OPTIMASI KLASIFIKASI *K-NEAREST NEIGHBORS* DENGAN SELEKSI FITUR MENGGUNAKAN *ANALYSIS OF VARIANCE* UNTUK PREDIKSI PENYAKIT LIVER

SKRIPSI

Oleh:
SALISATUN NUR LAILI
NIM. 19650149

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024

OPTIMASI KLASIFIKASI *K-NEAREST NEIGHBORS* DENGAN SELEKSI FITUR MENGGUNAKAN *ANALYSIS OF VARIANCE* UNTUK PREDIKSI PENYAKIT LIVER

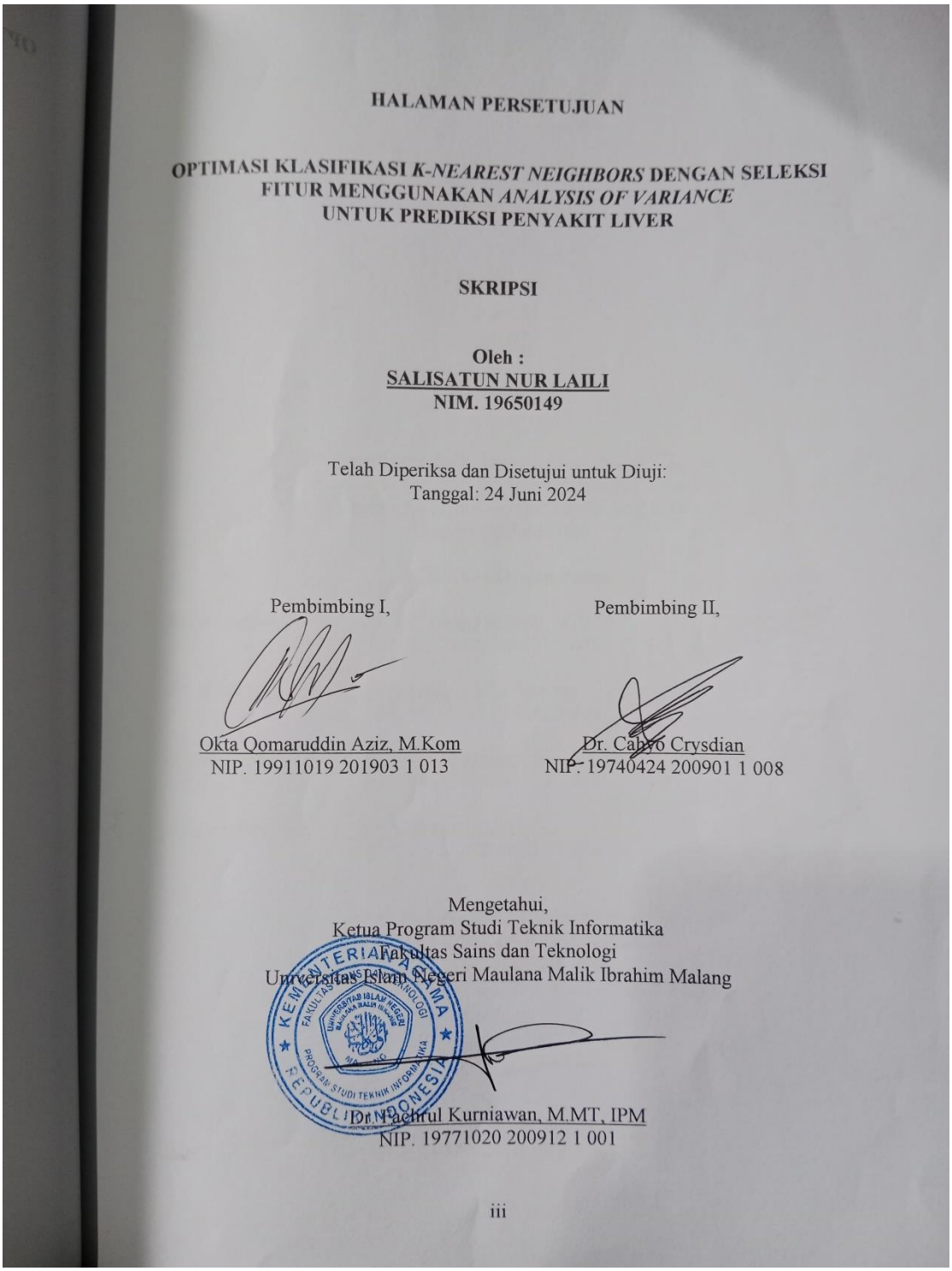
SKRIPSI

Diajukan Kepada :
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
SALISATUN NUR LAILI
NIM. 19650149

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024

HALAMAN PERSETUJUAN



HALAMAN PERSETUJUAN

OPTIMASI KLASIFIKASI *K-NEAREST NEIGHBORS* DENGAN SELEKSI FITUR MENGGUNAKAN *ANALYSIS OF VARIANCE* UNTUK PREDIKSI PENYAKIT LIVER

SKRIPSI

Oleh :
SALISATUN NUR LAILI
NIM. 19650149

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 24 Juni 2024

Pembimbing I,

Okta Qomaruddin Aziz, M.Kom
NIP. 19911019 201903 1 013

Pembimbing II,

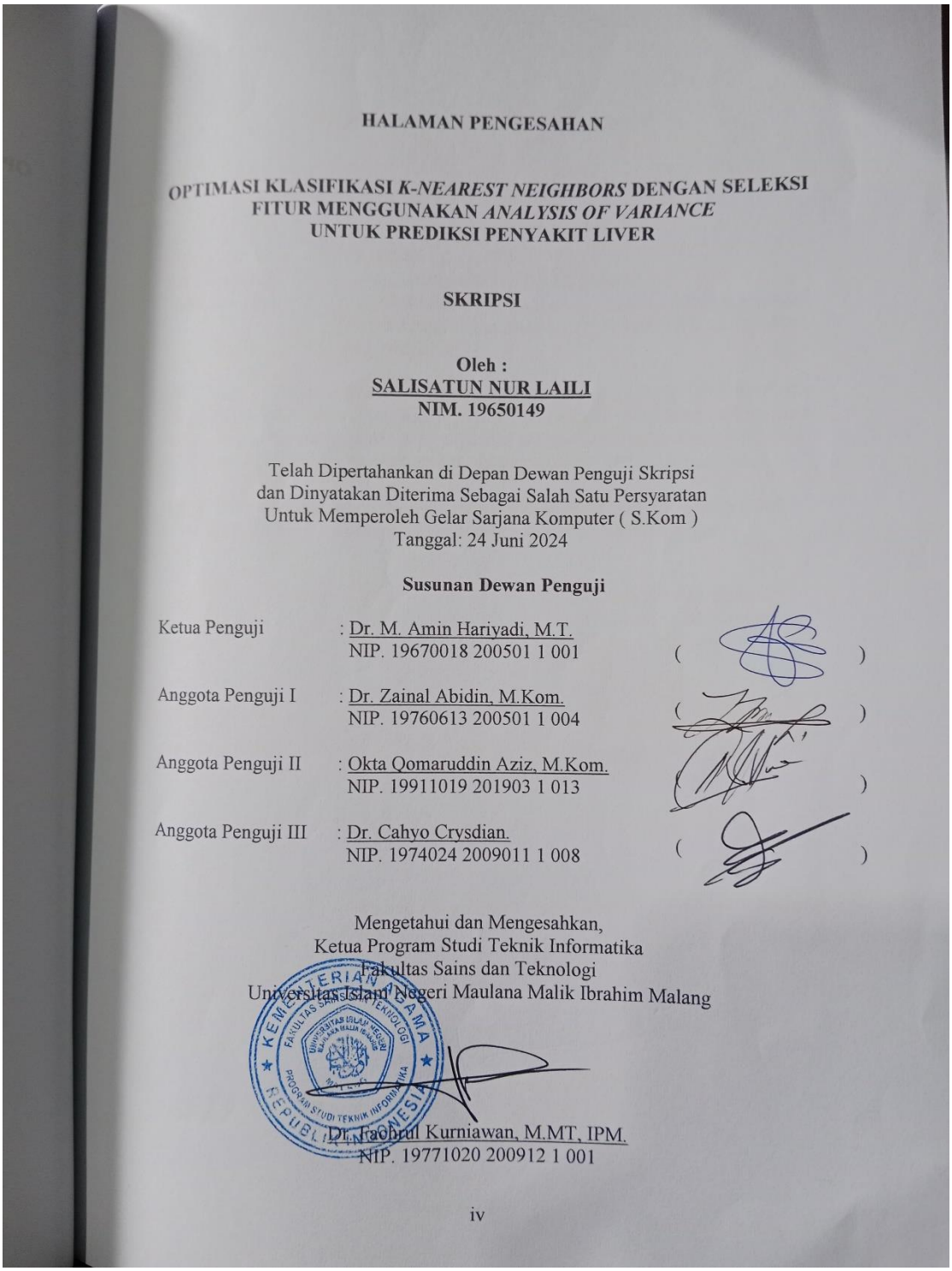
Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

iii

HALAMAN PENGESAHAN



HALAMAN PENGESAHAN

OPTIMASI KLASIFIKASI *K-NEAREST NEIGHBORS* DENGAN SELEKSI FITUR MENGGUNAKAN *ANALYSIS OF VARIANCE* UNTUK PREDIKSI PENYAKIT LIVER

SKRIPSI

Oleh :
SALISATUN NUR LAILI
NIM. 19650149

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 24 Juni 2024

Susunan Dewan Penguji

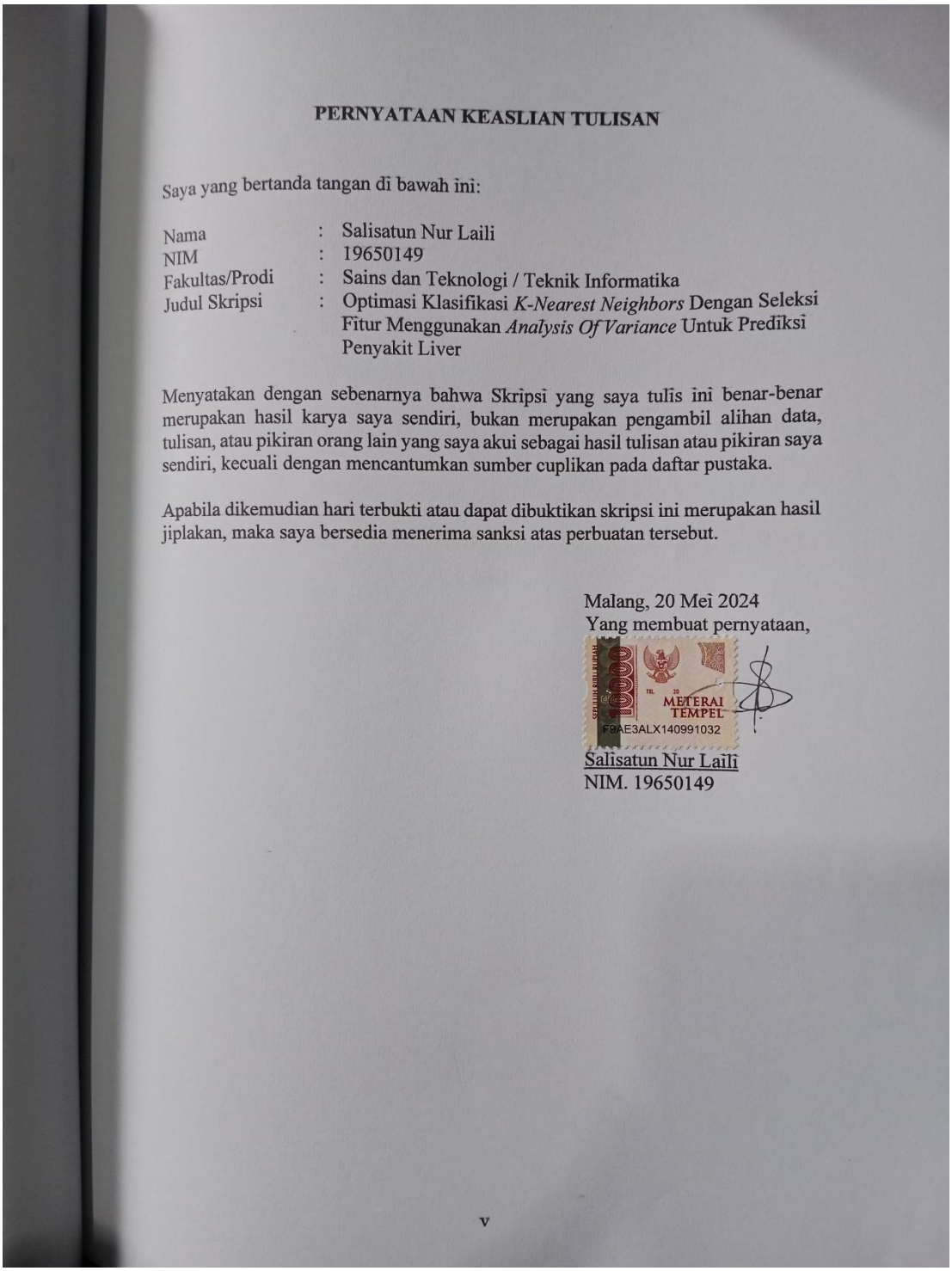
Ketua Penguji	: Dr. M. Amin Hariyadi, M.T. NIP. 19670018 200501 1 001	( )
Anggota Penguji I	: Dr. Zainal Abidin, M.Kom. NIP. 19760613 200501 1 004	( )
Anggota Penguji II	: Okta Qomaruddin Aziz, M.Kom. NIP. 19911019 201903 1 013	( )
Anggota Penguji III	: Dr. Cahyo Crysdian. NIP. 1974024 2009011 1 008	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM.
NIP. 19771020 200912 1 001

iv

PERNYATAAN KEASLIAN TULISAN



PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	:	Salisatun Nur Laili
NIM	:	19650149
Fakultas/Prodi	:	Sains dan Teknologi / Teknik Informatika
Judul Skripsi	:	Optimasi Klasifikasi *K-Nearest Neighbors* Dengan Seleksi Fitur Menggunakan *Analysis Of Variance* Untuk Prediksi Penyakit Liver

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Mei 2024
Yang membuat pernyataan,

METERAI TEMPEL
F9AE3ALX140991032

Salisatun Nur Laili
NIM. 19650149

v

HALAMAN MOTTO

“Fokus. Tidak Perlu Mendengarkan Pendapat Orang Lain”

HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamiin segala puji bagi Allah subhanahu wa ta’ala, serta shalawat dan salam kepada nabi agung Muhammad SAW. Penulis mempersembahkan skripsi ini untuk orang tua, kakak, serta teman yang telah memberikan banyak support, perhatian, kasih sayang serta arahan sehingga penulis dapat menyelesaikan skripsi ini.

KATA PENGANTAR

Assalamu'alaiakum Wr. Wb.

Bismillahirrahmaanirrahiim, puji syukur kehadirat Allah, dengan itu penulis mengucapkan terima kasih khususnya kepada:

1. Prof. Dr. M. Zainuddin, M.A. selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT. selaku Ketua Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Okta Qomaruddin Aziz, M.Kom. selaku Dosen Pembimbing I dan Dr. Cahyo Crysdian selaku Dosen Pembimbing II, yang membimbing penulis dengan penuh kesabaran dan keikhlasan dalam meluangkan waktunya untuk membimbing sehingga penulis dapat menyelesaikan tugas akhir ini.
5. Seluruh Dosen dan Staf di Program Studi Teknik Informatika, dengan ikhlas memberikan ilmu, bantuan, serta dorongan semangat selama perkuliahan.
6. Kedua orang tua penulis, Bapak Slamet, Ibu Munjayanah, dan kedua kakak saya yang telah memberikan banyak dukungan, doa serta bantuan sehingga penulis mampu menyelesaikan masa studi hingga mencapai gelar sarjana.
7. Keluarga besar Program Studi Teknik Informatika terutama Angkatan 2019 ALIEN “Alliance of Informatics Engineering” yang telah memberikan dukungan untuk saling menyelesaikan skripsi serta satu orang yang berarti bagi saya Deni Prabowo yang senantiasa meluangkan waktu untuk memberikan dorongan, motivasi, hingga materi, dalam masa perkuliahan hingga proses penyelesaian skripsi.
8. Diri saya sendiri, yang telah mampu melawan rasa malas dan bekerja keras untuk menyelesaikan skripsi ini. Terimakasih telah mampu kooperatif dan bertahan dalam menikmati proses pengerjaan skripsi.

Teriring doa, semoga amal baik yang telah diberikan kepada penulis mendapat balasan dari Allah SWT. Laporan Tugas Akhir ini telah ditulis dengan teliti dan sebaik-baiknya, namun saran dan kritikan yang membangun masih penulis harapkan untuk kemudian. Wassalamu'alaikum Wr. Wb.

Malang, 26 April 2023

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Laili, Salisatun Nur. 2024. **Optimasi Klasifikasi *K-Nearest Neighbors* Dengan Seleksi Fitur Menggunakan *Analysis of Variance* Untuk Prediksi Penyakit Liver**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Okta Qomaruddin Aziz, M.Kom (II) Dr. Cahyo Crysdian, MCS.

Kata Kunci: K-Nearest Neighbors, Analysis of Variance, Seleksi Fitur, Prediksi Penyakit Liver, Optimasi Model

Bertujuan untuk meningkatkan akurasi model klasifikasi K-Nearest Neighbors (K-NN) dalam memprediksi penyakit liver pada Indian Liver Patient Dataset (ILPD), penelitian ini fokus pada optimasi seleksi fitur. Teknik seleksi fitur yang digunakan adalah Analysis of Variance (ANOVA), yang berfungsi untuk memilih variabel yang paling signifikan dalam mempengaruhi hasil prediksi. Fitur yang terpilih kemudian melewati beberapa skenario uji coba dan digunakan sebagai input dalam model KNN, kemudian evaluasi model dilakukan dengan menggunakan confusion matrix. Penelitian menunjukkan dari seluruh skenario uji coba yang dilakukan, hasil rata-rata terbaik yang didapatkan ada pada rasio data 80:20 mengggunakan seleksi fitur yaitu diperoleh akurasi 70,17%, presisi 73,23%, *recall* 91,21%, dan *f1-score* 81,15%. Pada penelitian ini, hasil seleksi fitur yang lebih berpengaruh terhadap variabel target yaitu Total Bilirubin, Direct Bilirubin, dan Alkaline Phosphotase.

ABSTRACT

Laili, Salisatun Nur. 2024. **Optimizing K-Nearest Neighbors Classification with Feature Selection Using Analysis of Variance for Liver Disease Prediction**. Thesis. Informatics Engineering Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: (I) Okta Qomaruddin Aziz, M.Kom (II) Dr. Cahyo Crysdian, MCS.

Aiming to improve the accuracy of the K-Nearest Neighbors (K-NN) classification model in predicting liver disease on the Indian Liver Patient Dataset (ILPD), this research focuses on feature selection optimization. The feature selection technique used is Analysis of Variance (ANOVA), which serves to identify the most significant variables affecting the prediction outcomes. The selected features undergo several testing scenarios and are used as input for the KNN model. Model evaluation is conducted using a confusion matrix. The study shows that among all the testing scenarios, the best average results were obtained with an 80:20 data ratio using feature selection, achieving an accuracy of 70.17%, precision of 73.23%, recall of 91.21%, and an F1-score of 81.15%. In this study, the features that were most influential on the target variable, according to the feature selection results, were Total Bilirubin, Direct Bilirubin, and Alkaline Phosphotase.

Keywords: K-Nearest Neighbors, Analysis of Variance, Feature Selection, Liver Disease Prediction, Model Optimization

مستخلص البحث

ليلي، سالساتون نور. 2024. **تحسين تصنيف الجيران الأقرب (K–Nearest Neighbors) باستخدام اختيار الميزات بواسطة تحليل التباين للتنبؤ بأمراض الكبد**. البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج. المشرف الأول: أوكتا قمر الدين عزيز، الماجستير. المشرف الثاني: د. كاهيو كريسديان، ماجستير في علوم الكمبيوتر.

الكلمات الرئيسية: K–Nearest Neighbors, تحليل التباين، اختيار الميزات، التنبؤ بأمراض الكبد، تحسين النموذج.

يهدف هذا البحث إلى تحسين دقة نموذج تصنيف الجيران الأقرب (K-NN) في التنبؤ بأمراض الكبد باستخدام مجموعة بيانات مرضى الكبد الهندية(ILPD).ويركز على تحسين اختيار الميزات .التقنية المستخدمة لاختيار الميزات هي تحليل التباين (ANOVA). والتي تعمل على اختيار المتغيرات الأكثر أهمية في التأثير على نتائج التنبؤ .تم الميزات المختارة بعدة سيناريوهات تجريبية وتستخدم كمدخلات في نموذج ثم يتم تقييم النموذج باستخدام مصفوفة الارتباك أظهرت الدراسة أنه من بين جميع السيناريوهات التجريبية التي أجريت، كانت أفضل النتائج المتوسطة التي تم الحصول عليها عند نسبة بيانات80:20 باستخدام اختيار الميزات، حيث تم الحصول على دقة70.17%. ودقة(precision) 73.23%. واسترجاع. (recall)91.21%. ودرجة F1. بلغت. 81.15%. في هذه الدراسة، كانت نتائج اختيار الميزات الأكثر تأثيراً على المتغير الهدف هي إجمالي البيليروبين (Total Bilirubin) والبيليروبين المباشر(Direct Bilirubin) والفوسفاتاز القلوي (Alkaline Phosphatase).

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit liver atau hati merupakan salah satu penyakit yang serius dan dapat mengancam nyawa. Diagnosis dini dan pengobatan yang tepat sangat penting dalam upaya mengatasi penyakit ini. Penyakit liver adalah suatu gangguan pada setiap fungsi liver yang bertanggung jawab untuk fungsi-fungsi kritis dalam tubuh, dimana hilangnya fungsi-fungsi tersebut dapat menyebabkan kerusakan yang signifikan pada tubuh. Liver merupakan 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 (Pusporani et al., 2019).

Sivakrishnan & Pharm (2019) menjelaskan bahwa selama ini kebanyakan orang tidak menyadari bahwa apakah dirinya terkena penyakit liver atau tidak walaupun terdapat gejala. Banyak faktor-faktor penyebab rusaknya hati yang dimana hati sangat berpengaruh pada organ tubuh lainnya. Faktor tersebut disebabkan oleh virus, obat-obatan, alkohol, obesitas, diabetes atau serangan siatem kekebalan tubuh dan jika kondisi ini tidak segera ditangani dapat merusak saluran empedu pada hati sehingga menyebabkan kanker hati.

Secara garis besar klasifikasi penyakit menurut pandangan Islam, terdiri dari penyakit hati (rohani) dan penyakit jasmani. Diantara kedua penyakit itu ada pula yang disebut dengan penyakit alami, yaitu salah satu jenis penyakit jasmani yang tidak memerlukan tenaga medis dalam pengobatannya, seperti mengobati rasa

lapar, rasa haus, kedinginan, dan keletihan. Untuk mengatasi berbagai penyakit tersebut, Al-Qur’an menawarkan metode yang tepat, Allah berfirman, yang artinya:

وَنُنَزِّلُ مِنَ ٱلْقُرْءَانِ مَا هُوَ شِفَآءٌ وَرَحْمَةٌ لِّلْمُؤْمِنِينَ ۙ وَلَا يَزِيدُ ٱلظَّٰلِمِينَ إِلَّا خَسَارًا

“Dan kami turunkan sebagian dari Al-Qur’an suatu yang menjadi penawar dan rahmat bagi orang-orang yang beriman; dan Al-Qur’an itu tidaklah menambah manfaat kepada orang-orang zalim selain kerugian” (QS Al-Isra’/17:82)

Para ulama menafsirkan arti diturunkannya ayat tersebut bahwa Al-Qur'an suatu yang menjadi penawar dan rahmat bagi orang-orang yang beriman, dan Al-Qur'an itu tidaklah menambah kepada orang-orang yang zalim selain kerugian. Dalam Al-Qur'an sudah dituliskan penawar atau obat yang telah digolongkan atau diklasifikasikan sesuai penyakit. Dalam era perkembangan teknologi informasi, teknologi sangatlah berperan penting dalam dunia kesehatan. Salah satunya untuk analisis data dan pemodelan statistik yang menjadi salah satu alat yang efektif dalam mendukung diagnosis penyakit. Analisis data adalah proses menangkap informasi yang berguna dengan cara memeriksa, membersihkan, mengubah, dan memodelkan data menggunakan satu atau lebih teknik analisis data. Teknik ini dibagi menjadi dua jenis, yaitu berdasarkan matematika dan statistika dan berdasarkan machine learning dan artificial intelligence.

Machine learning merupakan mesin pembelajaran yang dikembangkan melalui kecerdasan buatan (Artificial Intelligence), sangat penting karena memungkinkan mesin untuk memperoleh kecerdasan mirip manusia tanpa pemrograman eksplisit. Algoritma machine learning sangat berperan penting untuk memudahkan seseorang dalam analisis sebuah data. Algoritma ini juga dapat

meningkatkan kemampuan dalam pengklasifikasian (Mohammed et al., 2016). Dalam hal ini machine learning memiliki kemampuan untuk memperoleh data yang ada dengan perintah ia sendiri. Proses kerja pada mesin pembelajran ini dimulai dengan memilih data berupa time series maupun data yang relevan, mesin ini memiliki tugas sebagai pengklasifikasi masalah, kelas target diketahui atau perlu prediksi (Herdiana & Geraldine, 2022).

Pada machine learning terdapat banyak metode diantaranya naïve bayes, decision tree, random forest, support vector machine, K-nearest neighbors (KNN), dan lainnya. Pada penelitian ini bertujuan untuk mengklasifikasi sebuah data ke dalam kelompok dengan metode KNN. K-Nearest Neighbor (K-NN) adalah salah satu metode dimana metode ini melakukan klasifikasi berdasarkan data training atau data pembelajaran dilihat dari jarak yang paling dekat dengan objek berdasarkan nilai k (Setianto et al., 2019). Dalam eksperimen penelitian ini digunakan model yang di implementasikan python untuk klasifikasi dataset ILPD (Indian Liver Patient Dataset) yang diambil dari UCI Machine Learning Repository.

Pada data terkait memerlukan pra-pemrosesan yang mencangkup pembersiham dan penghapusan variabel yang jelas tidak berpengaruh atau tidak relevan. Setelah data bersih termasuk indikator teknis diperoleh, data diproses lebih lanjut melalui penskalaan dan pengurangan dimensi (yaitu, pemilihan fitur, ekstraksi fitur, dan pembuatan fitur) untuk mendapatkan variabel yang relevan dan untuk menyaring variabel yang tidak relevan (Kumbure et al., 2022).

Gupta & Goel (2020) melakukan klasifikasi KNN menggunakan data penyakit diabetes untuk memprediksi sampel dataset baik pada kondisi diabetes maupun non-diabetes. Algoritma ini memberikan hasil terbaiknya dengan akurasi 87,01% dengan nilai skor f1 77,78%. Pengamatan menunjukkan bahwa nilai optimal K untuk pengklasifikasi KNN adalah 45 ketika fitur dipilih dengan metode pemilihan fitur ANOVA dan metode normalisasi min-max diterapkan untuk mengubah data.

Penelitian ini dilakukan untuk mengoptimasi klasifikasi KNN dengan seleksi fitur yang menggunakan metode uji ANOVA (Analysis of Variance) untuk mengetahui akurasi tertinggi dalam pengklasifikasi prediksi penyakit liver secara optimal. Sehingga penelitian ini dapat menjadikan acuan dalam mengevaluasi suatu penyakit dan mempermudah dalam mendiagnosis penyakit yang serupa. Sistem prediksi akan terus dikembangkan dan dievaluasi dengan tujuan untuk menciptakan akurasi diagnosis yang tepat dan efisien dalam diagnosis penyakit.

1.2 Pernyataan Masalah

Mengoptimasi algoritma klasifikasi K-Nearest Neighbors dengan seleksi fitur menggunakan Analysis of Variance untuk prediksi penyakit liver. Seleksi fitur ini meningkatkan interpretasi dan generalisasi model dengan fokus pada fitur-fitur yang paling berpengaruh terhadap variabel target.

1.3 Tujuan Penelitian

Mengukur tingkat akurasi klasifikasi penyakit liver menggunakan K-Nearest Neighbor dengan seleksi fitur menggunakan Analysis of Variance.

1.4 Batasan Masalah

Agar tercapainya maksud dan tujuan penelitian ini, maka terdapat batasan-batasan sebagai berikut :

1. Dataset yang digunakan dalam penelitian ini berupa dataset ILPD (Indian Liver Patient Dataset) yang diambil dari UCI Machine Learning.
2. Dataset yang digunnakan terdapat 11 atribut, yaiitu umur, gender, total bilirubin (TB), direct bilirubin (DB), Alkalin Phospotase (Alkphos), Alamine Aminotransferase (Sgpt), Aspartate Aminotransferase (Sgot), Total Protiens (TP), Albumin (ALB), Albumin and Globulin Ratio (A/G), dan dataset selector.

1.5 Manfaat Penelitian

Dari penelitian ini diharapkan bermanfaat bagi :

1. Bagi Tenaga Kesehatan dalam memprediksi penyakit liver lebih awal berdasarkan data yang tersedia.
2. Bagi pusat tenaga medis dalam membantu pasien dalam mendiagnosa penyakit liver berdasarkan gejala dan data yang tersedia.

BAB II

STUDI LITERATUR

2.1 Deteksi Penyakit Liver

Liver adalah organ penting dalam tubuh manusia yang memiliki banyak fungsi vital. Salah satu fungsi utamanya adalah sebagai pusat metabolisme, di mana liver membantu menguraikan zat-zat berbahaya dan memproses nutrisi dari makanan menjadi energi. Selain itu, liver juga berperan dalam produksi protein penting untuk pembekuan darah, serta menyimpan dan melepaskan glukosa sesuai kebutuhan tubuh. Kesehatan liver sangatlah penting untuk menjaga keseimbangan internal tubuh dan mencegah gangguan kesehatan serius.

Khairiah & Rismawan (2017) penyebab penyakit hati yang paling umum adalah konsumsi alkohol berlebihan, virus, kecanduan obat (khususnya dalam pembuluh darah), reaksi yang berlawanan dari berbagai macam obat (seperti analgesik, obat-obat anti peradangan, beberapa antibiotik, obat-obatan anti jamur dan penekan kekebalan. Jika penyakit hati tidak ditangani secara dini maka akan berkembang menjadi kanker hati dan dapat menimbulkan kematian.

Sivakrishnan & Pharm (2019) menjelaskan bahwa penyakit hati yang menyerang organ hati, terdapat beberapa jenisnya. Jenis-jenis penyakit hati ini yaitu:

a. Dataset Hepatitis A adalah peradangan (iritasi dan pembengkakan) hati dari virus hepatitis A. Virus hepatitis A kebanyakan ditemukan di tinja dan darah orang yang terinfeksi sekitar 15 - 45 hari sebelum gejala muncul dan selama

minggu pertama sakit. Gejala biasanya akan muncul 2 – 6 minggu setelah terpapar virus hepatitis A.

b. Hepatitis B adalah iritasi dan pembengkakan (radang) pada hati akibat infeksi virus hepatitis B (HBV). Hepatitis B menyebar melalui kontak dengan darah atau cairan tubuh (seperti air mani, cairan vagina, dan air liur) dari seseorang yang memiliki virus. Gejala hepatitis B mungkin tidak muncul hingga 6 bulan setelah infeksi.
c. Hepatitis C adalah penyakit virus yang menyebabkan pembengkakan (radang) hati. Infeksi hepatitis C disebabkan oleh virus hepatitis C (HCV). Hepatitis C menyebar melalui kontak dengan darah seseorang yang menderita hepatitis C.
d. Agen delta adalah sejenis virus yang disebut hepatitis D. Penyakit ini menyebabkan gejala hanya pada orang yang juga mengalami infeksi hepatitis B. Virus hepatitis D (HDV) hanya ditemukan pada orang yang membawa virus hepatitis B.
e. Hepatitis E adalah peradangan yang disebabkan oleh infeksi virus. Salah satu dari lima virus hepatitis manusia yang dikenal. A,B,C,D dan E. HEV sebagian besar ditularkan melalui kontaminasi feses pada air minum akibat sanitasi yang buruk, makanan yang terkontaminasi, seperti daging mentah atau setengah matang (misalnya: daging babi dan kerang) yang berasal dari hewan yang terinfeksi.
f. Penyakit hati berlemak adalah penumpukan lemak ekstra di sel-sel hati. Ini adalah tahap paling awal dari penyakit hati terkait alkohol. Biasanya tidak ada gejala. Jika gejala memang terjadi, mereka mungkin termasuk kelelahan,

kelemahan, dan penurunan berat badan. Hampir semua peminum berat memiliki penyakit hati berlemak. Namun, jika mereka berhenti minum, penyakit perlemakan hati biasanya akan hilang.

g. Hepatitis alkoholik menyebabkan hati membengkak dan menjadi rusak. Gejala mungkin termasuk kehilangan nafsu makan, mual, muntah, sakit perut, demam dan penyakit kuning. Hingga 35 persen peminum berat mengembangkan hepatitis alkoholik. Hepatitis alkoholik bisa ringan atau berat. Jika ringan, kerusakan hati dapat dibalik. Jika parah, dapat menyebabkan komplikasi serius termasuk gagal hati dan kematian.
h. Sirosis alkoholik adalah jaringan parut pada hati (jaringan parut yang keras menggantikan jaringan lunak yang sehat). Gejala sirosis mirip dengan hepatitis alkoholik. Antara 10 hingga 20 persen peminum berat terkena sirosis. Kerusakan akibat sirosis tidak dapat dipulihkan dan dapat menyebabkan gagal hati. Menghentikan konsumsi alkohol dapat membantu mencegah kerusakan lebih lanjut.

Indonesia merupakan negara dengan endemisitas tinggi Hepatitis B, terbesar kedua di negara South East Asian Region (SEAR) setelah Myanmar. Berdasarkan hasil Riset Kesehatan Dasar (Riskesdas), studi dan uji saring darah donor PMI maka diperkirakan di antara 100 orang Indonesia, 10 di antaranya telah terinfeksi Hepatitis B atau C. Sehingga saat ini diperkirakan terdapat 28 juta penduduk Indonesia yang terinfeksi Hepatitis B dan C, 14 juta di antaranya berpotensi untuk menjadi kronis, dan dari yang kronis tersebut 1,4 juta orang berpotensi untuk menderita kanker hati. Besaran masalah tersebut tentunya akan

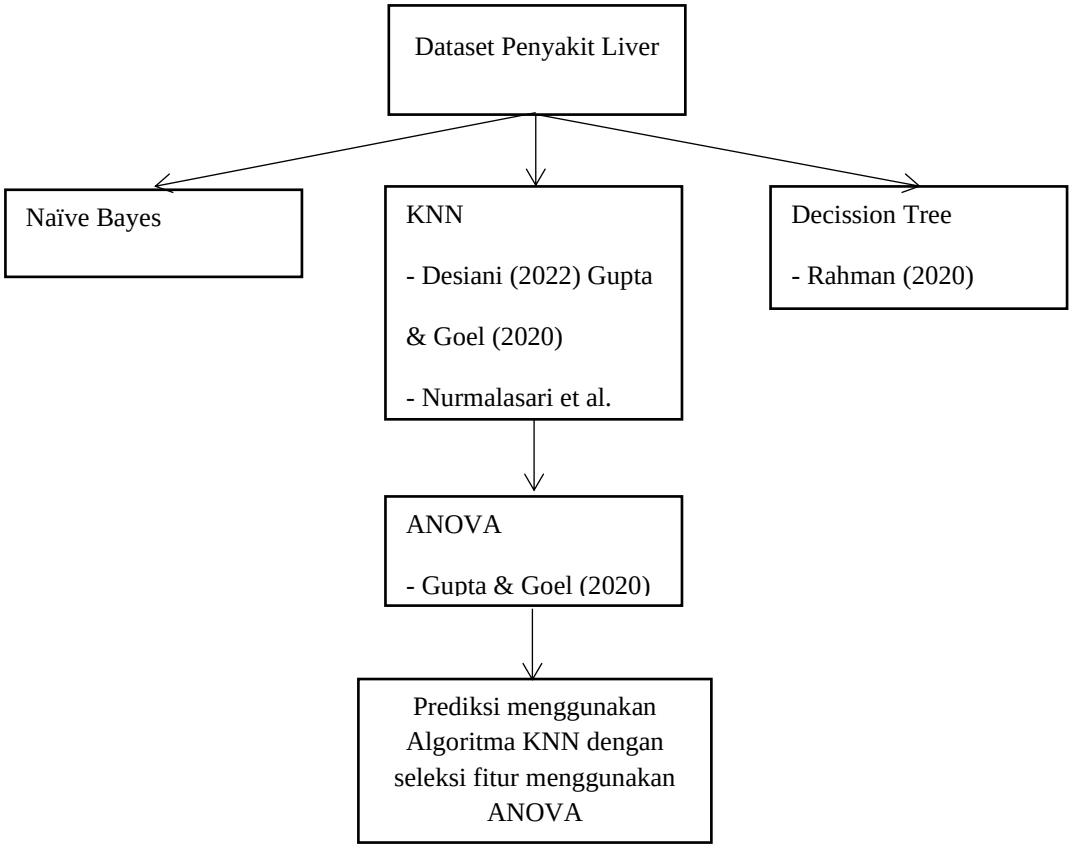
berdampak sangat besar terhadap masalah kesehatan masyarakat, produktifitas, umur harapan hidup, dan dampak sosial ekonomi lainnya (Rafsanjani et al., 2018).

Saragih (2022) melakukan diagnosa terhadap pasien di RSU Sari Mutiara Lubuk Pakam yang memiliki gejala-gejala penyakit hati menggunakan metode Naïve Bayes berbasis Web. Penelitian tersebut mendiagnosa dengan 18 data gejala penyakit gangguan hati yang dialami penderita penyakit hati yang menghasilkan 6 data jenis penyakit. Sistem ini dirancang karena di desa masih banyak masyarakat yang terkena penyakit dengan terbatasnya seorang dokter, jarak ke rumah sakit yang cukup jauh serta biaya konsultasi yang mahal, membuat masyarakat menunda untuk pergi ke rumah sakit. Hal ini tentu saja menjadi masalah bagi masyarakat di desa. Jadi, dibutuhkan sebuah teknologi yang dapat membantu masyarakat di desa, salah satunya teknologi yang berkembang saat ini adalah Sistem Pakar. Kemudian hasil dari sistem diagnosa tersebut menghasilkan nilai persentase tertinggi 0,005178%. Jadi dapat dikatakan bahwa pasien terkena penyakit gangguan hati jenis Hepatitis C.

Setiawati et al. (2019) melakukan klasifikasi penyakit liver pada dataset ILPD menggunakan Algoritma Decision Tree C4.5. Penelitian ini menunjukkan bahwa hanya 2 variabel (SPGT_AA dan Age) diantara 10 variabel pada dataset ILPD yang paling berpengaruh dalam penentuan klasifikasi penyakit liver. Penelitian ini juga menunjukkan hasil akurasi sebesar 72.67% dalam penentuan klasifikasi penyakit liver menggunakan Dataset ILPD.

Rahman (2020) telah melakukan pengujian pada data pasien penyakit liver menggunakan metode Decision Tree dan Naïve Bayes. Untuk mengetahui

komparasi algoritma yang paling baik dalam menentukan penyakit liver. Untuk mengukur kinerja kedua metode tersebut digunakan metode pengujian Split Validation dan Cross Validation. Dapat disimpulkan bahwa metode Decision Tree dalam klasifikasinya menghasilkan akurasi 70.29%. dan nilai AUC 0.757 yang termasuk dalam Fair Classification. Naïve Bayes menghasilkan akurasi 67.05% dan nilai AUC 0.714. Dengan demikian dapat disimpulkan metode yang memberikan pecahan untuk permasalahan dalam mengindentifikasi penyakit liver ialah Decision Tree.



Gambar 1.1 Theoritical Framework

2.2 *K - Nearest Neighbor* dalam Memprediksi Penyakit

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Algoritma k-NN menentukan nilai jarak pada pengujian data uji dengan data latih berdasarkan nilai terkecil dari nilai ketetanggaan terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan training samples (Swantika et al., 2023).

Nurmalasari et al. (2021) menunjukkan bahwa metode KNN memiliki performa lebih baik pada nilai confusion matrix dibandingkan metode naïve bayes dengan menggunakan dataset penyakit diabetes terdiri dari 521 data dengan 17 atribut. Proses preprocessing dilakukan menggunakan Replace Missing Value untuk menghilangkan data yang berulang. Setelah melakukan tahap preprocessing selanjutnya melakukan split data yaitu membagi data dengan rasio 80% untuk data training dan 20 % sebagai data testing yang akan digunakan untuk proses klasifikasi. Hasil dari perhitungan perfoma untuk algoritma Naïve Bayes dan KNN dengan menggunakan validasi data yaitu 10 k-fold yang telah dilakukan pada saat preprocessing dan perhitungan perfoma dengan confussion matrix.

Kustiyahningsih et al. (2021) melakukan pengklasifikasian penyakit sapi secara cepat dan akurat untuk membantu peternak sapi dalam mempercepat deteksi dan penanganan penyakit sapi. Penelitian ini menggunakan metode klasifikasi K-Nearest Neighbor (KNN) dengan seleksi fitur F-Score. Dataset yang digunakan adalah data penyakit sapi di Madura dengan total 350 data yang terdiri dari 21 fitur dan 7 kelas. Data dipecah menggunakan K-fold Cross Validation menggunakan k

= 5. Berdasarkan hasil pengujian diperoleh akurasi terbaik dengan jumlah fitur = 18 dan KNN (k = 3) yang menghasilkan akurasi sebesar 94.28571, recall sebesar 0,942857 dan presisi sebesar 0,942857.

Desiani (2022) melakukan penelitian menggunakan dua algoritma untuk melakukan klasifikasi dataset penyakit hati yaitu algoritma Naïve Bayes dan algoritma K-NN tujuannya untuk membandingkan kemudian menyimpulkan algoritma terbaik yang dapat digunakan dalam melakukan klasifikasi penyakit hati. Pada penerapan algoritma Naïve Bayes memperoleh akurasi sebesar 85%. Nilai presisi untuk terkena penyakit sebesar 88% dan untuk tidak terkena penyakit sebesar 83%. Nilai recall untuk terkena penyakit sebesar 81% dan untuk tidak terkena penyakit sebesar 90%. Pada penerapan algoritma K-Nearest Neigbhor memperoleh akurasi sebesar 100%. Nilai presisi untuk terkena penyakit sebesar 100% dan untuk tidak terkena penyakit sebesar 100%. Nilai recall untuk terkena penyakit sebesar 100% dan untuk tidak terkena penyakit sebesar 100%. Bahwa nilai presisi dan recall yang dihasilkan algoritma Naïve Bayes dan K-Nearest Neighbor memiliki perbedaan yang cukup stabil dan lumayan jauh sebagai prediksi penyakit hati karena K-NN mencapai 100% sedangkan Naïve Bayes 85% keatas yang artinya memiliki perbedaan atau selisih sekitar 15%. Algoritma K-NN memberikan nilai yang lebih tinggi dibandingkan dengan algoritma Naïve Bayes. Maka algoritma terbaik yang dapat digunakan adalah K-NN.

Wijaya et al. (2022) melakukan klasifikasi terhadap data kanker payudara, dimana data tersebut terdapat adalah data darah pengidap kanker payudara. Metode yang digunakan pada klasifikasi ini adalah K-Nearest Neighbor(KNN) dan

Gaussian Naive Bayes(GNB). Pengujian akurasi pada penelitian ini menggunakan Cross Validation dan evaluasi data ujia dengan Confusion Matrix. Dari penelitian ini didapatkan hasil pada 116 data darah kanker payudara, metode KNN menghasilkan akurasi 86,9% lebik baik dari pada GNB,dan untuk presisi dan recall, metode KNN menghasilkan presisi sebesar 87,3%, dan recallsebesar 86,7%, pengujian pada metode KNN menggunakan nilai K=4.

Oktavyani et al. (2023) melakukan perbandingan klasifikasi antara metode Naïve bayes, KNN dan Decision Tree. Dimana data dari penelitian adalah dataset laporan kesehatan penyakit Stroke berasal dari kaggle.com, pada penelitian ini akan diukur confusion matrix, precision, recall, accuracy, hingga f-measure kemudian juga dihitung root mean square error dari tiap-tiap metode, dari perhitungan tersebut metode KNN mendapatkan accuracy tertinggi hingga 95,20% sehingga dapat disimpulkan metode klasifikasi KNN lebih baik dari metode Naïve bayes maupun Decision Tree. Pada metode KNN juga memiliki nilai MSE yang terkecil.

Metode k-Nearest Neighbor (k-NN) adalah teknik klasifikasi pembelajaran mesin nonparametrik yaitu, tidak ada asumsi untuk distribusi data yang mendasarinya. Dengan kata lain, struktur model ditentukan dari dataset dengan menyimpan semua kasus yang tersedia dan memprediksi kasus baru berdasarkan ukuran kesamaan (Sarker et al., 2018).

2.3 *Analysis of Variance* dalam Memprediksi Penyakit

Analysis of Variance (ANOVA) adalah pengujian metode statistik yang digunakan untuk menguji perbedaan yang signifikan antara rata-rata kelompok data. Metode ANOVA satu arah digunakan untuk memilih fitur-fitur penting secara

statistic berdasarkan nilai F dan nilai p value. Pada nilai F, fitur hanya dipilih dengan hasil skor tertinggi untuk digunakan pengklasifikasi. Selanjutnya pada nilai p-value ditentukan berdasarkan fitur yang relevan dan membandingkannya dengan tingkat signifikansi. Jika nilai p kurang dari signifikansi maka fitur disimpan untuk proses uji lanjut dan begitu sebaliknya (Alassaf et al., 2022).

Gupta & Goel (2020) melakukan klasifikasi KNN menggunakan data penyakit diabetes untuk memprediksi sampel dataset baik pada kondisi diabetes maupun non-diabetes. Algoritma ini memberikan hasil terbaiknya dengan akurasi 87,01% dengan nilai skor f1 77,78%. Pengamatan menunjukkan bahwa nilai optimal K untuk pengklasifikasi KNN adalah 45 ketika fitur dipilih dengan metode pemilihan fitur ANOVA dan metode normalisasi min-max diterapkan untuk mengubah data.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Dataset

Pada penelitian ini, data yang digunakan diambil dari situs resmi UCI Machine Learning Repository "https://archive.ics.uci.edu/dataset/225/". Dalam dataset ILPD (Indian Liver Patient Dataset) terdapat data yang berjumlah 583 data pasien dengan 11 variabel diantaranya 1 variabel dependent dan 10 variabel independent. Dibawah ini merupakan dataset liver untuk variabel data ILPD terdapat kelas yaitu kelas 1 liver dan kelas 2 non liver yang akan digunakan sebagai sampel data untuk proses selanjutnya, berikut dapat dilihat pada Tabel 3.1.

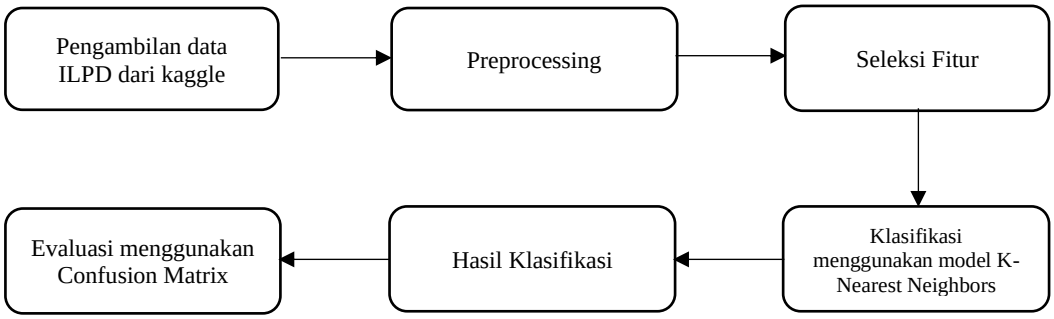
Tabel 3. 1 Dataset ILPD (Indian Liver Patient Dataset)

No	Age	Gender	TB	DB	AP	Alm	Asp	TP	Alb	AGR	Class
1	65	Female	0,7	0,1	187	16	18	6,8	3,3	0,9	1
2	62	Male	10,9	5,5	699	64	100	7,5	3,2	0,74	1
3	62	Male	7,3	4,1	490	60	68	7	3,3	0,89	1
4	58	Male	1	0,4	182	14	20	6,8	3,4	1	1
5	72	Male	3,9	2	195	27	59	7,3	2,4	0,4	1
6	46	Male	1,8	0,7	208	19	14	7,6	4,4	1,3	1
7	26	Female	0,9	0,2	154	16	12	7	3,5	1	1
8	29	Female	0,9	0,3	202	14	11	6,7	3,6	1,1	1
9	17	Male	0,9	0,3	202	22	19	7,4	4,1	1,2	2
10	55	Male	0,7	0,2	290	53	58	6,8	3,4	1	1
11	57	Male	0,6	0,1	210	51	59	5,9	2,7	0,8	1
12	72	Male	2,7	1,3	260	31	56	7,4	3	0,6	1
13	64	Male	0,9	0,3	310	61	58	7	3,4	0,9	2
14	74	Female	1,1	0,4	214	22	30	8,1	4,1	1	1
15	61	Male	0,7	0,2	145	53	41	5,8	2,7	0,87	1

Data ILPD (Indian Liver Patient Dataset) berisi 416 catatan pasien terkena liver dan 167 catatan pasien non liver. Kumpulan dataset ini dikumpulkan dari

wilayah timur laut Andhra Pradesh, India. Pada kumpulan dataset ILPD ini diketahui pasien penderita penyakit liver jenis hepatitis, disebabkan karena konsumsi alkohol yang berlebihan, menghirup gas berbahaya, konsumsi makanan yang terkontaminasi, dan obat-obatan. Data ILPD ini berisi 441 catatan pasien pria dan 142 catatan pasien wanita. Pada dataset ini memiliki 10 variabel independent diantaranya Age, Gender, Total Bilirubin (TB), Direct Bilirubin (DB), Alkaline Phosphotase(AP), Alamine Aminotransferase (Alm), Aspartate Aminotransferase (Asp), Total Protiens (TP), Albumin (ALB), dan Albumin Globulin Ratio (AGR).

Tiap variabel memiliki nilai kepentingan yang berbeda-beda sesuai dengan kebutuhan. Setiap variabel akan di uji seberapa besar pengaruh variabel terhadap kelas dan model yang digunakan. Dapat dilihat pada Gambar 3.1 terdapat desain penelitian.



Gambar 3. 1 Desain Penelitian

Pada Gambar 3.1 desain penelitian dimulai dengan pengenalan data ILPD yang mencakup deskripsi dataset dan sumbernya, dilanjutkan dengan memuat dan mengeksplorasi data untuk memahami struktur dan menampilkan informasi dasar tentang data seperti jumlah baris, kolom, tipe data. Selanjutnya, dilakukan prapemrosesan data seperti konversi kategori ke numerik. Kemudian seleksi fitur dilakukan menggunakan ANOVA untuk menentukan fitur yang signifikan.

Implementasi model K-Nearest Neighbors (KNN) melibatkan pembagian data menjadi data latih dan data uji. Evaluasi model dilakukan dengan mengukur akurasi, confusion matrix, dan classification report. Hasil analisis divisualisasikan melalui diagram batang.

Penelitian yang menggunakan ILPD Dataset sudah banyak dilakukan, pada tahun 2011 penelitian tentang membandingkan algoritma klasifikasi dengan menggunakan ILPD Dataset, dari penelitian tersebut menunjukkan bahwa algoritma KNN, Backpropagation, dan SVM memberikan hasil yang terbaik dari pada algoritma Naïve Bayes dan Decision Tree C4.5 (Ramana et al., 2011).

Ramana et al. (2012) pada penelitiannya dengan membandingkan dataset pasien penyakit liver dari ILPD dan UCLA. Dari dua dataset tersebut terdapat atribut yang sama yaitu Alkaline Phosphotase, Alamine Aminotransferase (SGPT), Aspartate Aminotransferase (SGOT). Dari 3 atribut tersebut diolah menggunakan ANOVA dan MANOVA, menghasilkan bahwa tidak ada perbedaan yang signifikan antara pasien non-hati dari dataset UCI dan India dengan pasien non-hati dari AS dan India.

Setiawati et al. (2019) penelitiannya menunjukkan bahwa hanya 2 variabel yaitu Alamine Aminotransferase dan Age diantara 10 variabel pada dataset ILPD yang paling berpengaruh dalam penentuan klasifikasi penyakit liver. Penelitian ini juga menunjukkan hasil akurasi sebesar 72.67% dalam penentuan klasifikasi penyakit liver menggunakan Dataset ILPD.

Pada penelitian ini dataset akan digunakan untuk melatih model klasifikasi menggunakan K-Nearest Neighbors (KNN) dengan seleksi fitur menggunakan

Analysis of Variance. Dataset pada penyakit liver ini memiliki deskripsi dari masing-masing atribut. Berikut dijelaskan pada Tabel 3.2:

Tabel 3. 2 Deskripsi Atribut

No	Atribut	Tipe Data	Deskripsi
1.	*Age*	*Numeric*	Umur pasien terkena penyakit liver
2.	Total Bilirubin (TB)	*Numeric*	Jumlah bilirubin langsung dan tidak langsung
3.	*Direct Bilirubin* (DB)	*Numeric*	Larutan air yang dikeluarkan dari hati menuju urine
4.	*Alkaline Phosphotase* (AP)	*Numeric*	Enzim dalam aliran darah yang bertugas membantu memecah protein dalam tubuh
5.	*Alanine Aminotransferase* (Alm)	*Numeric*	Mengukur jumlah enzim dalam darah
6.	*Aspartate Aminotransferase* (Asp)	*Numeric*	Protein yang dihasilkan oleh sel-sel hati
7.	Total *Protiens* (TP)	*Numeric*	Mengukur jumlah total dua jenis protein pada tubuh
8.	Albumin (ALB)	*Numeric*	Protein yang diproduksi oleh organ hati
9.	*Albumin Globulin Ratio* (AGR)	*Numeric*	Rasio yang membandingkan jumlah albumin dan globulin dalam darah
10.	*Class*	*Category*	Menderita liver/ tidak menderita liver

Atribut yang terdapat pada dataset ILPD ini memiliki fungsi dan pengaruh penting terhadap resiko terkena penyakit liver. Pada laman web "https://www.hepatitis.va.gov/hcv/patient/index.asp" situs resmi U.S. Department of Veterans Affair, dijelaskan pengaruh setiap tes fungsi hati yang menyebabkan liver. Pertama terdapat fungsi variabel Age (umur) yang dapat mempengaruhi kinerja organ tubuh manusia, semakin bertambahnya usia maka fungsi organ tubuh semakin menurun secara alamiah. Kemudian pada variabel Gender (jenis kelamin) dapat mempengaruhi hasil dari bilirubin. Dimana pria cenderung memiliki kadar bilirubin sedikit lebih tinggi dibandingkan wanita. Orang berkulit hitam cenderung memiliki kadar bilirubin lebih rendah dibandingkan orang dari ras lain.

Selanjutnya variabel Bilirubin adalah zat kekuningan yang dihasilkan oleh pemecahan (penghancuran) hemoglobin, komponen utama sel darah merah. Seiring bertambahnya usia sel darah merah, mereka dipecah secara alami di dalam tubuh. Bilirubin dilepaskan dari sel darah merah yang hancur dan diteruskan ke hati. Hati melepaskan bilirubin dalam cairan yang disebut empedu. Jika hati tidak berfungsi dengan baik, bilirubin tidak akan dikeluarkan dengan baik. Oleh karena itu, jika kadar bilirubin lebih tinggi dari yang diharapkan, hal ini mungkin berarti hati tidak berfungsi dengan baik.

Pada total bilirubin, menghasilkan dua jenis kadar bilirubin yaitu direct (langsung) dan indirect (tidak langsung). Direct bilirubin adalah bilirubin yang dibuat dari pemecahan sel darah merah kemudian berjalan dalam darah ke hati. Sedangkan indirect bilirubin adalah bilirubin yang telah mencapai hati dan mengalami perubahan kimia sehingga berpindah ke usus sebelum dikeluarkan melalui tinja. Untuk orang dewasa di atas 18 tahun, total bilirubin normalnya bisa mencapai 1,2 miligram per desiliter (mg/dl) darah. Sedangkan usia di bawah 18 tahun, kadar normalnya adalah 1 mg/dl. Hasil normal untuk direct bilirubin (langsung) harus kurang dari 0,3 mg/dl.

Selanjutnya Alkaline Fosfatase (sering disingkat menjadi alk phos) adalah enzim yang dibuat di sel hati dan saluran empedu. Tingkat alk phos adalah tes umum yang biasanya disertakan ketika tes hati dilakukan secara berkelompok. Tingkat alk phos yang tinggi tidak mencerminkan kerusakan atau peradangan hati. Tingkat alk phos yang tinggi terjadi ketika ada penyumbatan aliran di saluran empedu atau penumpukan tekanan di hati, sering kali disebabkan oleh batu empedu

atau jaringan parut di saluran empedu. Tidak sedikit pasien yang terkena liver memiliki kadar alkphos yang normal.

Alanine Aminotransferase atau ALT adalah salah satu dari dua enzim hati yang biasanya dikenal sebagai serum glutamat-piruvat transaminase, atau SGPT yang mana protein hanya dibuat oleh sel hati. Ketika sel-sel hati rusak, ALT bocor ke dalam aliran darah dan tingkat ALT dalam darah meningkat. Kadar ALT yang tinggi seringkali berarti ada kerusakan hati, namun pasien menderita penyakit hati dan sirosis yang sangat parah, tidak sedikit yang masih memiliki tingkat ALT yang normal.

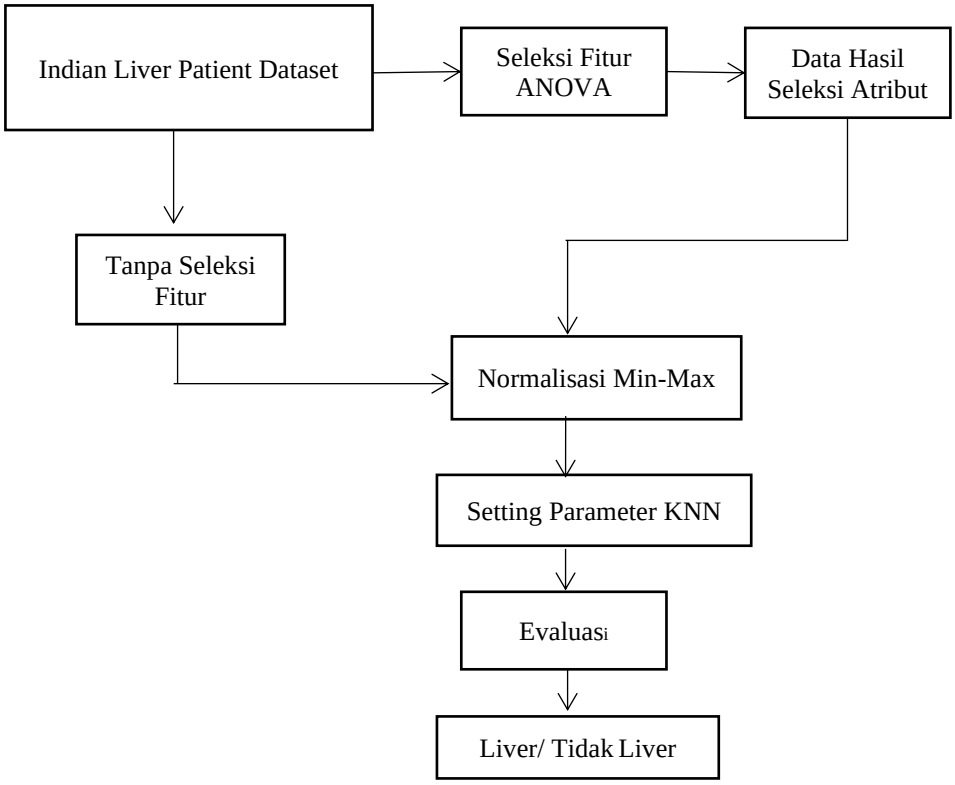
Aspartate Aminotransferase atau AST, adalah salah satu dari dua enzim hati. AST dikenal sebagai serum glutamic-oxaloacetic transaminase, atau SGOT. AST adalah protein yang dibuat oleh sel hati. Ketika sel-sel hati rusak, AST bocor ke aliran darah dan tingkat AST dalam darah menjadi meningkat. AST berbeda dengan ALT karena AST ditemukan di bagian tubuh selain hati yaitu termasuk jantung, ginjal, otot, dan otak. Ketika sel-sel di salah satu bagian tubuh rusak, AST dapat meningkat. Kadar AST tidak begitu membantu dibandingkan kadar ALT untuk memeriksa organ hati. Kadar AST yang tinggi dengan ALT yang normal mungkin berarti bahwa AST tersebut berasal dari bagian tubuh lain.

Total Protein adalah ukuran sejumlah protein yang berbeda dalam darah. Total protein dapat dibagi menjadi fraksi albumin dan globulin. Rendahnya kadar protein total dalam darah dapat terjadi karena adanya gangguan fungsi hati. Albumin adalah protein yang dibuat oleh hati, untuk mencegah cairan bocor keluar dari pembuluh darah ke jaringan. Kadar albumin yang rendah pada penderita

penyakit hati dapat menjadi tanda sirosis (penyakit hati stadium lanjut). Albumin Globulin Ratio atau rasio A/G dalah tes protein serum total mengukur semua protein dalam darah serta dapat memeriksa jumlah albumin dan dibandingkan dengan globulin. Orang yang sehat mempunyai lebih banyak albumin dibandingkan globulin.

3.2 Desain Sistem

Pada penelitian ini, akan dikembangkan sebuah sistem dengan beberapa tahapan. Sistem yang dibuat akan dikembangkan berdasarkan desain sistem pada Gambar 3.2, menggunakan bahasa pemrograman Phyton dan menggunakan dataset Indian liver sebagai input. Data akan diproses melalui beberapa tahapan seperti tahap split data, feature selection, dan normalization. Gambar 3.2 merupakan alur desain sistem pada penelitian Optimasi Klasifikasi K-Nearest Neighbors dengan Seleksi Fitur menggunakan Analysis of Variance untuk Prediksi Penyakit Liver:

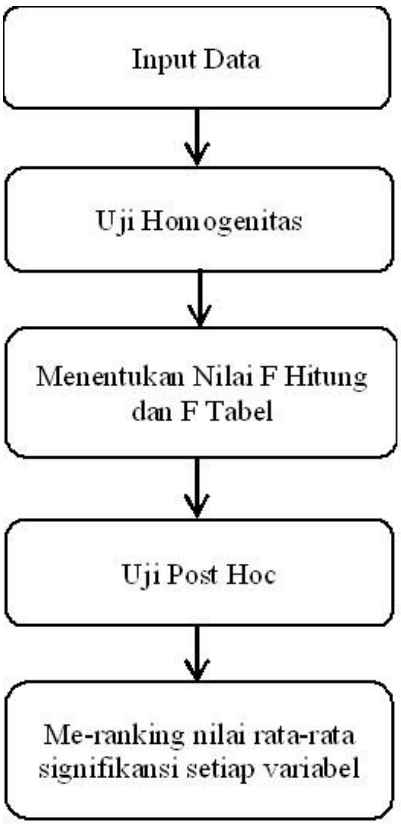


Gambar 3. 2 Desain Sistem

Pada Gambar 3.2 desain sistem dibuat dengan model klasifikasi menggunakan metode K-Nearest Neighbor. Pada alur tersebut, data yang sudah diambil akan melalui tahap proses seleksi atribut menggunakan *Analysis of Variance*, kemudian hasil yang didapat melewati proses normalisasi data menggunakan *min max*. Proses selanjutnya membangun parameter dengan metode K-Nearest Neighbor, kemudian melalui proses evaluasi yang diuji dengan menggunakan *confusion matrix* dan menghasilkan performa untuk memperoleh nilai akurasi, presisi, *recall*, dan *f1-score*. Setelah mendapatkan hasil keseluruhan, model yang digunakan pada penelitian akan di uji menggunakan *K-Fold Cross Validation* untuk mengukur kinerja model dengan lebih akurat.

3.3 *Analysis of Variance*

One-Way Analysis of Variance (ANOVA ragam satu arah) biasanya digunakan untuk menguji rata-rata/pengaruh perlakuan dari suatu percobaan yang menggunakan satu faktor, dimana satu faktor tersebut memiliki tiga atau lebih kelompok. Disebut satu arah karena peneliti dalam penelitiannya hanya berkepentingan dengan satu faktor saja atau juga dapat dikatakan *One-Way Anova* (analisis ragam satu arah) mengelompok data berdasarkan satu kriteria saja, misalnya ingin mengetahui ada perbedaan yang nyata antara rata-rata hitung sembilan kelompok data dan uji statistika yang digunakan uji F. Berikut alur *One way* ANOVA pada Gambar 3.3.



Gambar 3. 3 Alur *One Way* ANOVA

Dalam sample penelitian ini populasinya diketahui jumlahnya yaitu diasumsikan jumlah sampel 15 data. Kemudian membuat hipotesis dalam uraian kalimat H0 = tidak ada perbedaan nilai rata-rata performance yang siginifikan antara variabel data H1 = ada perbedaan nilai rata-rata performance yang siginifikan antara variabel data. Selanjutnya membuat hipotesis model statistic H0 : X1 = X2 = X3 dan menentukan taraf signifikan.

Sanggyu (2023) menurutnya tingkat signifikansi berarti kemungkinan menolak hipotesis nol karena penilaian yang salah meskipun hipotesis nol sebenarnya benar dalam pengujian hipotesis dan sesuai dengan kesalahan tipe 1. Saat melakukan uji hipotesis statistik, tingkat signifikansi dapat ditetapkan ke 1%, 5%, atau 10%. Jika tingkat signifikansi ditetapkan 5%, berarti hipotesis nol ditolak 5 kali dari 100 meskipun benar. Dengan kata lain, Anda 95% yakin bahwa Anda akan menguji hipotesis yang benar. Hal ini menyimpulkan bahwa ditolak di bawah tingkat signifikansi 5% sebagai hasil pengujian hipotesis statistik.

Pada penelitian ini menggunakan taraf signifikansinya yaitu 5% karena semakin kecil angka taraf signifikansi maka semakin besar tingkat kepercayaan pada penelitian. Dalam kaidah pengujian, jika Fhitung < Ftabel maka terima H0 dan jika Fhitung > Ftabel maka tolak H0. Dapat dilihat hasil uji homogenitas pada Gambar 3.4.

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Nilai	Based on Mean	11,500	8	126	,000
	Based on Median	4,897	8	126	,000
	Based on Median and with adjusted df	4,897	8	14,807	,004
	Based on trimmed mean	8,846	8	126	,000

ANOVA

Nilai

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	850321,541	8	106290,193	41,525	,000
Within Groups	322517,349	126	2559,662		
Total	1172838,891	134			

Gambar 3. 4 Hasil Uji Homogenitas

Berikut langkah-langkah dalam menghitung nilai F_{hitung} :

a. Membuat table penolong $\sum X_n$ dan $\sum(X_n)^2$, dilihat pada Tabel 3.3 dan Tabel 3.4.

Tabel 3. 3 Tabel Perhitungan untuk Mencari Nilai $\sum X_n$

Data	(X_1)	(X_2)	(X_3)	(X_4)	(X_5)	(X_6)	(X_7)	(X_8)	(X_9)
1	65	0,7	0,1	187	16	18	6,8	3,3	0,9
2	62	10,9	5,5	699	64	100	7,5	3,2	0,74
3	62	7,3	4,1	490	60	68	7	3,3	0,89
4	58	1	0,4	182	14	20	6,8	3,4	1
5	72	3,9	2	195	27	59	7,3	2,4	0,4
6	46	1,8	0,7	208	19	14	7,6	4,4	1,3
7	26	0,9	0,2	154	16	12	7	3,5	1
8	29	0,9	0,3	202	14	11	6,7	3,6	1,1
9	17	0,9	0,3	202	22	19	7,4	4,1	1,2
10	55	0,7	0,2	290	53	58	6,8	3,4	1
11	57	0,6	0,1	210	51	59	5,9	2,7	0,8
12	72	2,7	1,3	260	31	56	7,4	3	0,6
13	64	0,9	0,3	310	61	58	7	3,4	0,9
14	74	1,1	0,4	214	22	30	8,1	4,1	1
15	61	0,7	0,2	145	53	41	5,8	2,7	0,87
∑	820	34	14,1	3948	523	623	84,1	47,5	9,7

Tabel 3. 4 Tabel Perhitungan untuk Mencari Nilai $\sum(X_n)^2$

Data	$(X_1)^2$	$(X_2)^2$	$(X_3)^2$	$(X_4)^2$	$(X_5)^2$	$(X_6)^2$	$(X_7)^2$	$(X_8)^2$	$(X_9)^2$
1	4225	0,49	0,01	34969	256	324	46,24	10,89	0,81
2	3844	118,81	30,25	488601	4096	10000	56,25	10,24	0,5476
3	3844	53,29	16,81	240100	3600	4624	49	10,89	0,7921
4	3364	1	0,16	33124	196	400	46,24	11,56	1
5	5184	15,21	4	38025	729	3481	53,29	5,76	0,16
6	2116	3,24	0,49	43264	361	196	57,76	19,36	1,69
7	676	0,81	0,04	23716	256	144	49	12,25	1
8	841	0,81	0,09	40804	196	121	44,89	12,96	1,21
9	289	0,81	0,09	40804	484	361	54,76	16,81	1,44
10	3025	0,49	0,04	84100	2809	3364	46,24	11,56	1
11	3249	0,36	0,01	44100	2601	3481	34,81	7,29	0,64
12	5184	7,29	1,69	67600	961	3136	54,76	9	0,36
13	4096	0,81	0,09	96100	3721	3364	49	11,56	0,81
14	5476	1,21	0,16	45796	484	900	65,61	16,81	1
15	3721	0,49	0,04	21025	2809	1681	33,64	7,29	0,7569
$\sum$	49134	205,12	53,97	1342128	23559	35577	741,49	174,23	13,2166

b. Menjumlahkan total jawaban dari setiap kelompok.

$$X_1 = \sum X_1 + \sum X_2 + \cdots + \sum X_n$$

$$= 820 + 34 + \cdots + 9{,}7 \ = 6103{,}4$$

c. Menghitung jumlah kuadrat antarbaris (JKB)

$$JKB \ = \left\{\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \cdots + \frac{(\sum X_n)^2}{n_n}\right\} - \frac{(\sum X_T)^2}{N}$$

$$= \left\{\frac{(820)^2}{15} + \frac{(34)^2}{15} + \cdots + \frac{(9{,}7)^2}{15}\right\} - \frac{(6103{,}4)^2}{135}$$

$$= \{44826{,}667 + 77{,}06667 + \cdots + 6{,}272667\} - 275936{,}9745$$

$$= 1128769{,}331 - 275936{,}975 = 852832{,}3565$$

d. Menentukan nilai derajat kebebasan antar grup

$$dk_B = A - 1$$

$$= 9 - 1 = 8$$

e. Menghitung nilai ragam antar grup

$$S_1^2 = \frac{JKB}{dk_B}$$

$$= \frac{852832{,}3561}{8} = 106604{,}0445$$

f. Menghitung nilai kuadrat dalam antargrup

$$JKD = [\sum(X_1)^2 + \sum(X_2)^2 + .. + \sum(X_n)^2] - \left[\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + .. + \frac{(\sum X_n)^2}{n_n}\right]$$

$$= \left[49134 + 205{,}12 + \cdots + 13{,}2166\right] - \left[\frac{(820)^2}{15} + \frac{(34)^2}{15} + \cdots + \frac{(9{,}7)^2}{15}\right]$$

$$= [1451586{,}027] - [44826{,}667 + 77{,}06667 + \cdots + 6{,}272667]$$

$$= [1451586{,}027] - [1128769{,}331]$$

$$= [322816{,}6959]$$

g. Menentukan nilai derajat kebebasan dalam antargrup

$$dk_D = K - A$$

$$= 135 - 9 = 126$$

h. Menentukan nilai ragam dalam antargrup

$$S_2^2 = \frac{JKD}{dk_D}$$

$$= \frac{322816{,}6959}{126} = 2562{,}037269$$

i. Menghitung nilai F_{hitung}

$$F_{hitung} = \frac{S_1^2}{S_2^2}$$

$$= \frac{106604,0445}{2562,0373}$$

$$= 41,6091$$

Berikut langkah-langkah dalam menghitung nilai F_{tabel} :

a. Nilai F_{tabel} dapat dicari dengan menggunakan table *F*

b. Dimana : dk_A = *pembilang* = 8 (diambil dari jumlah 9 kelompok yang ada di sample kemudian dikurangi 1) , dk_B = *penyebut* = 126 (diambil dari jumlah data tiap kelompok yaitu 15 data dan dikurangi 1, kemudian dikalikan 9 kelompok sample) dan a = 0,05.

$$F_{tabel} = F_{(a)(dk_A,dk_B)}$$

$$= F_{(0,05)(8,126)}$$

$$= 2,01$$

Kemudian setelah F_{hitung} dan F_{tabel} mendapatkan hasilnya, maka langkah selanjutnya membuat tabulasi ragam untuk ANOVA satu arah seperti Tabel 3.5.

Tabel 3. 5 Tabulasi Ragam

Sumber	Jumlah Kuadrat	Derajat Kebebasan	Ragam	F Rasio
Antargrup	850321,541	8	106290,193	41,525
Galat	322517,349	126	2559,662	
Total	1172838,891	134		

Pada tabel tabulasi diatas dapat di ambil kesimpulannya bahwa F_{hitung} 41,61 > F_{tabel} 2,01 maka H_0 ditolak. Jadi ada perbedaan nilai rata-rata performance yang signifikan antara 9 kelompok data sample. Selanjutnya, setelah uji ANOVA dilakukan dan hasil hipotesis nol ditolak, yang berarti ada bukti signifikan yang menunjukkan bahwa setidaknya satu rata-rata kelompok berbeda dari yang lain, pengujian post hoc dapat dilakukan untuk mengidentifikasi kelompok mana yang

berbeda secara signifikan satu sama lain. Uji Post Hoc yang bertujuan meranking nilai variabel yang paling mempengaruhi dataset. Berikut Gambar 3.5 hasil dari uji Post Hoc.

Multiple Comparisons

Dependent Variable: Nilai

Tukey HSD

(I) Variabel	(J) Variabel	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Age	TB	52,33333	18,47399	,116	-5,9786	110,6453
	DB	53,59333	18,47399	,098	-4,7186	111,9053
	AP	-208,53333*	18,47399	,000	-266,8453	-150,2214
	Alm	19,80000	18,47399	,977	-38,5120	78,1120
	Asp	13,13333	18,47399	,999	-45,1786	71,4453
	TP	47,66000	18,47399	,206	-10,6520	105,9720
	ALB	51,30000	18,47399	,133	-7,0120	109,6120
	AGR	53,75333	18,47399	,096	-4,5586	112,0653
TB	Age	-52,33333	18,47399	,116	-110,6453	5,9786
	DB	1,26000	18,47399	1,000	-57,0520	59,5720
	AP	-260,86667*	18,47399	,000	-319,1786	-202,5547
	Alm	-32,53333	18,47399	,707	-90,8453	25,7786
	Asp	-39,20000	18,47399	,463	-97,5120	19,1120
	TP	-4,67333	18,47399	1,000	-62,9853	53,6386
	ALB	-1,03333	18,47399	1,000	-59,3453	57,2786
	AGR	1,42000	18,47399	1,000	-56,8920	59,7320
DB	Age	-53,59333	18,47399	,098	-111,9053	4,7186
	TB	-1,26000	18,47399	1,000	-59,5720	57,0520
	AP	-262,12667*	18,47399	,000	-320,4386	-203,8147
	Alm	-33,79333	18,47399	,663	-92,1053	24,5186
	Asp	-40,46000	18,47399	,419	-98,7720	17,8520
	TP	-5,93333	18,47399	1,000	-64,2453	52,3786
	ALB	-2,29333	18,47399	1,000	-60,6053	56,0186
	AGR	,16000	18,47399	1,000	-58,1520	58,4720
AP	Age	208,53333*	18,47399	,000	150,2214	266,8453
	TB	260,86667*	18,47399	,000	202,5547	319,1786
	DB	262,12667*	18,47399	,000	203,8147	320,4386
	Alm	228,33333*	18,47399	,000	170,0214	286,6453
	Asp	221,66667*	18,47399	,000	163,3547	279,9786
	TP	256,19333*	18,47399	,000	197,8814	314,5053
	ALB	259,83333*	18,47399	,000	201,5214	318,1453
	AGR	262,28667*	18,47399	,000	203,9747	320,5986

Alm	Age	-19,80000	18,47399	,977	-78,1120	38,5120
	TB	32,53333	18,47399	,707	-25,7786	90,8453
	DB	33,79333	18,47399	,663	-24,5186	92,1053
	AP	-228,33333*	18,47399	,000	-286,6453	-170,0214
	Asp	-6,66667	18,47399	1,000	-64,9786	51,6453
	TP	27,86000	18,47399	,850	-30,4520	86,1720
	ALB	31,50000	18,47399	,742	-26,8120	89,8120
	AGR	33,95333	18,47399	,657	-24,3586	92,2653
Asp	Age	-13,13333	18,47399	,999	-71,4453	45,1786
	TB	39,20000	18,47399	,463	-19,1120	97,5120
	DB	40,46000	18,47399	,419	-17,8520	98,7720
	AP	-221,66667*	18,47399	,000	-279,9786	-163,3547
	Alm	6,66667	18,47399	1,000	-51,6453	64,9786
	TP	34,52667	18,47399	,636	-23,7853	92,8386
	ALB	38,16667	18,47399	,501	-20,1453	96,4786
	AGR	40,62000	18,47399	,413	-17,6920	98,9320
TP	Age	-47,66000	18,47399	,206	-105,9720	10,6520
	TB	4,67333	18,47399	1,000	-53,6386	62,9853
	DB	5,93333	18,47399	1,000	-52,3786	64,2453
	AP	-256,19333*	18,47399	,000	-314,5053	-197,8814
	Alm	-27,86000	18,47399	,850	-86,1720	30,4520
	Asp	-34,52667	18,47399	,636	-92,8386	23,7853
	ALB	3,64000	18,47399	1,000	-54,6720	61,9520
	AGR	6,09333	18,47399	1,000	-52,2186	64,4053
ALB	Age	-51,30000	18,47399	,133	-109,6120	7,0120
	TB	1,03333	18,47399	1,000	-57,2786	59,3453
	DB	2,29333	18,47399	1,000	-56,0186	60,6053
	AP	-259,83333*	18,47399	,000	-318,1453	-201,5214
	Alm	-31,50000	18,47399	,742	-89,8120	26,8120
	Asp	-38,16667	18,47399	,501	-96,4786	20,1453
	TP	-3,64000	18,47399	1,000	-61,9520	54,6720
	AGR	2,45333	18,47399	1,000	-55,8586	60,7653
AGR	Age	-53,75333	18,47399	,096	-112,0653	4,5586
	TB	-1,42000	18,47399	1,000	-59,7320	56,8920
	DB	-,16000	18,47399	1,000	-58,4720	58,1520
	AP	-262,28667*	18,47399	,000	-320,5986	-203,9747
	Alm	-33,95333	18,47399	,657	-92,2653	24,3586
	Asp	-40,62000	18,47399	,413	-98,9320	17,6920
	TP	-6,09333	18,47399	1,000	-64,4053	52,2186
	ALB	-2,45333	18,47399	1,000	-60,7653	55,8586

Gambar 3. 5 Hasil Uji Post Hoc

Gambar 3.5 Hasil dari uji post hoc, semakin kecil nilai rata-rata signifikansi pada variabel terhadap satu sama lain, maka semakin berpengaruh variabel tersebut. Fitur-fitur yang memiliki p-value kurang dari 0.05 dianggap signifikan dan dipilih

untuk analisis lebih lanjut. Hasil seleksi fitur menunjukkan bahwa atribut terpilih yaitu Age, AP, dan Asp

3.4 Normalisasi *Min Max*

Normalization merupakan suatu proses untuk mengubah skala dari nilai atribut pada data sehingga nilai tersebut berada pada rentang tertentu (Nasution et al., 2019). Implementasi proses normalisasi pada penelitian ini menggunakan teknik Min Max Scaling, yang mana membagi selisih antara nilai variabel terbesar dan nilai variabel terkecil dengan seluruh nilai variabel yang telah dikurangi nilai variabel terkecil. Tujuan dari dilakukannya normalisasi ini untuk mendapatkan nilai rentang, sehingga mampu menghasilkan citra yang memiliki nilai piksel dengan rentang antara 0 dan 1. Output proses normalisasi ini akan digunakan pada proses selanjutnya yaitu proses segmentation. Perhitungan untuk proses normalisasi menggunakan Min Max Scaling sebagai berikut (Pramana et al., 2020).

$$Inormalisasi = \frac{I - Imin}{Imax - Imin} \quad (3.1)$$

Keterangan :
Inormalisasi : Nilai yang dinormalisasi
I : Nilai asli dalam data
Imin : Nilai terkecil dalam data
Imax : Nilai terbesar dalam data

Berikut adalah hasil proses normalisasi yang telah diimplementasikan menggunakan Min Max Scaling yang dapat dilihat pada Tabel 3.6.

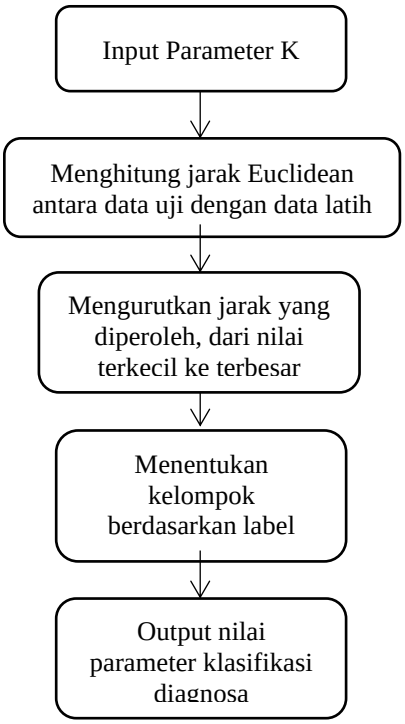
Tabel 3. 6 Hasil Normalisasi Min Max

No	Age	AP	Asp	Alb	AGR
1	0,842105	0,075812	0,078652	0,45	0,555556
2	0,789474	1	1	0,4	0,377778
3	0,789474	0,622744	0,640449	0,45	0,544444
4	0,719298	0,066787	0,101124	0,5	0,666667
5	0,964912	0,090253	0,539326	0	0
6	0,508772	0,113718	0,033708	1	1
7	0,157895	0,016245	0,011236	0,55	0,666667
8	0,210526	0,102888	0	0,6	0,777778
9	0	0,102888	0,089888	0,85	0,888889
10	0,666667	0,261733	0,52809	0,5	0,666667
11	0,701754	0,117329	0,539326	0,15	0,444444
12	0,964912	0,207581	0,505618	0,3	0,222222
13	0,824561	0,297834	0,52809	0,5	0,555556
14	1	0,124549	0,213483	0,85	0,666667
15	0,77193	0	0,337079	0,15	0,522222

3.5 *K-Nearest Neighbor*

K-Nearest Neighbor (KNN) adalah sebuah algoritma pembelajaran mesin yang digunakan untuk masalah klasifikasi. Pada proses ini akan dilakukan pengujian terhadap data uji dan data latih guna menemukan model klasifikasi yang menghasilkan akurasi paling tinggi. Tahap mining pada penelitian kali ini akan dilakukan dengan menggunakan metode K-Nearest Neighbor (k-NN). Dalam proses pengolahan data peneliti akan dilakukan dengan melakukan perhitungan secara matematis dengan cara manual dan penghitungan menggunakan aplikasi SPSS sebagai alat bantu klasifikasi. Dengan menggunakan metode k-NN akan didapatkan euclidean distance yang nantinya akan diurutkan berdasarkan jarak ketetangga-an terdekat untuk mendapatkan model. Pada tahap ini akan didapatkan model klasifikasi dari data yang telah diolah berupa tabel *Confusion Matrix* yang

akan menampilkan hasil pengujian yang dihasilkan oleh perhitungan manual dan akan disesuaikan dengan hasil pada data asli. Alur proses pada metode *K-Nearest Neighbor* (KNN) digambarkan pada Gambar 3.6 berikut.



Gambar 3. 6 Alur algoritma KNN

Alur pada metode K-Nearest Neighbors dimulai dari input parameter k, kemudian menghitung jarak *Euclidean* antara data uji dengan data latih. Selanjutnya langkah yang dapat dilakukan untuk mengklasifikasikan data dengan menggunakan metode K-Nearest Neighbor adalah sebagai berikut :

1. Menyiapkan data latih dan data uji
2. Menentukan nilai k
3. Melakukan perhitungan nilai jarak antara data latih dengan data uji. Dengan rumus penghitung jarak euclidean ada pada persamaan (3.2) (Hastie et al., 2009):

$$d_{(p,q)} = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2 + \cdots +(p_n-q_n)^2} \quad (3.2)$$

Keterangan:

$d_{(p,q)}$: Jarak Euclidean antara data latih dan data uji
p = $(p_1,\ p_2,\ \ldots, p_n)$: Data latih
q = $(q_1,\ q_2,\ \ldots, q_n)$: Data uji

4. Mengelompokkan data berdasarkan perhitungan jarak.
5. Mengelompokkan data berdasarkan nilai tetangga terdekat.
6. Memilih nilai yang sering muncul dari tetangga terdekat sebagai acuan prediksi data selanjutnya.

Pada sample yang digunakan dalam penelitian ini berupa data penyakit liver dengan 15 sample data dan 9 variabel. Data telah melalui tahap seleksi fitur menggunakan uji ANOVA serta data telah dinormalisasikan menggunakan *min max*. Selanjutnya data di uji menggunakan algoritma *K-Nearest Neighbor*. Berikut merupakan perhitungan dengan rumus Euclidean pada persamaan (3.2).

$$d_i = \sqrt{(0{,}842105623 - 0{,}771929825)^2 + (0{,}075812274 - 0)^2 + (0{,}078651685 - 0{,}337078652)^2}$$

$$d_i = \sqrt{0{,}07745659}$$

$$d_i = 0{,}278310241$$

Tabel 3.7 menunjukkan hasil perhitungan KNN dengan metode ANOVA, dimana data latih 90% dan data uji 10%. Nilai K yang digunakan K=1, K=3, dan K=5 hasil data uji yang sesuai dengan ketetanggaan terdekat yaitu Y atau pasien terkena liver.

Tabel 3.7 Perhitungan KNN dengan ANOVA

JARAK EU	Rank	K=1	K=3	K=5
0,278310241	3		Y	Y
1,199905205	14			
0,692929516	11			
0,250809441	2		Y	Y
0,293754499	5			Y
0,417394046	9			
0,695324722	12			
0,662858866	10			
0,817046516	13			
0,340690051	7			
0,244119848	1	Y	Y	Y
0,329753997	6			
0,357715569	8			
0,287757261	4			Y
0	?	Y	Y	Y

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Skenario Uji Coba

Pada penelitian ini, untuk mengetahui evaluasi hasil dataset yang ada dibagi menjadi dua subset yaitu data latih dan data uji. Dalam persentase yang dilakukan Harafani & Al-Kautsar, (2021), data dibagi menjadi 90% data latih dan 10% data uji, 80% data latih dan 20% data uji, serta 70% data latih dan 30% data uji. Hasil prediksi yang didapat, kemudian dilakukan perhitungan untuk mendapatkan nilai akurasi, presisi, recall, dan f-1 score menggunakan metode confusion matrix. Selanjutnya untuk menentukan nilai K terbaik dilakukannya proses K-Fold Cross Validation. Pada skenario uji coba ini dilakukan untuk mengetahui pengaruhnya terhadap tingkat akurasi yang dihasilkan.

4.2 Data Penelitian

Pada penelitian ini akan digunakan dataset yang tersedia untuk umum pada situs UCI Machine Learning "https://archive.ics.uci.edu/dataset/225/". Dataset memuat 583 data dengan 10 atribut. Dimana 10 atribut berupa input dan 1 atribut berupa kelas. Kelas yang dimaksud disini yaitu menyatakan seorang pasien tersebut terdiagnosa penyakit liver atau tidak. Target yang bernilai '1' untuk pasien mengalami penyakit liver dan '2' untuk pasien tidak mengalami penyakit liver. Dataset yang telah diimport dapat dilihat pada Gambar 4.1.

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio	Dataset
0	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.90	1
1	62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
2	62	Male	7.3	4.1	490	60	68	7.0	3.3	0.89	1
3	58	Male	1.0	0.4	182	14	20	6.8	3.4	1.00	1
4	72	Male	3.9	2.0	195	27	59	7.3	2.4	0.40	1

```
[[65.    0.7   0.1  ...  6.8   3.3   0.9 ]
 [62.   10.9   5.5  ...  7.5   3.2   0.74]
 [62.    7.3   4.1  ...  7.    3.3   0.89]
 ...
 [52.    0.8   0.2  ...  6.4   3.2   1.  ]
 [31.    1.3   0.5  ...  6.8   3.4   1.  ]
 [38.    1.    0.3  ...  7.3   4.4   1.5 ]]
```

Gambar 4. 1 Dataset Import

Setelah pengumpulan data selesai, langkah selanjutnya yaitu melakukan proses seleksi fitur menggunakan metode *Analysis of Variance* untuk memperoleh fitur-fitur yang lebih berpengaruh terhadap prediksi.

4.3 Hasil Seleksi Fitur

Pada tahap seleksi fitur ini peneliti menggunakan metode *one way* ANOVA untuk seleksi atribut. Analysis of Variance (ANOVA) digunakan untuk menilai pentingnya setiap fitur terhadap variabel target. Fitur-fitur yang memiliki p-value kurang dari 0.05 dianggap signifikan dan dipilih untuk analisis lebih lanjut. Seleksi fitur ini digunakan untuk menyeleksi atribut mana yang lebih berpengaruh pada kelas dataset. Gambar 4.2 hasil dari seleksi fitur menggunakan *one way* ANOVA.

```
Selected Feature Indices: [1 2 3]

array([False,  True,  True,  True, False, False, False, False, False])
```

Gambar 4. 2 Hasil Seleksi Atribut menggunakan ANOVA

Dapat disimpulkan bahwa nilai “True” adalah atribut yang lebih mempengaruhi dataset sedangkan “False” adalah atribut yang tidak mempengaruhi dataset. Terdapat 3 atribut terpilih yaitu angka 1, 2 dan 3 yang artinya Total

Bilirubin, Direct Bilirubin, dan Alkaline Phosphotase. Langkah selanjutnya atribut terpilih akan dilakukan uji normalisasi menggunakan metode *min-max* untuk memperoleh angka dengan nilai rentang yang sama dan memudahkan analisis data.

4.4 Hasil Normalisasi *Min-Max*

Selanjutnya setelah dilakukan tahap seleksi fitur, data akan dilakukan normalisasi menggunakan metode *Min-Max*. Pada *normalization* data bertujuan untuk membuat semua data berada dalam skala yang sama dalam rentang 0 hingga 1. Hasil dataset yang telah dinormalisasi menggunakan code Python ada pada Tabel 4.1 sebagai berikut:

Tabel 4. 1 Hasil Normalisasi menggunakan *MinMax*

Age	**Total Bilirubin (TB)**	***Direct Bilirubin*** **(DB)**	**…**	**Total *Protiens* (TP)**	**Albumin (ALB)**	***Albumin Globulin Ratio*** **(AGR)**
0.28571429	0.00402145	0.00510204	…	0.44927536	0.54347826	0.44
0.53571429	0.03351206	0.06122449	…	0.62318841	0.32608696	0.08
0.64285714	0.02010724	0.03571429	…	0.26086957	0.19565217	0.12
0.42857143	0.00670241	0.00510204	…	0.84057971	0.76086957	0.28
0.64285714	0.02546917	0.0255102	…	0.56521739	0.56521739	0.32
0.69047619	0.01340483	0.02040816	…	0.65217391	0.36956522	0.08

Tabel diatas menunjukkan perubahan pada nilai-nilai setiap fitur dengan nilai angka rentang 0 sampai 1. Selanjutnya data yang telah dinormalisasi, dilakukan proses pembagian data (*split data*) yaitu data dibagi menjadi data latih dan data uji. Kemudian dilanjutkan proses pengujian menggunakan model yang telah ditetapkan pada penelitian ini.

4.5 Menghitung Kinerja Sistem

Pada tahap ini bertujuan untuk memberikan hasil performa dari suatu model klasifikasi dan gambaran menyeluruh tentang hasil prediksi model. Data pada penelitian ini terdapat total 583 data dengan diagnosa liver dan non liver. Selanjunya data dibagi menjadi dua yaitu data latih dan data uji. Data latih merupakan data yang nantinya untuk melatih mesin dalam mengenali pola pada sistem. Data uji merupakan data untuk menguji hasil yang telah dilakukan pelatihan terhadap mesin. Dalam uji coba ini menggunakan model *K-Nearest Neighbor*, seleksi fitur dan *setting* parameter k harus ditentukan terlebih dahulu untuk mengoptimalkan klasifikasi.

Setelah sistem dibangun, dilakukan evaluasi untuk menghitung akurasi, presisi, *recall*, dan f-1 *score*. Evaluasi terhadap model digunakan metode *confusion matrix*, dimana hasil prediksi dari model klasifikasi akan dibandingkan dengan hasil sebenarnya (aktual) pada data uji atau validasi. Pada tahap ini akan menampilkan jumlah data yang diklasifikasikan dengan benar atau salah dalam setiap kategori, baik kategori positif maupun negatif. *Confusion matrix* terdiri dari empat kategori, yaitu *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Dapat dilihat Tabel 4.2 merupakan tabel dari *confusion matrix*.

Tabel 4. 2 Tabel Confusion Matrix

Kelas	Diklasifikasikan sebagai Positif	Diklasifikasikan sebagai Negatif
+	True Positive (TP)	False Positive (FP)
_	False Negative (FN)	True Negative (TN)

Tabel 4.2 merupakan tabel *confusion matrix* yang mana masing-masing memiliki pengertian sebagai berikut:

1. TP (*True Positive*) : Jumlah data penyakit liver yang diklasifikasikan benar sebagai positif oleh model klasifikasi.
2. FP (*False Positive*) : Jumlah data penyakit liver yang diklasifikasikan salah sebagai positif oleh model klasifikasi.
3. TN (*True Negative*) : Jumlah data non liver yang diklasifikasikan benar sebagai negatif oleh model klasifikasi.
4. FN (*False Negative*) : Jumlah data non liver yang diklasifikasikan salah sebagai negatif oleh model klasifikasi.

Berdasarkan kategori tersebut, kemudian dihitung akurasi, presisi, *recall*, dan *F1-score*. Adapun rumus untuk menghitung akurasi, presisi, dan recall masing-masing ditunjukkan pada rumus persamaan (4.1), (4.2), (4.3), dan (4.4) (Goutte & Gaussier, 2005).

$$. Accuracy = \frac{TP+TN}{(TP+FP+TN+FN} \quad (4.1)$$

Akurasi merupakan pengukuran performa yang terdiri dari presisi dan *recall*. Presisi adalah rasio prediksi positif yang benar dengan hasil keseluruhan dari prediksi positif.

$$Presisi = \frac{TP}{TP+FP} \quad (4.2)$$

$$Recall = \frac{TP}{TP+FN} \quad (4.3)$$

Recall adalah rasio antara prediksi yang benar-benar positif (*true positive*) dengan data global yang sebenarnya positif. *F*1-*Score* adalah perbandingan antara presisi rata-rata dan recall.

$$F1 - score = 2 \times \frac{Presisi * Recall}{Presisi + Recall} \quad (4.4)$$

4.6 Uji Hasil Pengujian

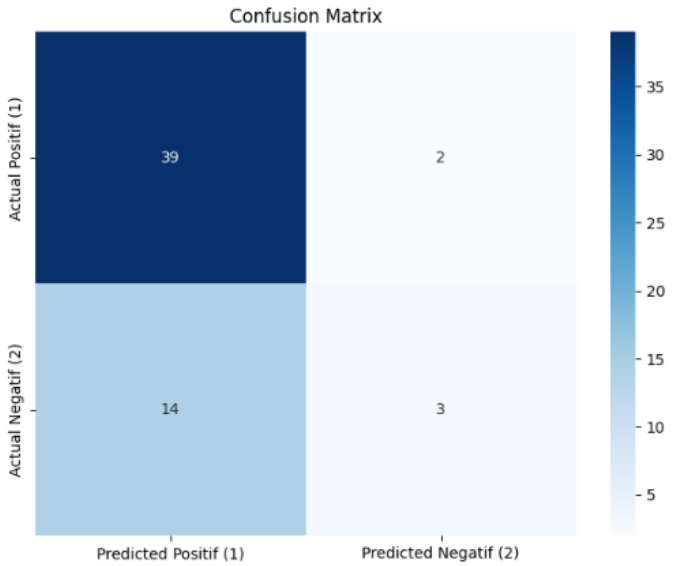
4.6.1 Uji Coba Rasio Data 90:10

Skenario uji coba yang pertama menggunakan ratio data 90:10 dengan nilai k tetangga terdekat 1 sampai 5 dimana pengujian algoritma KNN ini menggunakan seleksi fitur ANOVA dan tanpa seleksi fitur. Dibawah ini merupakan hasil uji coba algoritma KNN dengan seleksi fitur menggunakan ANOVA. Hasil *confusion matrix* yang didapatkan pada rasio 90:10 untuk akurasi, *precision*, *recall*, dan *f1-score* tersebut ditunjukkan pada Tabel 4.3.

Tabel 4. 3 Nilai *Accuracy, Precision, Recall, F1-Score* pada Rasio Data 90:10 dengan ANOVA

Pengujian Nilai K-	Rasio Data 90:10			
	Accuracy	*Precision*	*Recall*	*F1-Score*
1	60,34%	70,45%	75,60%	72,94%
2	65,52%	69,81%	90,24%	78,72%
3	62,07%	69,38%	82,92%	75,55%
4	70,69%	70,68%	100%	82,82%
5	72,41%	73,58%	95,12%	82,97%
Rata-rata	66,21%	70,78%	88,78%	78.6%

Pada Tabel 4.3 menunjukkan akurasi tertinggi didapatkan pada pengujian nilai k-5 sebesar 72,41% dengan menggunakan seleksi fitur ANOVA yang artinya mampu memprediksi dengan baik oleh sistem. Selanjutnya nilai k-5 pada *confusion matrix* ditunjukkan pada Gambar 4.3.



Gambar 4. 3 *Confusion Matrix* 90:10 pada k-5 dengan ANOVA

Dalam dataset angka "1" sebagai liver dan angka "2" sebagai non liver. Hasil nilai *confusion matrix* menunjukkan terdapat 39 data TP (*True Positif*) yang mampu memprediksi liver dengan benar oleh sistem, 14 data FP (*False Positif*) yang salah memprediksi liver tetapi kenyataan non liver, 2 data FN (*False Negatif*) yang salah memprediksi non liver tetapi kenyataan liver, dan 3 data TN (*True Negatif*) sistem berhasil memprediksi data dengan benar sebagai non liver.

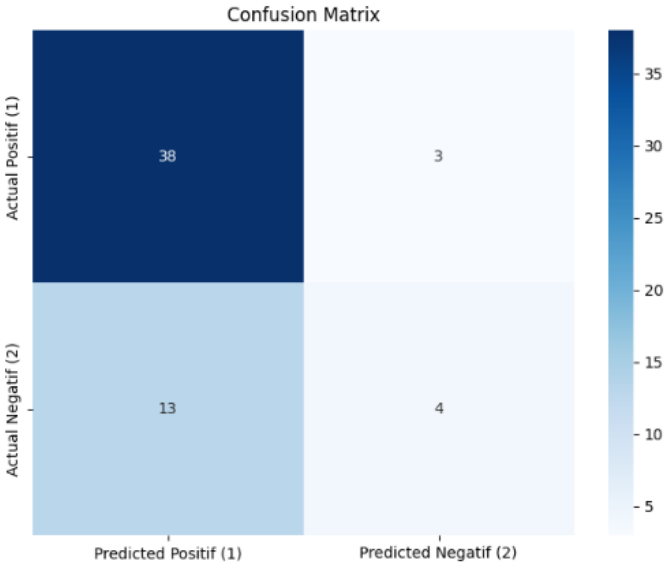
Selanjutnya ratio data 90:10 dengan nilai k tetangga terdekat 1 sampai 5 dimana pengujian algoritma KNN ini tanpa menggunakan seleksi fitur. Dibawah ini merupakan hasil uji coba algoritma KNN tanpa seleksi fitur. Hasil yang didapatkan pada rasio 90:10 untuk akurasi, *precision*, *recall*, dan *f1-score* tersebut ditunjukkan pada Tabel 4.4.

Tabel 4. 4 Nilai *Accuracy, Precision, Recall, F1-Score* pada Rasio Data 90:10 tanpa Seleksi Fitur

Pengujian Nilai K-	Rasio Data 90:10			
	Accuracy	*Precision*	*Recall*	*F1-Score*
1	65,52%	75,60%	75,60%	75,60%

Pengujian Nilai K-	Rasio Data 90:10			
	Accuracy	*Precision*	*Recall*	*F1-Score*
2	72,41%	74,50%	92,68%	82,60%
3	63,79%	72,72%	78,04%	75,29%
4	67,24%	72%	87,80%	79,12%
5	65,52%	73%	80,4%	76,74%
Rata-rata	66,9%	73,56%	82,90%	77,87%

Pada Tabel 4.4 menunjukkan akurasi tertinggi didapatkan pada pengujian nilai k-2 sebesar 72,41% tanpa menggunakan seleksi fitur yang artinya mampu memprediksi dengan baik oleh sistem, namun dalam kondisi tertentu algoritma KNN tanpa seleksi fitur menghasilkan akurasi lebih rendah dibandingkan menggunakan seleksi fitur. Selanjutnya nilai k-2 pada *confusion matrix* ditunjukkan pada Gambar 4.4.



Gambar 4. 4 *Confusion Matrix* 90:10 pada k-2 tanpa Seleksi Fitur

Pada Gambar 4.4 hasil nilai *confusion matrix* menunjukkan terdapat 38 data TP (*True Positif*) yang mampu memprediksi liver dengan benar oleh sistem, 13 data FP (*False Positif*) yang salah memprediksi liver tetapi kenyataan non liver, 3 data

FN (*False Negatif*) yang salah memprediksi non liver tetapi kenyataan liver, dan 4 data TN (*True Negatif*) sistem berhasil memprediksi data dengan benar sebagai non liver.

Hasil yang diberikan pada rasio data 90:10 hampir sama, namun lebih baik oleh K-Nearest Neighbors dengan seleksi fitur dibandingkan tanpa seleksi fitur dengan terdapat perbedaan selisih 2,44% pada nilai *recall* yang lebih besar yaitu 95,15%.

4.6.2 Uji Coba Rasio Data 80:20

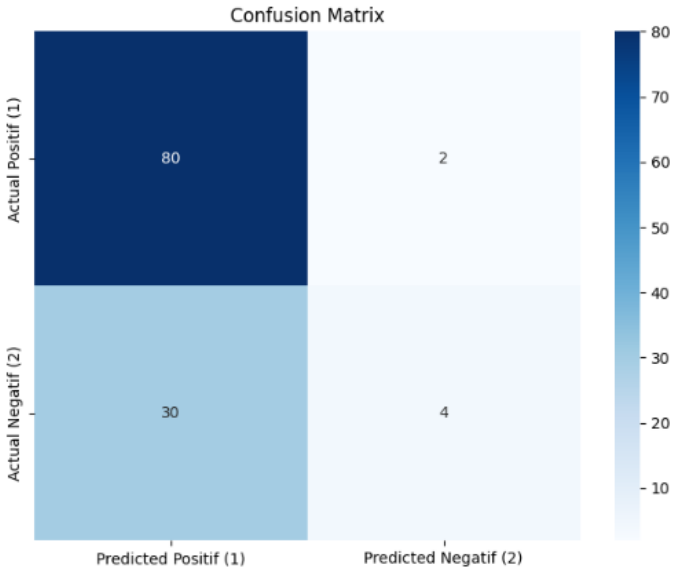
Skenario uji coba yang kedua menggunakan ratio data 80:20 dengan nilai k tetangga terdekat 1 sampai 5 dimana pengujian algoritma KNN ini menggunakan seleksi fitur ANOVA dan tanpa seleksi fitur. Dibawah merupakan hasil uji coba model KNN dengan seleksi fitur ANOVA. Hasil yang didapatkan pada rasio 80:20 untuk akurasi, *precision*, *recall*, dan *f1-score* ditunjukkan pada Tabel 4.5.

Tabel 4. 5 Nilai *Accuracy, Precision, Recall, F1-Score* pada Rasio Data 80:20 dengan ANOVA

Pengujian Nilai K-	Rasio Data 80:20			
	Accuracy	*Precision*	*Recall*	*F1-Score*
1	70,69%	76%	85,36%	80,45%
2	72,41%	72,72%	97,56%	83,33%
3	66,38%	71,71%	86,58%	78,45%
4	72,41%	72,72%	97,56%	83,33%
5	68,97%	73%	89%	80,21%
Rata-rata	70,17%	73,23%	91,21%	81,15%

Pada Tabel 4.5 menunjukkan akurasi tertinggi didapatkan pada pengujian nilai k-2 dan k-4 memiliki akurasi yang sama sebesar 72,41% dengan menggunakan

seleksi fitur ANOVA yang artinya mampu memprediksi dengan baik oleh sistem. Selanjutnya nilai k-2 pada *confusion matrix* ditunjukkan pada Gambar 4.5..



Gambar 4. 5 *Confusion Matrix* 80:20 pada k-2 dengan ANOVA

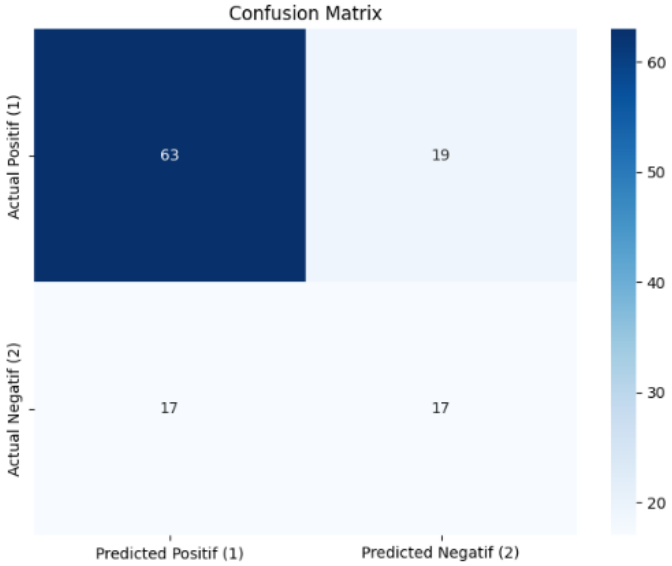
Hasil nilai *confusion matrix* pada Gambar 4.5 menunjukkan terdapat 80 data TP (*True Positif*) yang mampu memprediksi liver dengan benar oleh sistem, 30 data FP (*False Positif*) yang salah memprediksi liver tetapi kenyataan non liver, 2 data FN (*False Negatif*) yang salah memprediksi non liver tetapi kenyataan liver, dan 4 data TN (*True Negatif*) sistem berhasil memprediksi data dengan benar sebagai non liver.

Selanjutnya ratio data 80:20 dengan nilai k tetangga terdekat 1 sampai 5 dimana pengujian algoritma KNN ini tanpa menggunakan seleksi fitur. Dibawah ini hasil uji coba algoritma KNN tanpa seleksi fitur. Hasil yang didapatkan pada rasio 80:20 untuk akurasi, *precision*, *recall*, dan *f1-score* ditunjukkan pada Tabel 4.6.

Tabel 4. 6 Nilai *Accuracy, Precision, Recall, F1-Score* pada Rasio Data 80:20 tanpa Seleksi Fitur

Pengujian Nilai K-	Rasio Data 80:20			
	Accuracy	*Precision*	*Recall*	*F1-Score*
1	68,97%	78,75%	76,82%	77,8%
2	68,10%	71,42%	91,46%	80,21%
3	68,10%	75,86%	80,48%	78,10%
4	65,52%	71%	86,58%	78,02%
5	62,93%	72,41%	76,82%	74,55%
Rata-rata	66,72%	73,89%	82,43%	77,74%

Pada Tabel 4.6 menunjukkan akurasi tertinggi didapatkan pada pengujian nilai k-1 sebesar 68,97% tanpa menggunakan seleksi fitur yang artinya kurang mampu dalam memprediksi penyakit liver dengan baik oleh sistem, akurasi yang dihasilkan jauh lebih rendah dibandingkan menggunakan seleksi fitur. Selanjutnya nilai k-1 pada *confusion matrix* ditunjukkan pada Gambar 4.6.



Gambar 4. 6 *Confusion Matrix* 80:20 pada k-1 tanpa Seleksi Fitur

Hasil nilai *confusion matrix* pada Gambar 4.6 menunjukkan terdapat 63 data TP (*True Positif*) yang mampu memprediksi liver dengan benar oleh sistem, 17 data FP (*False Positif*) yang salah memprediksi liver tetapi kenyataan non liver, 19 data FN (*False Negatif*) yang salah memprediksi non liver tetapi kenyataan liver, dan 17

data TN (*True Negatif*) sistem berhasil memprediksi data dengan benar sebagai non liver. Hasil yang diberikan pada rasio data 80:20 berhasil diprediksi jauh lebih baik oleh K-Nearest Neighbors dengan seleksi fitur dibandingkan tanpa seleksi fitur dengan nilai akurasi selisih 3,44%, *recall* 97,56%, dan *f1-score* 83,33% yang lebih besar.

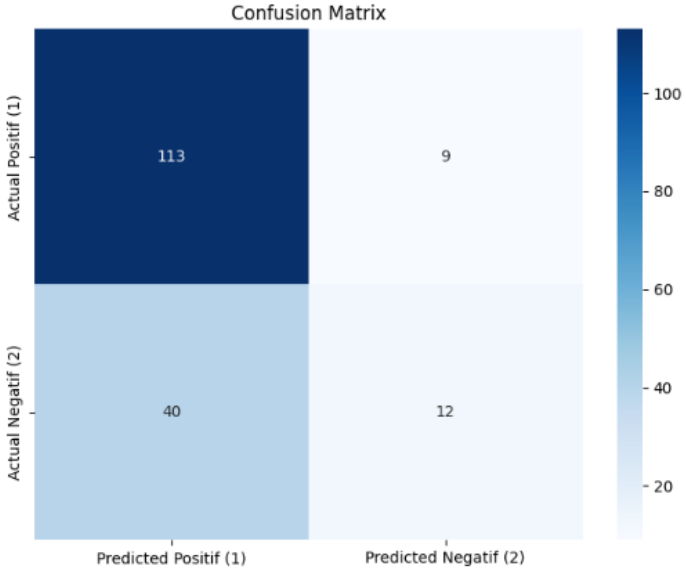
4.6.3 Uji Coba Rasio Data 70:30

Skenario uji coba yang ketiga menggunakan ratio data 70:30 dengan nilai k tetangga terdekat 1 sampai 5 dimana pengujian KNN ini menggunakan seleksi fitur ANOVA dan tanpa seleksi fitur. Dibawah merupakan hasil uji coba algoritma KNN dengan seleksi fitur menggunakan ANOVA. Hasil yang didapatkan pada rasio 70:30 untuk akurasi, *precision*, *recall*, dan *f1-score* ditunjukkan pada Tabel 4.7.

Tabel 4. 7 Nilai *Accuracy, Precision, Recall, F1-Score* pada Rasio Data 70:30 dengan ANOVA

Pengujian Nilai K-	Rasio Data 70:30			
	Accuracy	*Precision*	*Recall*	*F1-Score*
1	64,94%	71,03%	84,42%	77,15%
2	68,97%	70,23%	96,72%	81,37%
3	68,97%	72,66%	89,34%	80,14%
4	70,11%	71%	96,72%	81,94%
5	71,84%	73,85%	92,62%	82,18%
Rata-rata	68,97%	71,75%	91,96%	80,56%

Pada Tabel 4.7 menunjukkan akurasi tertinggi didapatkan pada pengujian nilai k-5 sebesar 71,84% dengan menggunakan seleksi fitur ANOVA yang artinya mampu memprediksi penyakit liver dengan baik oleh sistem. Selanjutnya nilai k-5 pada *confusion matrix* ditunjukkan pada Gambar 4.7.



Gambar 4. 7 *Confusion Matrix* 70:30 pada k-5 dengan ANOVA

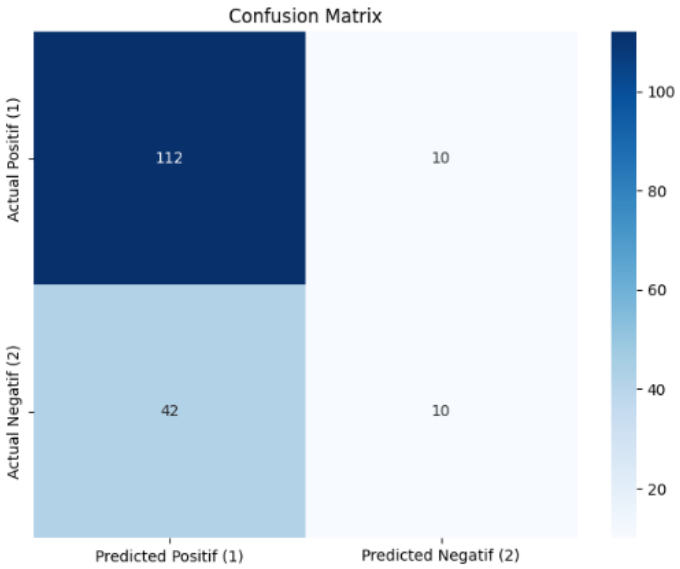
Hasil nilai *confusion matrix* pada Gambar 4.7 menunjukkan terdapat 113 data TP (*True Positif*) yang mampu memprediksi liver dengan benar oleh sistem, 40 data FP (*False Positif*) yang salah memprediksi liver tetapi kenyataan non liver, 9 data FN (*False Negatif*) yang salah memprediksi non liver tetapi kenyataan liver, dan 12 data TN (*True Negatif*) sistem berhasil memprediksi data dengan benar sebagai non liver.

Selanjutnya ratio data 70:30 dengan nilai k tetangga terdekat 1 sampai 5 dimana pengujian algoritma KNN ini tanpa menggunakan seleksi fitur. Dibawah ini merupakan hasil uji coba algoritma KNN tanpa seleksi fitur. Hasil yang didapatkan pada rasio 70:30 untuk akurasi, *precision*, *recall*, dan *f1-score* tersebut ditunjukkan pada Tabel 4.8.

Tabel 4. 8 Nilai *Accuracy, Precision, Recall, F1-Score* pada Rasio Data 70:30 tanpa Seleksi Fitur

Pengujian Nilai K-	Rasio Data 70:30			
	Accuracy	*Precision*	*Recall*	*F1-Score*
1	70,11%	77,34%	81,14%	79,2%
2	70,11%	72,72%	91,80%	81,15%
3	66,09%	73,68%	80,32%	76,86%
4	67,24%	71,52%	88,52%	79,12%
5	63,22%	71,01%	80,32%	75,38%
Rata-rata	67,35%	73,25%	84,42%	78,34%

Pada Tabel 4.8 menunjukkan akurasi tertinggi didapatkan pada pengujian nilai k-1 dan k-2 yang memiliki kesamaan akurasi sebesar 70,11% tanpa menggunakan seleksi fitur yang artinya mampu dalam memprediksi penyakit liver dengan baik oleh sistem, akan tetapi akurasi yang dihasilkan lebih rendah dibandingkan menggunakan seleksi fitur. Selanjutnya nilai k-1 pada *confusion matrix* ditunjukkan pada Gambar 4.8.



Gambar 4. 8 *Confusion Matrix* 70:30 pada k-1 tanpa Seleksi Fitur

Hasil nilai *confusion matrix* pada Gambar 4.8 menunjukkan terdapat 112 data TP (*True Positif*) yang mampu memprediksi liver dengan benar oleh sistem,

42 data FP (*False Positif*) yang salah memprediksi liver tetapi kenyataan non liver, 10 data FN (*False Negatif*) yang salah memprediksi non liver tetapi kenyataan liver, dan 10 data TN (*True Negatif*) sistem berhasil memprediksi data dengan benar sebagai non liver.

Hasil yang diberikan pada rasio data 70:30 berhasil diprediksi jauh lebih baik oleh K-Nearest Neighbors dengan seleksi fitur dibandingkan tanpa seleksi fitur dengan nilai akurasi, *recall*, presisi, dan *f1-score* yang lebih besar. Artinya pada rasio data 70:30 ini sangat berpengaruh pada nilai yang dihasilkan oleh model KNN menggunakan seleksi fitur.

4.7 K - *Fold Cross Validation*

Dalam proses membagi dataset, biasanya untuk mengukur kinerja model dengan lebih akurat dilakukannnya proses *cross validation*. *Cross validation* ini terbagi menjadi beberapa jenis, seperti *Holdout Cross Validation, K-Fold Cross Validation, Stratified K-Fold Cross Validation, Time Series Split Cross Validation*, dan sebagainya. Namun kali ini peneliti akan menggunakan *K-Fold Cross Validation* dalam mengukur kinerja model dan proses ini dilakukan sebanyak 5 kali pada data. Metode ini bertujuan untuk memberikan gambaran yang lebih akurat tentang performa model dengan membagi dataset menjadi beberapa subset atau *folds*. Proses ini mengurangi kemungkinan overfitting dan memberikan estimasi yang lebih stabil dari performa model. Berikut hasil dari *K-Fold Cross Validation* dapat dilihat pada Tabel 4.9.

Tabel 4. 9 Tabel Hasil Akurasi dari *K-Fold Cross Validation*

Fold ke-	Hasil Akurasi
1	67%
2	64%
3	66%
4	66%
5	65%
Rata-rata	66%
Standar Deviasi	0,010198

Gambar diatas menunjukkan hasil rata-rata akurasi dengan *K-Fold Cross Validation* memiliki hasil akurasi 66% dengan standar deviasi 0,010198 menggunakan model K-Nearest Neighbors. Semakin kecil nilai standar deviasi yang dihasilkan maka semakin terpusat penyebaran nilai-nilai item pada data dengan mean. Pada penelitian ini, nilai standar deviasi 0,010198 menunjukkan bahwa variasi hasil akurasi antara fold cukup rendah, mengindikasikan konsistensi performa model di berbagai subset data. Selanjutnya rata-rata akurasi 66% menunjukkan bahwa, secara keseluruhan, model dapat memprediksi dengan benar 66% dari kasus dalam dataset.

4.8 Pembahasan

Berdasarkan hasil dari beberapa skenario uji coba dan seleksi fitur yang dilakukan terhadap 583 data penyakit liver. Pada hasil seleksi fitur menggunakan Analysis of Variance digunakan untuk menilai pentingnya setiap fitur terhadap variabel target. Fitur-fitur yang memiliki p-value kurang dari 0.05 dianggap signifikan dan dipilih untuk analisis lebih lanjut. Hasil seleksi fitur menunjukkan bahwa fitur Total Bilirubin, Direct Bilirubin, dan Alkaline Phosphotase, signifikan terhadap prediksi penyakit liver. Kemudian hasil seleksi tersebut digunakan pada beberapa uji coba KNN dengan seleksi fitur dan KNN tanpa seleksi fitur diperoleh

akurasi terbaik pada masing-masing rasio yaitu hasil akurasi menggunakan seleksi fitur lebih tinggi dibandingkan tanpa seleksi dengan nilai rentang k-1 hingga k-5. Hasil klasifikasi KNN tanpa ANOVA dan hasil KNN dengan ANOVA dapat dilihat pada Tabel 4.10 dan Tabel 4.11.

Tabel 4. 10 Hasil Klasifikasi KNN tanpa ANOVA

Nilai K-(Ketetanggaan)	Rasio Data		
	90/10	80/20	70/30
1	65,52%	68,97%	70,11%
2	72,41%	68,10%	70,11%
3	63,79%	68,10%	66,09%
4	67,24%	65,52%	67,24%
5	65,52%	62,93%	63,22%

Tabel 4. 11 Hasil Klasifikasi KNN dengan ANOVA

Nilai K-(Ketetanggaan)	Rasio Data		
	90/10	80/20	70/30
1	60,34%	70,69%	64,94%
2	65,52%	72,41%	68,97%
3	62,07%	66,38%	68,97%
4	70,69%	72,41%	70,11%
5	72,41%	68,97%	71,84%

Dari seluruh skenario uji coba yang dilakukan, hasil rata-rata terbaik yang didapatkan ada pada rasio data 80:20 mengggunakan seleksi fitur dengan nilai rata-rata yaitu diperoleh akurasi 70,17%, presisi 73,23%, *recall* 91,21%, dan f1-score 81,15%. Dari hasil uji coba tersebut menunjukkan bahwa akurasi tertinggi diperoleh dengan dilakukannnya pemilihan atribut dengan seleksi fitur menggunakan ANOVA. Hal ini diketahui bahwa dari percobaan pertama, kedua, dan ketiga menunjukkan adanya peningkatan akurasi terhadap sistem. Percobaan kedua ini menunjukkan bahwa pemilihan atribut dengan seleksi fitur memiliki pengaruh terhadap kinerja sistem. Kemudian dari hasil uji coba kedua dilanjutkan

untuk pengujian dengan menggunakan teknik *K-fold Cross Validation*, dengan pemilihan nilai k sebesar 5, dan dilakukan iterasi sebanyak 5 kali. Dimana dari hasil pengujian ini, hasil menunjukkan rata-rata akurasi *K-fold Cross Validation* memberikan gambaran yang lebih akurat tentang kinerja model dibandingkan metode validasi sederhana seperti train-test split. Dalam penelitian ini, model memiliki akurasi rata-rata 66% dengan variasi yang rendah, menunjukkan bahwa model memiliki performa yang stabil pada berbagai subset data. Metode ini sangat penting dalam memastikan bahwa model tidak hanya bekerja baik pada satu subset data, tetapi memiliki generalisasi yang baik pada data yang tidak terlihat sebelumnya.

Pengukuran hasil evaluasi model yang dilakukan pada saat uji coba menghasilkan kesimpulan bahwa sistem berhasil dengan baik mengoptimasikan pengklasifikasian penyakit liver ke dalam kelas liver dan non liver menggunakan algoritma K-Nearest Neighbors dengan seleksi fitur menggunakan *Analysis of Variance*. Serta pengujian dengan teknik *K-fold Cross Validation* dengan pemilihan nilai k-5 dengan hasil standar deviasi yang kecil sehingga menghasilkan akurasi terbaik. Dalam penelitian ini, hasil akurasi yang diperoleh dengan metode K-Nearest Neighbors sangat dipengaruhi oleh beberapa faktor. Diantaranya penggunaan atribut pada data, rasio pembagian data, proses preprocessing dengan tepat, penggunaan fitur seleksi, dan pengujian dengan penggunaan nilai k- pada teknik *K-fold Cross Validation*. Penggunaan pemilihan fitur dengan metode ANOVA dan pemilihan nilai k untuk teknik *fold Cross Validation* pada dataset yang telah diuji mempengaruhi hasil nilai akurasi yang diperoleh. Hal ini dapat

diamati pada hasil uji coba seleksi fitur semakin sedikit atribut yang digunakan bahwa semakin tinggi tingkat akurasi yang dihasilkan.

Selain itu, nilai akurasi juga bergantung pada proses preprocessing yang tepat dan data yang digunakan hasil dari proses pelabelan yang dilakukan pada data training, hal ini dikarenakan metode K-Nearest Neighbors termasuk ke dalam pendekatan supervised learning, di mana hasil prediksi kelas berdasarkan pada pola yang dihasilkan dari pengolahan data training yang diimplementasikan pada model. Dari hasil penelitian yang telah dilakukan, sistem dapat diterapkan untuk kepentingan klasifikasi khususnya dalam memprediksi kesehatan. Sebagaimana firman Allah subhanahu wa ta'ala dalam Surah Al-Hijr Ayat 21 dalam Al Qur'an yang berbunyi:

وَإِن مِّن شَيْءٍ إِلَّا عِندَنَا خَزَائِنُهُۥ وَمَا نُنَزِّلُهُ إِلَّا بِقَدَرٍ مَّعْلُومٍ

"Tidak ada sesuatu pun melainkan di sisi Kamilah perbendaharaannya dan Kami tidak menurunkannya melainkan dengan ukuran tertentu." (QS. Al-Hijr :21)

Menurut tafsir Jalalain mengenai ayat ini yaitu (Dan tiada) tidak ada (sesuatu pun melainkan pada sisi Kamilah khazanahnya) huruf min adalah zaidah; yang dimaksud adalah kunci-kunci perbendaharaan segala sesuatu itu (dan Kami tidak menurunkannya melainkan dengan ukuran-ukuran yang tertentu) sesuai dengan kepentingan-kepentingannya. Tafsir tersebut menunjukkan bahwa segala sesuatu yang Allah ciptakan dengan ketelitian dan kebijaksanaan. Maka, dengan adanya sistem klasifikasi ini diharapkan mampu melakukan klasifikasi sesuai ukurannya yang akurat, seperti Allah menciptakan sesuatu sesuai dengan ukuannya. Salah satu tujuan ANOVA yaitu untuk mendapatkan ukuran

yang baik dengan cara seleksi fitur untuk memperoleh fitur yang signifikan. Dengan demikian mampu mengurangi resiko kesalahan dalam pengambilan keputusan karena salahnya informasi yang digunakan untuk klasifikasi data dalam proses pengembangan aplikasi. Salah satunya sistem dapat memprediksi penyakit liver dan dapat membantu memudahkan tenaga kesehatan mendapatkan informasi.

Kenyataan memang membuktikan, kebanyakan manusia terserang penyakit disebabkan kurang memperhatikan norma-norma kesehatan yang berlaku. Salah satunya peyakit liver didapati dengan makan-makanan dan minum-minuman yang diharamkan oleh Allah. Sebagaimana firman Allah subhanahu wa ta'ala dalam Surah Al-Baqarah Ayat 168 dalam Al Qur'an yang berbunyi:

يٰٓاَيُّهَا النَّاسُ كُلُوْا مِمَّا فِى الْاَرْضِ حَلٰلًا طَيِّبًاۖ وَّلَا تَتَّبِعُوْا خُطُوٰتِ الشَّيْطٰنِۗ اِنَّهٗ لَكُمْ عَدُوٌّ مُّبِيْنٌ

"Wahai manusia, makanlah sebagian (makanan) di bumi yang halal lagi baik dan janganlah mengikuti langkah-langkah setan. Sesungguhnya ia bagimu merupakan musuh yang nyata." (QS. Al-Baqarah :168)

Dari ayat diatas menurut tafsir Jalalain diturunkannya tentang bagaimana orang-orang diperintahkan untuk makan-makanan yang halal lagi baik dan jangan mengikuti langkah setan. Dan orang-orang yang bertaqwa akan memperoleh apa yang diperoleh oleh hamba-hamba Allah, mereka adalah orang-orang yang selalu mengikuti petunjuk Allah dan selalu menggunakan akal yang sehat. Maka kita sebagai orang mukmin perlu diperhatikan makanan dan minuman yang akan masuk kedalam tubuh kita, agar tubuh tetap sehat dan tidak memberikan dampak negative yang serius bagi kesahatan.

BAB V

PENUTUP

4.9 Kesimpulan

Berdasarkan hasil uji yang diperoleh, dapat disimpulkan bahwa K-Nearest Neighbors dengan seleksi fitur menggunakan *Analysis of Variance* mampu memprediksi penyakit liver dengan baik sehingga menghasilkan nilai akurasi tinggi dalam penelitian ini. Pada beberapa uji coba dengan rasio data yang berbeda, nilai yang dihasilkan memiliki tingkat akurasi yang besar dibandingkan tanpa seleksi fitur. Penelitian ini menggunakan rasio data sebesar 90%, 80%, dan 70% pada data yang berjumlah 583 data pasien dengan nilai k yang digunakan 1 sampai 5. Dari seluruh skenario uji coba yang dilakukan, hasil rata-rata terbaik yang didapatkan ada pada rasio data 80:20 mengggunakan seleksi fitur dengan nilai rata-rata yaitu diperoleh akurasi 70,17%, presisi 73,23%, *recall* 91,21%, dan f1-score 81,15%. Pada penelitian ini, hasil seleksi fitur yang lebih berpengaruh terhadap variabel target yaitu Total Bilirubin, Direct Bilirubin, dan Alkaline Phosphotase.

4.10 Saran

Saran yang dapat penulis sampaikan untuk penelitian selanjutnya adalah:

1. Dapat melakukan pengujian dengan memperluas variasi model K-Nearest Neighbors dan Analysis of Variance untuk meningkatkan akurasi prediksi.
2. Dapat melakukan penelitian terkait dengan menggunakan metode *Machine Learning* lain untuk memprediksi penyakit liver.

DAFTAR PUSTAKA

Alassaf, M., & Qamar, A. M. (2022). Improving sentiment analysis of Arabic tweets by One-Way ANOVA. *Journal of King Saud University-Computer and Information Sciences*, *34*(6), 2849–2859.

Desiani, A. (2022). Perbandingan Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Klasifikasi Penyakit Hati. *SIMKOM*, *7*(2), 104–110. https://doi.org/10.51717/simkom.v7i2.96

Gupta, S. C., & Goel, N. (2020). Performance enhancement of diabetes prediction by finding optimum K for KNN classifier with feature selection method. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 980–986. https://doi.org/10.1109/ICSSIT48917.2020.9214129

Harafani, H., & Al-Kautsar, H. A. (2021). MENINGKATKAN KINERJA K-NN UNTUK KLASIFIKASI KANKER PAYUDARA DENGAN SELEKSI FITUR. *Jurnal Pendidikan Teknologi dan Kejuruan*, *18*(1).

Herdiana, Y., & Geraldine, W. A. (n.d.). PENERAPAN MACHINE LEARNING DENGAN MODEL LINEAR REGRESSION TERHADAP ANALISIS KUALITAS HASIL PETIK THE DI PT. PERKEBUNAN NUSANTARA VIII KEBUN SEDEP. *Jurnal Informatika*, *09*.

Khairiah, L., & Rismawan, T. (2017). *SISTEM PAKAR DIAGNOSIS PENYAKIT HATI DENGAN METODE DEMPSTER SHAFER BERBASIS ANDROID*. *05*(2).

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, *197*, 116659. https://doi.org/10.1016/j.eswa.2022.116659.

Kustiyahningsih, Y., Nurhamidin, Q. A., & Purnama, J. (2021). *Feature Selection and K-nearest Neighbor for Diagnosis Cow Disease*. *05*(02).

Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine Learning* (0 ed.). CRC Press. https://doi.org/10.1201/9781315371658

Nurmalasari, M. D., Kusrini, K., & Sudarmawan, S. (2021). Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor untuk Membangun Pengetahuan Diagnosa Penyakit Diabetes. *Jurnal Komtika (Komputasi dan Informatika)*,

5(1), 52–59. https://doi.org/10.31603/komtika.v5i1.5140

Oktavyani, A. R., Wicaksono, A., Seanne, A. F., Karolin, A. D., Putra, R. S., & Kurniawan, M. (n.d.). *Perbandingan Metode Naive Bayes, K-NN dan Decision Tree Terhadap Dataset Healthcare Stroke*.

Pramana, I. G. B. B. A., Widiartha, I. M., & Astuti, L. G. (2020). Implementation Learning Vector Quantization (LVQ) for Chronic Kidney Disease Classification. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, *9*(2), 241. https://doi.org/10.24843/JLK.2020.v09.i02.p11

Pusporani, E., Qomariyah, S., & Irhamah, I. (2019). Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning. *Inferensi*, *2*(1), 25. https://doi.org/10.12962/j27213862.v2i1.6810

Rafsanjani, R. G., Hidayat, N., & Dewi, R. K. (n.d.). *Diagnosis Penyakit Hati Menggunakan Metode Naive Bayes Dan Certainty Factor*.

Rahman, N. T. (2020). ANALISA ALGORITMA DECISION TREE DAN NAÏVE BAYES PADA PASIEN PENYAKIT LIVER. *JURNAL FASILKOM*, *10*(2), 144–151. https://doi.org/10.37859/jf.v10i2.2087

Saragih, N. B. (2022). *Sistem Pakar Mendiagnosa Penyakit Gangguan Hati Pada Manusia Menggunakan Metode Naïve Bayes Berbasis WEB*.

Sarker, I., Faruque, Md., Alqahtani, H., & Kalim, A. (2018). K-Nearest Neighbor Learning based Diabetes Mellitus Prediction and Analysis for eHealth Services. *ICST Transactions on Scalable Information Systems*, *0*(0), 162737. https://doi.org/10.4108/eai.13-7-2018.162737

Setianto, Y. A., Kusrini, K., & Henderi, H. (2019). Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur. *Creative Information Technology Journal*, *5*(3), 232. https://doi.org/10.24076/citec.2018v5i3.179

Setiawati, I., Permana, A., & Hermawan, A. (2019). IMPLEMENTASI DECISION TREE UNTUK MENDIAGNOSIS PENYAKIT LIVER. *Journal of Information System Management (JOISM)*, *1*(1), 13–17. https://doi.org/10.24076/JOISM.2019v1i1.17

Sivakrishnan, D. S., & Pharm, M. (n.d.). LIVER DISEASES-AN OVERVIEW. *World Journal of Pharmacy and Pharmaceutical Sciences*, *8*(1).

Swantika, I. M. A., Kanata, B., & Suksmadana, I. M. B. (2023). PERANCANGAN SISTEM UNTUK MENGETAHUI KUALITAS BIJI KOPI BERDASARKAN WARNA DENGAN K-NEAREST NEIGHBOR. *Jurnal*

Bakti Nusa, *1*(2. Oktober 2020), 25–36.

Wijaya, J. T., Oktavianto, H., & Al Faruq, H. A. (2022). Perbandingan Algoritma K-Nearest Neighbor (Knn) Dan Gaussian Naive Bayes (Gnb) Dalam Klasifikasi Breast Cancer Coimbra. *Jurnal Smart Teknologi*, *3*(3), 233–237.

Goutte, C., & Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and FScore, with Implication for Evaluation. Lecture Notes in Computer Science, 3408, 345–359. https://doi.org/10.1007/978-3- 540-31865-1_25

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning. Springer Series in Statistics.* doi:10.1007/978-0-387-84858-7

Ramana, Bendi & Surendra, M & Babu, Prasad & Bala Venkateswarlu, Nagasuri. (2012). *A Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis.* Diakses 13 Agustus 2023 dari https://archive.ics.uci.edu/dataset/225/

Kwak S. Are Only *p*-Values Less Than 0.05 Significant? A *p*-Value Greater Than 0.05 Is Also Significant! J Lipid Atheroscler. 2023 May;12(2):89-95. doi: 10.12997/jla.2023.12.2.89. Epub 2023 May 3. PMID: 37265851; PMCID: PMC10232224.

LAMPIRAN

Lampiran 1: Hasil Prediksi Rasio Data 90:10

Rasio Data 90:10				
No	**Seleksi Fitur**		**Tanpa Seleksi Fitur**	
	Real	**Prediksi**	**Real**	**Prediksi**
1	1	1	1	1
2	2	2	2	1
3	1	1	1	1
4	2	2	2	2
5	2	1	2	1
6	2	1	2	1
7	2	1	2	2
8	2	1	2	2
9	1	1	1	1
10	1	1	1	1
11	2	2	2	2
12	1	1	1	2
13	2	1	2	1
14	1	1	1	1
15	1	1	1	2
16	1	1	1	1
17	2	1	2	1
18	1	1	1	1
19	1	1	1	1
20	1	1	1	1
21	1	1	1	1
22	1	1	1	1
23	1	1	1	2
24	1	1	1	2
25	1	1	1	1
26	1	1	1	1
27	1	1	1	1
28	1	2	1	2
29	1	1	1	1
30	1	1	1	1
31	2	1	2	1
32	1	1	1	1
33	1	1	1	2
34	1	1	1	1
35	2	1	2	1
36	2	1	2	1
37	2	1	2	1
38	1	1	1	1
39	1	1	1	1
40	1	1	1	1
41	2	1	2	2
42	1	1	1	1
43	1	1	1	1
44	1	1	1	1
45	1	1	1	1
46	1	1	1	1
47	2	1	2	1

48	1	1	1	1
49	1	2	1	1
50	1	1	1	1
51	2	1	2	1
52	1	1	1	1
53	1	1	1	1
54	1	1	1	2
55	1	1	1	1
56	1	1	1	2
57	1	1	1	1
58	2	1	2	1

Lampiran 2: Hasil Prediksi Rasio Data 80:20

Rasio Data 80:20				
No	**Seleksi Fitur**		**Tanpa Seleksi Fitur**	
	Real	**Prediksi**	**Real**	**Prediksi**
1	1	1	1	1
2	2	2	2	1
3	1	1	1	1
4	2	1	2	2
5	2	1	2	2
6	2	1	2	1
7	2	1	2	1
8	2	1	2	2
9	1	1	1	1
10	1	1	1	2
11	2	1	2	2
12	1	1	1	2
13	2	1	2	1
14	1	1	1	1
15	1	1	1	2
16	1	1	1	1
17	2	1	2	1
18	1	1	1	1
19	1	1	1	1
20	1	1	1	1
21	1	1	1	1
22	1	1	1	1
23	1	1	1	2
24	1	1	1	2
25	1	1	1	1
26	1	1	1	1
27	1	1	1	1
28	1	1	1	2
29	1	1	1	1
30	1	1	1	1
31	2	2	2	1
32	1	1	1	1
33	1	1	1	2
34	1	1	1	1
35	2	1	2	1

36	2	1	2	1
37	2	1	2	1
38	1	1	1	2
39	1	1	1	1
40	1	2	1	1
41	2	1	2	1
42	1	1	1	1
43	1	1	1	1
44	1	1	1	1
45	1	2	1	2
46	1	1	1	1
47	2	1	2	2
48	1	1	1	1
49	1	1	1	1
50	1	1	1	1
51	2	1	2	1
52	1	2	1	1
53	1	1	1	1
54	1	1	1	2
55	1	1	1	1
56	1	2	1	1
57	1	1	1	1
58	2	1	2	1
59	2	1	2	2
60	1	1	1	1
61	2	1	2	1
62	1	1	1	1
63	1	1	1	1
64	1	1	1	1
65	1	2	1	2
66	2	2	2	2
67	1	1	1	1
68	2	2	2	1
69	1	1	1	1
70	1	2	1	1
71	2	1	2	1
72	2	2	2	1
73	1	1	1	1
74	1	1	1	1
75	2	1	2	2
76	1	1	1	1
77	1	1	1	1
78	2	2	2	1
79	1	1	1	1
80	1	1	1	2
81	2	1	2	1
82	1	1	1	2
83	1	1	1	1
84	1	1	1	1
85	1	1	1	1
86	2	2	2	1
87	1	2	1	1
88	1	1	1	2

89	1	1	1	1
90	2	1	2	1
91	1	1	1	1
92	1	2	1	1
93	2	1	2	1
94	1	1	1	2
95	2	1	2	2
96	1	1	1	1
97	1	1	1	1
98	1	1	1	1
99	1	1	1	2
100	2	1	2	1
101	2	1	2	1
102	2	1	2	1
103	1	1	1	1
104	1	1	1	1
105	1	1	1	1
106	1	1	1	1
107	1	1	1	1
108	1	1	1	1
109	2	1	2	2
110	1	1	1	2
111	1	1	1	2
112	1	1	1	1
113	1	1	1	2
114	1	1	1	1
115	1	2	1	1
116	1	1	1	1

Lampiran 3: Hasil Prediksi Rasio Data 70:30

Rasio Data 70:30				
No	**Seleksi Fitur**		**Tanpa Seleksi Fitur**	
	Real	**Prediksi**	**Real**	**Prediksi**
1	1	1	1	1
2	2	1	2	1
3	1	1	1	1
4	2	1	2	2
5	2	1	2	2
6	2	2	2	1
7	2	1	2	1
8	2	1	2	2
9	1	1	1	1
10	1	1	1	2
11	2	2	2	1
12	1	1	1	2
13	2	1	2	1
14	1	1	1	1
15	1	1	1	2
16	1	1	1	1
17	2	1	2	1
18	1	1	1	1
19	1	1	1	1

20	1	1	1	1
21	1	1	1	1
22	1	1	1	1
23	1	2	1	2
24	1	1	1	1
25	1	1	1	1
26	1	1	1	1
27	1	1	1	1
28	1	2	1	2
29	1	1	1	1
30	1	2	1	2
31	2	1	2	1
32	1	1	1	1
33	1	1	1	2
34	1	1	1	1
35	2	1	2	1
36	2	1	2	1
37	2	1	2	1
38	1	1	1	2
39	1	1	1	1
40	1	1	1	1
41	2	1	2	1
42	1	1	1	1
43	1	1	1	2
44	1	1	1	1
45	1	1	1	2
46	1	1	1	1
47	2	1	2	2
48	1	1	1	1
49	1	1	1	1
50	1	1	1	1
51	2	1	2	1
52	1	1	1	1
53	1	1	1	1
54	1	1	1	2
55	1	1	1	1
56	1	1	1	1
57	1	1	1	1
58	2	1	2	1
59	2	1	2	2
60	1	1	1	1
61	2	2	2	1
62	1	1	1	1
63	1	1	1	1
64	1	1	1	2
65	1	1	1	1
66	2	2	2	2
67	1	1	1	1
68	2	2	2	2
69	1	1	1	1
70	1	2	1	1
71	2	1	2	1
72	2	1	2	1

73	1	1	1	1
74	1	1	1	1
75	2	1	2	2
76	1	1	1	1
77	1	1	1	1
78	2	2	2	1
79	1	1	1	1
80	1	1	1	2
81	2	1	2	1
82	1	1	1	2
83	1	1	1	1
84	1	2	1	1
85	1	1	1	1
86	2	1	2	1
87	1	1	1	1
88	1	1	1	2
89	1	1	1	1
90	2	1	2	1
91	1	1	1	1
92	1	1	1	1
93	2	1	2	1
94	1	1	1	2
95	2	1	2	1
96	1	1	1	1
97	1	1	1	1
98	1	1	1	1
99	1	1	1	2
100	2	1	2	1
101	2	1	2	1
102	2	1	2	1
103	1	1	1	1
104	1	1	1	1
105	1	1	1	1
106	1	1	1	1
107	1	1	1	1
108	1	1	1	1
109	2	1	2	1
110	1	1	1	2
111	1	2	1	1
112	1	1	1	1
113	1	1	1	2
114	1	1	1	1
115	1	2	1	1
116	1	1	1	1
117	1	1	1	1
118	1	1	1	1
119	1	1	1	1
120	1	1	1	2
121	2	1	2	1
122	1	1	1	1
123	2	2	2	1
124	1	1	1	1
125	1	1	1	1

126	1	1	1	1
127	1	1	1	1
128	1	1	1	1
129	1	2	1	1
130	1	1	1	1
131	1	1	1	2
132	2	1	2	1
133	2	2	2	1
134	2	1	2	1
135	1	1	1	1
136	1	1	1	1
137	2	1	2	1
138	1	1	1	1
139	2	1	2	1
140	1	1	1	1
141	1	1	1	1
142	1	1	1	1
143	1	1	1	1
144	1	1	1	1
145	1	1	1	2
146	1	1	1	1
147	2	1	2	2
148	2	1	2	1
149	2	2	2	1
150	1	1	1	1
151	2	1	2	1
152	1	1	1	1
153	1	1	1	1
154	1	1	1	1
155	2	2	2	2
156	1	1	1	1
157	1	1	1	1
158	1	1	1	1
159	2	1	2	1
160	1	1	1	1
161	1	1	1	1
162	2	1	2	2
163	2	2	2	1
164	1	1	1	1
165	1	2	1	2
166	1	1	1	1
167	1	1	1	1
168	1	1	1	1
169	1	1	1	2
170	1	1	1	1
171	2	2	2	1
172	2	1	2	1
173	1	1	1	1
174	2	1	2	2