ILDP (*Indian Liver Patient Dataset*), Terdapat 583 observasi dan 11 fitur yang digunakan.
- e. Aplikasi diimplementasikan berbasis *website*.

### 1.4 Tujuan Penelitian

Adapun Tujuan dari penelitian ini adalah:

- a. Membangun aplikasi yang dapat digunakan untuk membantu bidang medis dalam mendiagnosa penyakit hati menggunakan algoritma *K-Nearest Neighbor* (K-NN).
- b. Mengetahui k optimal menggunakan metode *elbow method* dari algoritma *K-Nearest Neighbor* (K-NN).
- c. Mengetahui tingkat akurasi dari algoritma *K-Nearest Neighbor* (K-NN) pada klasifikasi penyakit *liver*.

## 1.5 Manfaat Penelitian

Adapun Manfaat dari penelitian ini adalah:

- a. Meningkatkan pengetahuan penulis dalam mengklasifikasi risiko penyakit *Liver* menggunakan metode *K-Nearest Neighbor (K-NN)*.
- b. Studi ini diharap dapat membantu dan memfasilitasi profesional kesehatan dalam mengklasifikasi penyakit hati.

## 1.6 Metodologi Penelitian

Metode penelitian yang digunakan pada penelitian ini diantaranya studi literatur, pengumpulan data observasi, analisis kebutuhan, perancangan aplikasi, implementasi aplikasi, pengujian aplikasi dan penarikan kesimpulan.

## 1.7 Sistematika Penulisan

Sistematika penulisan yang penulis gunakan pada penelitian ini adalah sebagai berikut:

### **BAB I      Pendahuluan**

Bab ini berisi latar belakang, rumusan masalah, pembatasan masalah tujuan penelitian, metodologi dan sistematika penulisan skripsi

### **BAB II     Landasan Teori**

Pada bab ini membahas tentang teori-teori keilmuan yang mendasari masalah yang diteliti, yang terdiri dari teori-teori umum dan teori-teori khusus.

**BAB III Metode Penelitian**

Bab ini berisi studi literatur, pengumpulan data, analisis kebutuhan, perancangan sistem, implementasi sistem, dan pengujian sistem.

**BAB IV Analisis dan Perancangan Sistem**

Bab analisis dan perancangan aplikasi berisi pembahasan analisis, pemodelan dan perancangan aplikasi klasifikasi penyakit *liver* berbasis *website*. Pembahasan ditujukan untuk menguraikan kebutuhan-kebutuhan dalam pengembangan aplikasi dan pemodelan.

**BAB V Hasil dan Pengujian**

Bab ini berisi uraian tentang hasil, analisis, dan pengujian aplikasi klasifikasi penyakit *Liver* menggunakan metode *K-Nearest Neighbor* (K-NN). Pengujian pada penelitian ini menggunakan tingkat akurasi.

**BAB VI Kesimpulan dan Saran**

Bab ini berisi tentang kesimpulan dari penelitian yang telah dilakukan dan saran atau rekomendasi untuk perbaikan, pengembangan, kesempurnaan atau kelengkapan penelitian yang dilakukan.

## BAB VI KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dari penelitian yang telah dilakukan dan saran atau rekomendasi untuk perbaikan, pengembangan, kesempurnaan atau kelengkapan penelitian yang dilakukan.

### 6.1 Kesimpulan

Dari hasil penelitian diatas dapat diambil kesimpulan berdasarkan rumusan masalah sebagai berikut:

1. Kesimpulan dari penelitian ini adalah penulis dapat membangun aplikasi untuk mendiagnosa penyakit hati dan aplikasi yang dibangun dapat membantu bidang medis dalam mengklasifikasi penyakit hati dengan baik walaupun ada kekurangan dalam tingkat akurasi.
2. Kesimpulan dari evektifitas menggunakan *elbow method* mendapatkan hasil k optimal dari *error rate* pada k 4 dengan *best accuracy* 0.7894736842105263, dengan ini peneliti dengan akurat mendapatkan k dengan akurasi terbaik.
3. Dari hasil ujicoba akurasi dapat diambil kesimpulan bahwa data percobaan ketiga yang menjadi akurasi terbaik dengan akurasi *training* 80,5% *testing* 78,9% dan akurasi setelah *Tuning* pada data *training* 82,2% *testing* 77,1%, walaupun pada data percobaan keempat dengan menyeimbangkan data mengalami banyak peningkatan pada data *training* 98,8% dan setelah *tuning* menjadi 100% akan tetapi pada data *testing* mengalami penurunan yang sangat signifikan hingga 69,2% dan setelah *tuning* menjadi 62% bisa diambil kesimpulan bahwa data setelah penyeimbangan data menjadi *Overfitting* yang mana merupakan keadaan dimana model berusaha untuk mempelajari seluruh detail termasuk *noise* yang ada dalam data dan berusaha untuk mengikutsertakan semua data sehingga prediksi menjadi kurang bagus, skor AUC dan Akurasi dari setiap metode adalah metode *K-Nearest Neighbor* (K-NN) merupakan metode terbaik yang dapat dipakai dalam memprediksi penyakit liver .

## 6.2 Saran

Dari hasil penelitian diatas peneliti dapat memberikan beberapa saran untuk penelitian selanjutnya sebagai berikut:

1. Menambah sample data untuk *training* agar lebih banyak data yang dipelajari karena dalam metode *K-Nearest Neighbor* (K-NN) semakin banyak data semakin baik dalam mengkalsifikasi karena *K-Nearest Neighbor* (K-NN) mengklasifikasi dari k tetangga jadi semakin banyak k tetangga semakin baik juga metode *K-Nearest Neighbor* (K-NN).
2. Dalam mengatasi *overfitting* bisa dengan mengurangi *noise* atau *outlier*, pada penelitian ini dilakukan pembatasan menggunakan IQR dengan dilakukan penanganan menggunakan nilai mean meskipun tidak banyak yang berubah karena takut merusak hasil prediksi cara ini bukanlah solusi terbaik, penelitian selanjutnya bisa menggunakan perhitungan yang lebih kompleks dalam penanganan outlier.
3. Saran untuk penelitian yang menggunakan dataset ILDP (*Indian Liver Patient Dataset*) yang diambil dari website The UCI Learning Repository agar tidak melakukan *balancing* data karena hasil setelah menangani data *imbalance* mengakibatkan *overfitting* pada hasil prediksi.

## DAFTAR PUSTAKA

- [1] A. Noviriandini, P. Handayani, and Syahrani, "Prediksi Penyakit Liver Dengan Menggunakan Metode," *Pros. TAU SNAR-TEK Semin. Nas. Rekayasa dan Teknol.*, no. November, pp. 75–80, 2019.
- [2] Rudianto, "Penentuan Penyakit Peradangan Hati Dengan Menggunakan Neural Network Backpropagation," *Indones. J. Comput. Inf. Technol. Vol 1 No 1*, vol. 1, no. 1, pp. 27–33, 2016.
- [3] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016, doi: 10.15294/sji.v3i1.4610.
- [4] M. S. Mustafa and I. W. Simpen, "Implementasi Algoritma K-Nearest Neighbor ( KNN ) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba," *Pros. Semin. Ilm. Sist. Inf. Dan Teknol. Inf.*, vol. VIII, no. 1, pp. 1–10, 2019, [Online]. Available: <https://ejurnal.dipanegara.ac.id/index.php/sisiti/article/view/1-10>
- [5] W. Anggraini, "Deep Learning Untuk Deteksi Wajah Yang Berhijab Menggunakan Algoritma Convolutional Neural Network (Cnn) Dengan Tensorflow," *Sustain.*, vol. 4, no. 1, pp. 1–9, 2020.
- [6] M. M. Hidayat, "Data Mining Data mining," *Min. Massive Datasets*, vol. 2, no. January 2013, pp. 5–20, 2015, [Online]. Available: [https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book\\_part](https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part)
- [7] D. Turban, Efraim ; Aronson, Jay E ; Liang, Ting peng ; Prabantini, "Decision Support Systems And Intelligent Systems : ( Sistem Pendukung Keputusan Dan Sistem Cerdas ) / Efraim Turban," 2005.
- [8] T. Kenikan, K. Di, S. D. N. Citamiang, and K. Sukabumi, "Penerapan algoritma c.45 untuk klasifikasi tingkat kenikan kelas di sdn citamiang 2 kota sukabumi," 2020, [Online]. Available: [https://repository.bsi.ac.id/index.php/unduh/item/302730/19162343\\_Aris-Setiawan-new.pdf](https://repository.bsi.ac.id/index.php/unduh/item/302730/19162343_Aris-Setiawan-new.pdf)
- [9] E. T. L. Kusrini, "Algoritma data mining".
- [10] G. A. Marcoulides, *Discovering Knowledge in Data: an Introduction to Data Mining*, vol. 100, no. 472. 2005. doi: 10.1198/jasa.2005.s61.
- [11] D. Putra and A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 84–92, 2020.
- [12] E. Indrayuni, "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *J. Khatulistiwa Inform.*, vol. 7, no. 1, 2019, doi: 10.31294/jki.v7i1.5740.

- [13] J. Tandi, “Pola Penggunaan Obat Pada Pasien Penyakit Hati Yang Menjalani Rawat Inap Di Rumah Sakit Umum Daerah Undata Palu,” *Perspekt. J. Pengemb. Sumber Daya Insa.*, vol. 2, no. 2, pp. 218–223, 2017.
- [14] F. Gorunescu, *Data Mining: Concepts, models and techniques*. 2011.
- [15] R. K. Ngantung and M. A. I. Pakereng, “Model Pengembangan Sistem Informasi Akademik Berbasis User Centered Design Menerapkan Framework Flask Python,” *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1052, 2021, doi: 10.30865/mib.v5i3.3054.
- [16] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
- [17] R. Yuliana Sari, H. Oktavianto, and H. Wahyu Sulisty, “Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia,” *J. Smart Teknol.*, vol. 3, no. 2, pp. 2774–1702, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- [18] R. S. Darmawan and A. Suwastono, “Model Sistem Antrian Berbasis Arduino Uno R3 Dan Raspberry Pi Menggunakan Python,” *Univ. Gadjah Mada*, no. September, pp. 297–304, 2015.
- [19] A. F. K. Sibero, “Web Programming Power Pack,” 2015.
- [20] A. Géron, *Hands-on Machine Learning whith Scikit-Learing, Keras and Tensorfow*. 2019.

## BIOGRAFI PENULIS



Nama Lengkap : Ihya' Nashirudin Abrar  
Tempat, tanggal lahir : Pontianak, 3 Agustus 1999  
Jenis Kelamin : Laki-laki  
Agama : Islam  
Alamat : Jalan Adisucipto, Gang Permata Hijau, Perumahan  
Adigriya Karya No: E11.  
Nomor Telpon : 0895-4149-89092  
Email : 181220082@unmuhpnk.ac.id  
Kata Motifasi : "SUKSES !!!"