

**ANALISIS DETEKSI DINI PENYAKIT JANTUNG DENGAN
PENDEKATAN REGRESI LOGISTIK PADA DATA PASIEN**

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Mendapatkan Gelar Sarjana
Komputer(S.Kom) Program Studi Informatika



IKHLASHUL AMAL

105841105220

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MAKASSAR

2024



UNIVERSITAS MUHAMMADIYAH MAKASSAR FAKULTAS TEKNIK

GEDUNG MENARA IQRA LT. 3

Jl. Sultan Alauddin No. 259 Telp. (0411) 866 972 Fax (0411) 865 588 Makassar 90221

Website: www.unismuh.ac.id, e_mail: unismuh@gmail.com

Website: <http://teknik.unismuh.makassar.ac.id>

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

PENGESAHAN

Skripsi atas nama Ikhlashul Amal dengan nomor induk Mahasiswa 105 84 11052 20, dinyatakan diterima dan disahkan oleh Panitia Ujian Tugas Akhir/Skripsi sesuai dengan Surat Keputusan Dekan Fakultas Teknik Universitas Muhammadiyah Makassar Nomor : 0008/SK-Y/55202/0991004/2024, sebagai salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar pada hari Sabtu tanggal 26 Agustus 2024.

Panitia Ujian :

Makassar, 21 safar 1446 H
26 Agustus 2024 M

1. Pengawas Umum

a. Rektor Universitas Muhammadiyah Makassar

Dr. Ir. H. Abd. Rakhim Nanda, ST., MT., IPU

b. Dekan Fakultas Teknik Universitas Hasanuddin

Prof. Dr. Eng. Muhammad Isran Ramli, ST., MT.

2. Penguji

a. Ketua : Dr. Ir. Hafsah/Nirwana, M.T.

b. Sekretaris : Lukman, S.Kom., M.T.

3. Anggota

1. Lukman Anas, S.Kom., M.T.

2. Titin Wahyuni, S.Pd., M.T.

3. Muhyiddin A M Hayyat, S.Kom., MT.

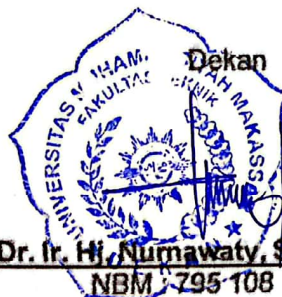
Mengetahui :

Pembimbing I

Rizki Yusliana Bakti, S.T., M.T

Pembimbing II

Fahrir Irhamna Rachman, S.Kom., M.T



Dr. Ir. Hj. Numawaty, ST., MT., IPM.

NBM : 795-108



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan untuk memenuhi syarat ujian guna memperoleh gelar Sarjana Komputer (S.Kom) Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar.

Judul Skripsi : **ANALISIS DETEKSI DINI PENYAKIT JANTUNG DENGAN PENDEKATAN REGRESI LOGISTIK PADA DATA PASIEN**

Nama : Ikhlashul Amal

Stambuk : 105841105220

Makassar, 28 Agustus 2024

Telah Diperiksa dan Disetujui
Oleh Dosen Pembimbing;

Pembimbing I

Pembimbing II

Rizki Yusliana Bakti, S.T., M.T

Fahrir Irhamna Rachman, S.Kom., M.T

Mengetahui,

Ketua Program Studi Informatika



Muhyiddin A. M. Hayat, S.Kom., MT.

NBM : 1504 577

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan risiko penyakit jantung pada pasien dengan menggunakan metode clustering dan regresi logistik. Data yang digunakan mencakup 640 pasien di RSUD Haji Makassar, di mana setiap pasien diklasifikasikan ke dalam tiga kategori risiko: rendah, tinggi, dan sangat tinggi. Metode Elbow digunakan untuk menentukan jumlah cluster optimal, dan hasil analisis menunjukkan bahwa tiga cluster merupakan jumlah yang paling sesuai untuk data ini. Cluster 0 mencakup pasien dengan risiko rendah, Cluster 1 mencakup pasien dengan risiko tinggi, dan Cluster 2 mencakup pasien dengan risiko sangat tinggi. Selanjutnya, regresi logistik diterapkan untuk memprediksi risiko penyakit jantung berdasarkan variabel kesehatan yang tersedia, seperti usia, tekanan darah, dan kadar kolesterol. Hasil akurasi dari metode regresi logistic tersebut sebesar 98.44%. Hasil penelitian ini menunjukkan bahwa kombinasi metode clustering dan regresi logistik efektif dalam mengelompokkan pasien dan memprediksi tingkat risiko penyakit jantung, yang dapat membantu dalam pengambilan keputusan medis dan perencanaan perawatan kesehatan yang lebih tepat.

Kata kunci: *Clustering*, Regresi Logistik, *Elbow*, *K-Means*, Risiko Penyakit Jantung

ABSTRACT

This study aims to classify heart disease risk in patients using clustering and logistic regression methods. The data includes 640 patients from RSUD Haji Makassar, where each patient is classified into three risk categories: low, high, and very high. The Elbow method was used to determine the optimal number of clusters, and the analysis results indicate that three clusters are the most suitable for this data. Cluster 0 includes patients with low risk, Cluster 1 includes patients with high risk, and Cluster 2 includes patients with very high risk. Subsequently, logistic regression was applied to predict heart disease risk based on available health variables such as age, blood pressure, and cholesterol levels. The accuracy of the logistic regression method was 98.44%. The results of this study show that the combination of clustering and logistic regression methods is effective in grouping patients and predicting heart disease risk levels, which can assist in medical decision-making and more precise healthcare planning.

Keywords: *Clustering, Logistic Regression, Elbow, K-Means, Heart Disease Risk*

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu'alaikum Warahmatullah Wabarakatuh

Puji syukur penulis panjatkan kehadiran Allah Subhanahu Wata'ala atas limpahan rahmat dan hidayah-Nya sehingga proposal skripsi yang berjudul **“Analisis Deteksi Dini Penyakit Jantung Dengan Pendekatan Regresi Logistik Pada Data Pasien”**.

Proposal skripsi ini diajukan untuk memenuhi salah satu persyaratan yang harus ditempuh dalam rangka menyelesaikan studi di Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar. Meskipun proposal skripsi ini telah selesai, hal ini tidak berarti bahwa proposal tersebut sudah dalam bentuk yang sempurna. Oleh karena itu, saran dan kritikan dari pembaca sangat diharapkan untuk meningkatkan kualitas proposal skripsi ini.

Penulis menyadari bahwa selama dalam penyusunan proposal skripsi ini banyak pihak yang telah membantu dan memberikan dukungannya baik secara material maupun moral. Demikian juga segala bantuan yang penulis peroleh selama ini dibangku perkuliahan sehingga penulis merasa sangat bersyukur dan mengucapkan banyak terima kasih kepada segenap pihak yang telah membantu penulis. Oleh karena itu, penghargaan yang setinggi-tingginya kami hanturkan dengan hormat kepada:

1. **Kedua Orang Tua kami tercinta**, penulis mengucapkan terima kasih yang sebesar-besarnya atas segala limpahan kasih sayang, doa dan dukungan baik secara moral maupun materi.
2. **Ibu Dr. Ir. Hj. Nurnawaty, ST., MT.,IPM**, sebagai Dekan Fakultas Teknik Universitas Muhammadiyah Makassar.
3. **Bapak Muhyiddin AM Hayat, S.Kom.,MT**, sebagai Ketua Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Makassar.
4. **Bapak Fahrin Irhamna Rachman S.Kom.,MT**, selaku Dosen Pembimbing 2 Skripsi.

5. **Ibu Rizki Yusliana Bakti, S.T., MT**, selaku Dosen Pembimbing 1 Skripsi.
6. Bapak/Ibu Dosen dan Staff Administrasi Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar.
7. Saudara M.syahdan dan Saudari Andi Raodatul Adawiyah Ramli, yang telah membantu dalam penyusunan skripsi ini.
8. Sahabat – sahabat penulis di Fakultas Teknik terkhusus Fitra M natsir, Rina Permata Sari, M. Syahdan, Iqbal, Aryansyah Iskandar, yang selalu belajar dan berjuang bersama dengan rasa persaudaraan yang tinggi banyak membantu serta memberi dukungan dalam menyelesaikan tugas skripsi ini.

Penulis mengharapkan tugas proposal skripsi ini dapat bermanfaat bagi pengembangan ilmu pengetahuan khususnya dibidang keinformatikaan, Aamiin.

“Billahi fii sabilil haq fastabiqul khairat”

“Wassalamu’alaikum Warahmatullah Wabarakatuh”

Makassar, 20 Agustus 2024

IKHLASHUL AMAL

DAFTAR ISI

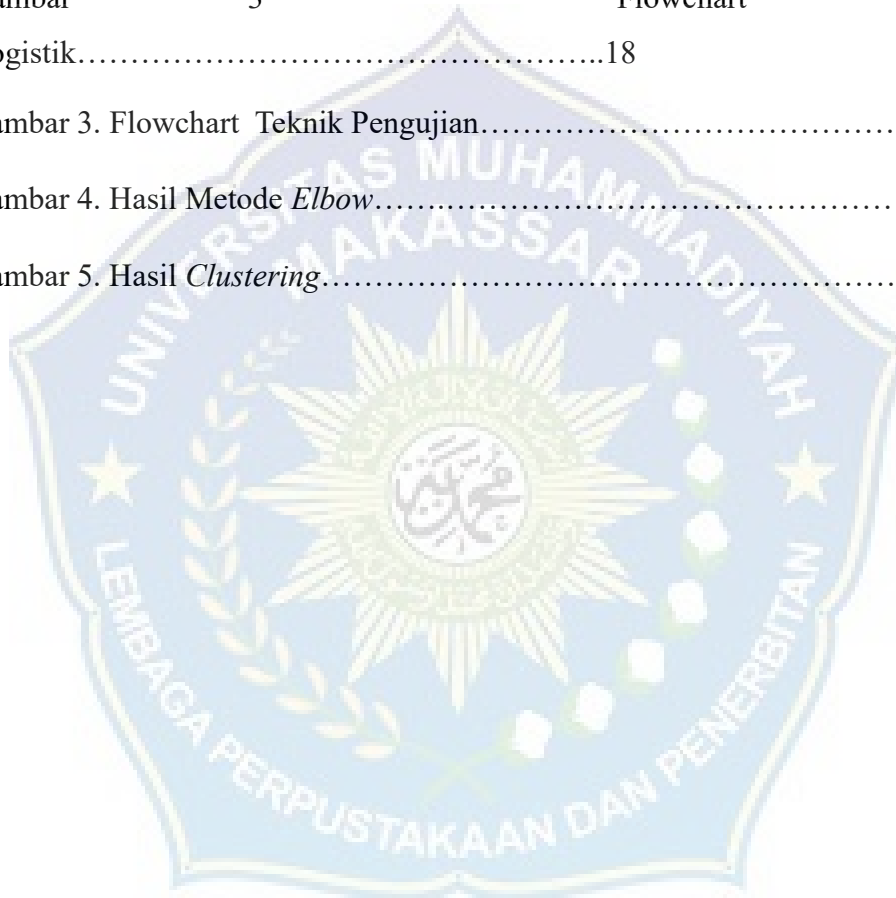
ABSTRAK	ii
KATA PENGANTAR	iv
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	ix
DAFTAR ISTILAH	x
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah.....	1
B. Rumusan Masalah	3
C. Tujuan Penelitian	3
D. Manfaat Penelitian	3
E. Ruang Lingkup penelitian	4
F. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
A. Landasan Teori.....	6
B. Penelitian Terkait	13
C. Kerangka Pikir	15
BAB III METODE PENELITIAN.....	16
A. Tempat dan Waktu Penelitian	16
B. Alat dan Bahan.....	16
C. Perancangan Sistem	16
D. Teknik Pengujian	21
E. Teknik Analisis Data.....	22
BAB IV HASIL DAN PEMBAHASAN	24
A. Pengumpulan Data	24
B. Propeccing Data.....	26
C. Implementasi Metode Elbow	27
D. Menentukan Hasil Clustering	30
E. Implementasi Model Regresi Logistik.....	33
F. Pengujian Sistem	40

BAB V KESIMPULAN DAN SARAN.....	45
A. Kesimpulan	45
B. Saran.....	45
DAFTAR PUSTAKA	45
LAMPIRAN.....	48



DAFTAR GAMBAR

Gambar	1.	Diagram	Kerangka
Pikir.....			15
Gambar	2.	Flowchart	Teknik
sistem.....			17
Gambar	3	Flowchart	Regresi
Logistik.....			18
Gambar 3. Flowchart		Teknik Pengujian.....	21
Gambar 4. Hasil Metode		<i>Elbow</i>	29
Gambar 5. Hasil		<i>Clustering</i>	31



DAFTAR TABEL

Tabel 1. Data Mentah.....	24
Tabel 2. Data Olahan.....	25
Tabel 3. Hasil <i>Systolic dan Diastolic</i>	26
Tabel 4. Hasil <i>Clustering</i>	32
Tabel 5. Hasil Akurasi.....	34
Tabel 6. Prediksi Cluster.....	38
Tabel 7. <i>Clustering</i>	39
Tabel 8. Jumlah Prediksi.....	40
Tabel 9. Laporan Klasifikasi.....	41



DAFTAR ISTILAH

Akurasi	ukuran yang digunakan untuk menilai kinerja suatu model atau sistem dalam klasifikasi atau prediksi.
<i>Clustering</i>	teknik dalam analisis data yang bertujuan untuk mengelompokkan sekumpulan objek atau data ke dalam beberapa kelompok (cluster) berdasarkan kemiripan atau kedekatan mereka satu sama lain
Data mining	alat yang memungkinkan pengguna mengakses data dalam jumlah besar dengan cepat.
Deteksi dini	proses identifikasi atau pengenalan suatu kondisi atau penyakit pada tahap awal, sebelum gejala klinis yang jelas muncul.
<i>Elbow</i>	<i>Elbow</i> adalah metode yang digunakan untuk menentukan jumlah cluster optimal dalam algoritma <i>clustering</i> , khususnya dalam metode <i>K-Means</i> .
Jantung	istilah umum yang mencakup sejumlah kondisi yang mempengaruhi jantung dan pembuluh darah. Penyakit ini seringkali disebabkan oleh penumpukan plak di arteri (aterosklerosis), yang dapat menghambat aliran darah dan memicu masalah serius seperti serangan jantung atau stroke.
<i>Machine learning</i>	salah satu cabang kecerdasan buatan yang menggunakan bahasa pemrograman untuk membuat komputer berperilaku cerdas seperti manusia.

Pasien orang yang berkonsultasi mengenai permasalahan kesehatannya untuk memperoleh pelayanan kesehatan yang diperlukan baik secara langsung maupun tidak langsung di rumah sakit.

Python bahasa pemrograman tingkat tinggi yang dirancang untuk memudahkan pengembangan perangkat lunak dengan menekankan pada keterbacaan kode dan sintaks yang bersih.

Scikit learn sebuah pustaka yang sangat berguna dan powerful untuk pembelajaran mesin menggunakan Python



BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Penyakit kronis adalah kondisi kesehatan yang berlangsung lama, biasanya lebih dari tiga bulan, dan sering memerlukan penanganan serta perawatan berkepanjangan. Penyakit ini dapat berkembang secara perlahan dan mempengaruhi kemampuan penderitanya untuk melakukan aktivitas sehari-hari. Penyakit kronis merupakan penyebab kematian utama di Indonesia. Menurut Kementerian Kesehatan Republik Indonesia (2016), proporsi kematian akibat penyakit tidak menular meningkat dari 41,7% pada tahun 1995 menjadi 49,9% pada tahun 2001 dan sebesar 59,5% pada tahun 2007. Penyebab kematian utama adalah stroke (15,4%), disusul hipertensi, diabetes, kanker, dan PPOK. Penyakit kronis biasanya berlangsung dalam jangka waktu yang lama, bisa sehari-hari, berbulan-bulan, bahkan bertahun-tahun (Sri et al., 2021).

Penyakit jantung adalah istilah umum yang mencakup sejumlah kondisi yang mempengaruhi jantung dan pembuluh darah. Penyakit ini seringkali disebabkan oleh penumpukan plak di arteri (aterosklerosis), yang dapat menghambat aliran darah dan memicu masalah serius seperti serangan jantung atau stroke. Menurut WHO dan CDC, penyakit jantung adalah penyebab utama kematian di Inggris, Amerika Serikat, Kanada dan Australia. Jumlah orang dewasa yang terdiagnosis penyakit jantung adalah 26,6 juta (11,3%) dari populasi orang dewasa (Utomo & Mesran, 2020).

Deteksi dini penyakit jantung kronis sangat penting karena memungkinkan intervensi lebih awal yang dapat mencegah perkembangan penyakit dan komplikasi serius seperti serangan jantung dan stroke. Pengobatan pada tahap awal biasanya lebih efektif, mengurangi biaya perawatan kesehatan jangka panjang, dan meningkatkan kualitas hidup pasien. Dengan deteksi dini, faktor

risiko seperti hipertensi dan kolesterol tinggi dapat dikelola lebih baik, dan pasien dapat lebih cepat mengadopsi gaya hidup sehat yang mendukung kesehatan jantung(Huda, 2022).

Deteksi dini penyakit jantung sangat penting untuk mengurangi risiko penyakit jantung. Namun, masih banyak orang yang belum tahu bagaimana cara mengurangi risiko tersebut. Penelitian ini menggunakan algoritma regresi logistik, yang merupakan metode analisis statistik untuk memprediksi kemungkinan terjadinya suatu peristiwa dengan hasil biner (dua pilihan) berdasarkan satu atau lebih variabel bebas. Metode ini sangat berguna untuk klasifikasi, seperti menentukan apakah seorang pasien berisiko terkena penyakit jantung (ya/tidak). Regresi logistik adalah teknik machine learning yang efektif dalam memprediksi probabilitas kejadian, termasuk dalam analisis data pasien untuk deteksi dini penyakit jantung. Dengan regresi logistik, faktor risiko utama dapat diidentifikasi dan individu dapat dikelompokkan berdasarkan tingkat risiko mereka, sehingga intervensi pencegahan bisa lebih tepat sasaran dan efisien. Penelitian ini bertujuan untuk meningkatkan kesadaran masyarakat tentang faktor risiko penyakit jantung dan pentingnya deteksi dini melalui teknologi machine learning (Salam et al., 2023).

Penelitian ini mengembangkan sistem analisis interaktif yang menyediakan pemodelan dan manipulasi data. Sistem ini dapat mendukung berbagai layanan kesehatan masyarakat untuk memberikan informasi kepada masyarakat mengenai faktor risiko penyakit jantung melalui deteksi dini. Setelah mengkaji literatur dari penelitian sebelumnya berjudul "GAMBARAN PENGETAHUAN BANTUAN HIDUP DASAR (BHD) PADA ANGGOTA KELUARGA YANG MEMILIKI FAKTOR RISIKO PENYAKIT JANTUNG DI DENPASAR TIMUR" (Pertiwi et al., 2021), penelitian ini berbeda karena berfungsi sebagai pendukung bagi penelitian yang sedang dilakukan. Tujuan utama penelitian ini adalah melakukan klasifikasi data secara efektif, dengan

judul "ANALISIS DETEKSI DINI PENYAKIT KRONIS DENGAN PENDEKATAN REGRESI LOGISTIK PADA DATA PASIEN".

B. Rumusan Masalah

Berdasarkan pada latar belakang masalah di atas, maka rumusan masalah pada penelitian ini yaitu “

1. Bagaimana penerapan metode clustering dan regresi logistik dalam mengklasifikasikan risiko penyakit jantung pada pasien di RSUD Haji Makassar?
2. Bagaimana tingkat akurasi nilai dari metode regresi logistik untuk melakukan klasifikasi deteksi dini penyakit jantung di RSUD Haji Makassar?

C. Tujuan Penelitian

Berdasarkan pada latar belakang masalah, rumusan masalah di atas, maka tujuan yang ingin dicapai dari penelitian ini yaitu :

1. Menerapkan metode clustering dan regresi logistik untuk mengklasifikasikan risiko penyakit jantung pada pasien di RSUD Haji Makassar.
2. untuk mendapatkan nilai akurasi yang baik penggunaan metode regresi logistic dalam mengelompokkan dan mengklasifikasikan pasien dengan risiko penyakit jantung di RSUD Haji Makassar

D. Manfaat Penelitian

Berdasarkan pada uraian latar belakang , rumusan masalah dan tujuan penelitian di atas, maka manfaat penelitian ini sebagai berikut

1. Bagi penulis, Penelitian ini memberikan kesempatan bagi penulis untuk mengembangkan keterampilan dalam analisis data dan penerapan metode

machine learning, seperti clustering dan regresi logistik, yang sangat relevan dalam bidang ilmu data dan kesehatan.

2. Bagi mahasiswa, Penelitian ini dapat menjadi referensi akademis yang berguna bagi mahasiswa lain yang sedang mengerjakan tugas akhir atau penelitian serupa di bidang kesehatan dan ilmu data.
3. Bagi universitas, Penelitian ini dapat meningkatkan reputasi universitas sebagai institusi yang mendukung penelitian berkualitas tinggi di bidang ilmu data dan kesehatan.

E. Ruang Lingkup penelitian

1. Pengambilan data pasien dari Rumah Sakit Haji Makassar sebagai sumber data utama.
2. Investigasi faktor-faktor yang berpengaruh terhadap akurasi deteksi, dengan fokus pada data pasien yang berasal dari Rumah Sakit Haji Makassar.
3. Eksplorasi kontribusi variabel klinis terhadap keberhasilan deteksi dini penyakit kronis (jantung) menggunakan Regresi Logistik, menggunakan data pasien yang diperoleh dari Rumah Sakit Haji Makassar.

F. Sistematika Penulisan

BAB 1 PENDAHULUAN

Pada bab ini, dimulai dengan menggaris bawahi pentingnya deteksi dini penyakit kronis jantung dan menjelaskan metode Regresi Logistik dalam analisis data medis. Bab ini juga merumuskan masalah dan tujuan penelitian untuk memandu studi ini, dengan penekanan pada manfaatnya bagi praktik klinis dan kemajuan ilmu pengetahuan. Selain itu, bab ini juga mencakup batasan penelitian dan tata cara penulisan untuk memberikan gambaran menyeluruh tentang struktur dan konten penelitian.

BAB II TINJAUAN PUSTAKA

Dalam bab ini, disajikan tinjauan umum tentang penyakit jantung dan metode analisis data medis yang relevan. Selanjutnya, dibahas juga konsep dan penerapan regresi logistik dalam analisis data, serta ulasan penelitian terdahulu

yang relevan dengan topik ini. Tujuan dari pembahasan ini adalah untuk memberikan landasan teoritis yang kuat bagi penelitian yang dilakukan.

BAB III METODOLOGI PENELITIAN

Pada bab ini, dijelaskan desain penelitian yang digunakan, termasuk pendekatan dan rancangan studi yang dipilih. Proses pengumpulan data dari Rumah Sakit Haji Makassar serta tahapan analisis data menggunakan regresi logistik juga diuraikan secara terperinci. Selain itu, dijabarkan pula metode validasi yang diterapkan untuk memastikan keakuratan hasil penelitian.

BAB IV HASIL DAN PEMBAHASAN

Pada bagian ini, diberikan deskripsi tentang data pasien yang terkumpul dari Rumah Sakit Haji Makassar, termasuk karakteristiknya. Selain itu, hasil analisis kinerja regresi logistik dalam mendeteksi penyakit jantung pada tahap awal juga dibahas secara rinci, memberikan gambaran tentang efektivitas metode yang digunakan dalam penelitian ini. Bab ini dilanjutkan dengan interpretasi hasil dan implikasi temuan untuk praktik klinis serta saran untuk penelitian berikutnya..

BAB V PENUTUP

Bab ini menekankan kesimpulan dari penelitian dan memberikan gambaran umum tentang temuan utama. Selain itu, disusun kembali manfaat penelitian dan disampaikan saran untuk penelitian lanjutan. Bab ini juga mengulas kembali batasan penelitian serta menyoroti kontribusi penelitian ini bagi bidang studi yang bersangkutan.

BAB II

TINJAUAN PUSTAKA

A. Landasan Teori

1. Penyakit Jantung

Menurut (Rabbani et al., 2023) Penyakit jantung merupakan sekelompok penyakit yang mempengaruhi struktur, fungsi atau pembuluh darah jantung, yang merupakan penyebab utama kematian di seluruh dunia. Di Indonesia, kejadian penyakit jantung meningkat dari 0,5 persen populasi pada tahun 2013 menjadi 1,5 persen pada tahun 2018, menurut laporan RISKEDAS. Diperkirakan kondisi ini semakin meningkat setiap tahunnya. Penyakit jantung merupakan sekelompok penyakit yang mempengaruhi struktur, fungsi atau pembuluh darah jantung, yang merupakan penyebab utama kematian. Di Indonesia, kejadian penyakit jantung meningkat dari 0,5 persen populasi pada tahun 2013 menjadi 1,5 persen pada tahun 2018, menurut laporan RISKEDAS. Keadaan ini diperkirakan semakin meningkat setiap tahunnya..

Menurut dr. Airindya Bella (Bella, 2022), terdapat beberapa jenis penyakit jantung beserta gejala yang dialami, antara lain:

- a. Penyakit Jantung Koroner (PJK): Terjadi ketika arteri yang membawa darah ke jantung mengeras dan menyempit akibat penumpukan kolesterol dan pembekuan darah. Gejala yang mungkin timbul adalah nyeri dada, kesulitan bernapas, keringat dingin, dada berdebar, dan mual.
- b. Serangan Jantung: Terjadi ketika aliran darah ke jantung terhenti sepenuhnya, menyebabkan kerusakan pada sel-sel otot jantung. Serangan jantung umumnya dipicu oleh penyakit jantung koroner. Gejala yang mungkin timbul adalah nyeri dada, sesak napas, dan keringat dingin.
- c. Aritmia: Terjadi ketika impuls listrik yang mengatur ritme jantung mengalami gangguan, menyebabkan jantung tidak berfungsi dengan

benar. Gejala yang mungkin timbul adalah jantung berdebar atau palpitasi, rasa tidak nyaman pada dada, pusing, lemas, napas pendek, dan pingsan.

- d. Gagal Jantung: Terjadi ketika jantung kehilangan kemampuan untuk memompa darah secara efektif. Gejala yang mungkin timbul adalah napas sesak ketika berbaring, napas sesak saat beraktivitas fisik, batuk dengan lendir putih, bengkak di kaki, perut, dan pergelangan kaki, pusing, lemas, sulit berkonsentrasi, detak jantung tak beraturan, dan hilangnya nafsu makan.
- e. Penyakit Katup Jantung: Terjadi ketika katup jantung tidak berfungsi secara normal, mengganggu aliran darah menuju dan dari jantung. Gejala yang mungkin timbul adalah nyeri dada saat beraktivitas atau menghirup udara dingin, lemas dan pusing, serta palpitasi atau jantung berdebar.

Penyakit jantung merupakan penyakit kondisi ketika jantung mengalami gangguan yang dapat menyerang secara tiba-tiba dan menyebabkan kematian tanpa memandang usia dan jenis kelamin (Huda, 2022). Tingginya angka kematian akibat penyakit jantung disebabkan oleh rendahnya kesadaran masyarakat untuk melakukan pemeriksaan kesehatan jantung secara rutin dan pola hidup yang tidak sehat. Penyakit jantung dapat mencakup berbagai macam kelainan, antara lain masalah pada pembuluh darah jantung, irama jantung, masalah katup jantung, atau kelainan bawaan..

Adapun yang disampaikan kepada (Ryfai et al., 2022) Penyakit jantung merupakan suatu kondisi dimana organ jantung mengalami kesulitan dalam menjalankan fungsinya membawa darah ke seluruh tubuh. Masalah jantung ada banyak jenisnya seperti masalah kardiovaskular, penyempitan atau penyumbatan pembuluh darah, masalah katup jantung, irama jantung dan lain-lain. Jantung dianggap bekerja dengan baik ketika darah dapat mengalir masuk dan keluar dengan lancar.

2. Deteksi Dini Penyakit Jantung

Pada dasarnya penyakit jantung dapat dicegah dengan beberapa faktor, termasuk pola hidup sehat. Selain itu, deteksi dini penyakit jantung juga diperlukan untuk mencegah kematian pasien. Salah satu pilihan untuk melakukan

deteksi dini adalah dengan memanfaatkan teknologi informasi. (Nurmasani & Pristyanto, 2021). Data mining merupakan salah satu bidang yang dapat digunakan sebagai alat deteksi dini penyakit jantung. Data mining dapat diartikan sebagai kombinasi berbagai disiplin ilmu yang mencakup sistem database, statistik, pembelajaran mesin, visualisasi, dan kecerdasan data.

Penyakit jantung dapat diprediksi dan dideteksi sejak dini jika semua orang mengetahui faktor risiko yang dapat menyebabkan terjadinya penyakit jantung. Faktor risiko penyakit jantung ada dua, yaitu faktor yang tidak dapat diubah dan faktor yang dapat diubah. (Kasron et al., 2022). Usia merupakan faktor risiko penyakit jantung, karena kualitas anatomi dan fungsi struktur pembuluh darah menurun seiring bertambahnya usia..

Menurut (Rahmat et al., 2023) Penyakit jantung adalah salah satu penyebab utama kematian di seluruh dunia. Oleh karena itu, deteksi dini dan klasifikasi kemungkinan penyakit jantung sangat penting dalam pencegahan dan pengobatannya.

3. Data *Mining*

Penambangan data adalah alat yang memungkinkan pengguna mengakses data dalam jumlah besar dengan cepat. Lebih khusus lagi, data mining adalah alat dan aplikasi yang menggunakan analisis data statistik. Penambangan data adalah proses penambangan atau penggalian informasi besar, yang sebelumnya tidak diketahui, namun dapat dimengerti dan berguna dari database besar. Informasi yang ditemukan melalui data mining dapat digunakan untuk membuat keputusan bisnis yang sangat penting.(Ahmad et al., 2022).

Menurut (Amalia et al., 2021) Data mining adalah bidang interdisipliner yang menggabungkan teknik dari berbagai disiplin ilmu seperti pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi. Tujuannya adalah untuk mengidentifikasi dan mengekstrak informasi berguna dari database besar, untuk mendukung pengambilan keputusan dan analisis. Pada saat yang sama, data mining juga dapat diartikan sebagai proses pencarian pola pada data dalam jumlah besar yang disimpan dalam penyimpanan. Proses ini menggunakan teknologi pengenalan pola, teknik statistik dan matematika.

Penambangan data adalah cara yang berguna untuk mengekstrak informasi berharga dari data dalam jumlah besar dengan menggunakan pengetahuan dari bidang statistik, matematika, dan pengenalan pola. Proses ini melibatkan penggalian dan identifikasi data besar untuk menemukan informasi yang berguna bagi bisnis. Data mining dapat digunakan untuk klasifikasi, prediksi dan evaluasi untuk memperoleh informasi yang berguna. Data mining membantu dalam perencanaan dan menyediakan data yang akurat untuk membuat prediksi berdasarkan tren masa lalu dan kondisi saat ini. Penggunaan data mining juga dapat membantu perusahaan mengalokasikan dana dengan lebih efisien, karena otomatisasi pengambilan keputusan dapat mengurangi biaya. (Damuri et al., 2021).

4. Pasien

Pasien adalah setiap orang yang mengalami gangguan kesehatan, baik psikis maupun organik. Pengertian pasien tidak hanya terbatas pada yang berada di rumah sakit saja, karena jika pengertian pasien hanya berlaku pada rumah sakit, maka ketika seseorang tidak lagi berada di rumah sakit, misalnya rawat jalan, maka hilanglah segala hak dan kewajibannya (Disemadi & Pardede, 2021).

Menurut (Salam et al., 2023) Pasien adalah setiap orang yang berkonsultasi mengenai permasalahan kesehatannya untuk memperoleh pelayanan kesehatan yang diperlukan baik secara langsung maupun tidak langsung di rumah sakit. Yang dimaksud dengan pasien juga mencakup perseorangan, keluarga, kelompok, dan komunitas sosial yang menerima layanan refraksi/optometri optik sesuai Peraturan Kementerian Kesehatan Republik Indonesia Nomor 41 Tahun 2015.

Menurut Peraturan Menteri Kesehatan Republik Indonesia Nomor 269/Menkes/Per/III/2008 Pasien adalah setiap orang yang melakukan konsultasi masalah kesehatan untuk memperoleh pelayanan kesehatan yang diperlukan baik secara langsung maupun tidak langsung kepada dokter atau dokter gigi (Finamore et al., 2021).

n tersebut dapat disimpulkan bahwa pasien adalah individu dengan gangguan kesehatan fisik dan mental yang mencari dan menerima pelayanan kesehatan, termasuk pelayanan optometri, dari tenaga medis. Istilah pasien mencakup orang-

orang yang dirawat di rumah sakit atau melalui konsultasi langsung atau tidak langsung..

5. *Macnine learning*

Pembelajaran mesin (ML) merupakan salah satu cabang kecerdasan buatan yang menggunakan bahasa pemrograman untuk membuat komputer berperilaku cerdas seperti manusia. ML menganalisis data yang tersedia, termasuk data dalam jumlah besar (Big Data), untuk menemukan pola pembelajaran mesin tertentu. Machine Learning memiliki tiga tipe, yaitu *Supervised Learning*, *Unsurpervised Learning*, dan *Reiforcement Learning* (Hayadi & Damanik, 2022).

Machine learning merupakan cabang kecerdasan buatan dan ilmu komputer yang berfokus pada penggunaan data dan algoritme untuk meniru cara manusia belajar dan secara bertahap dapat meningkatkan akurasi. Semakin baik algoritme pembelajaran mesin yang digunakan, semakin baik pula keputusannya.. (Faiza et al., 2022)

6. Scikit learn

Scikit-learn adalah sebuah pustaka yang sangat berguna dan powerful untuk pembelajaran mesin menggunakan Python. Pustaka ini menyediakan berbagai alat untuk pembelajaran mesin dan pemodelan statistik, termasuk berbagai metode klasifikasi, regresi, dan algoritma pengelompokan seperti *Support Vector Machines* (SVM), *Random Forests*, *Gradient Boosting*, dan *k-Means*. Scikit-learn adalah modul yang memungkinkan pemrograman machine learning berbasis python dan didistribusikan di bawah lisensi 3-Clause BSD (Nafisah Nurul Hakim, 2020).

Saat menggunakan bahasa pemrograman Python untuk mempraktikkan metode pembelajaran yang diawasi, peneliti sering kali menggunakan perpustakaan dan kerangka kerja populer seperti scikit-learn, TensorFlow, dan Keras. Alat-alat ini memungkinkan peneliti untuk dengan mudah menerapkan berbagai algoritma pembelajaran yang diawasi. Para peneliti menjelaskan langkah-langkah yang terlibat dalam proses pembelajaran, termasuk pemrosesan data, pemilihan fitur, pelatihan model, dan evaluasi kinerja. (Mestika et al., 2022).

7. Metode Elbow

Elbow adalah metode yang digunakan untuk menentukan jumlah cluster optimal dalam algoritma clustering, khususnya dalam metode K-Means. *Clustering* adalah proses pengelompokan data dengan cara observasi, atau pengelompokan kelas berdasarkan objek. Fungsi dari algoritma *clustering* yaitu mengelompokkan data berdasarkan ciri-ciri mereka dan mengukur sejauh mana kesamaan antara data dalam satu kelompok, walaupun setiap algoritma memiliki keunggulan dan kelemahan yang berbeda (Syahfitri et al., 2023).

Hasil *cluster* yang baik mempengaruhi metode yang digunakan, karakteristik, kumpulan data, struktur kepadatan data, ukuran data, dan jumlah *cluster* yang digunakan. Ada banyak metode berbeda yang digunakan untuk *clustering*, salah satu metode yang digunakan untuk menentukan jumlah cluster adalah metode elbow. Metode elbow digunakan untuk menentukan jumlah *cluster* yang baik, hasil *cluster* yang baik dapat digunakan untuk memaksimalkan hasil *cluster* (Syahfitri et al., 2023).

8. Regresi Logistik

Regresi logistik adalah teknik statistik yang digunakan untuk memodelkan hubungan antara variabel terikat biner dan satu atau lebih variabel bebas. Tujuannya adalah untuk memprediksi probabilitas suatu peristiwa menggunakan fungsi logit. Struktur regresi logistik terdiri dari tiga bagian utama: variabel dependen biner yang menunjukkan berhasil atau gagal suatu peristiwa, fungsi logit yang merupakan transformasi logaritmik dari rasio probabilitas, dan variabel independen yang digunakan untuk memprediksi nilai binomial. (Maharany Shandra Ayu Hapsary et al., 2021).

Regresi logistik adalah metode statistik yang digunakan untuk memprediksi kelas biner dalam analisis regresi. Variabel target regresi logistik bersifat dikotomis atau biner, yaitu terdiri dari dua kategori atau nilai, sering kali ditandai dengan 1 (ya, berhasil, dll.) atau 0 (tidak, gagal, dll.). Regresi logistik merupakan algoritma klasifikasi yang umum digunakan di berbagai bidang, termasuk kesehatan, untuk memprediksi kelas atau nilai biner berdasarkan variabel masukan yang diberikan. Fungsi logistik, juga dikenal sebagai fungsi sigmoid,

menyediakan kurva berbentuk S yang dapat mengambil bilangan bernilai nyata dan membuat grafiknya sebagai nilai antara 0 dan 1.(Tyasnurita & Pamungkas, 2020).

Menurut (Patunduk et al., 2022) Regresi logistik adalah suatu metode analisis statistik yang digunakan untuk menggambarkan hubungan antara suatu variabel terikat (dengan dua kategori atau lebih) dan satu atau lebih variabel bebas, yang mungkin dalam skala kontinu atau kategorikal. Jenis regresi logistik meliputi regresi logistik biner, regresi logistik multinomial, dan regresi logistik ordinal. Dalam statistik, regresi logistik, disebut juga model logistik atau model logit, digunakan untuk memprediksi probabilitas (probabilitas) suatu peristiwa dengan menggunakan data dari fungsi log kurva logistik. Regresi logistik juga dapat diartikan sebagai pendekatan untuk membangun model prediktif. Dalam regresi logistik, peneliti memprediksi variabel dependen dalam skala dikotomis.

Prosedur dalam membuat pemodelan klasifikasi dengan Regresi Logistik sebagai berikut:

1. Pengumpulan Data: Kumpulkan data yang akan digunakan dalam model, termasuk variabel dependen biner (misalnya, 0 atau 1) dan variabel independen yang relevan.
2. Praproses Data:
 - a. Pembersihan Data: Hilangkan atau perbaiki data yang hilang atau tidak konsisten.
 - b. Transformasi Data: Lakukan transformasi pada data jika diperlukan, seperti normalisasi atau standarisasi.
 - c. Pemisahan Data: Bagi data menjadi set pelatihan (training set) dan set pengujian (testing set).
3. Eksplorasi Data: Lakukan eksplorasi data untuk memahami karakteristiknya, termasuk distribusi variabel dan korelasi antara variabel.
4. Pemilihan Variabel Independen: Pilih variabel independen yang relevan untuk model berdasarkan analisis eksplorasi data dan literatur terkait.
5. Pendefinisian Model Regresi Logistik:
 - a. Model regresi logistik dapat dinyatakan dengan rumus logit:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$$

- 1). p adalah probabilitas kejadian dari variable dependen.
- 2). β_0 adalah intercept (konstanta).
- 3). $\beta_1, \beta_2, \dots, \beta_n$ adalah koefisien regresi untuk variabel independen x_1, x_2, \dots, x_n

B. Penelitian Terkait

1. “Permodelan Regresi Logistik Biner terhadap Analisis Penderita Penyakit Jantung Koroner di RSUD Dr SOEGIRI Lamongan” di lakukan oleh (Salsabya & Wulandari, 2023)

Kajian mengenai variabel-variabel yang diduga mempengaruhi penyakit jantung koroner dilakukan oleh Dr. Soegiri Lamongan. Data milik Dr. Soegiri Lamongan. Terdapat delapan variabel prognostik yaitu usia, jenis kelamin, hiperkolesterolemia, status merokok, hipertensi, diabetes, obesitas dan riwayat keluarga. Variabel respon terdiri dari dua kategori yaitu pasien jantung dengan penyakit jantung koroner dan pasien jantung tanpa penyakit jantung koroner. Hasil analisis regresi logistik biner menunjukkan bahwa jenis kelamin, hipertensi, diabetes dan merokok berpengaruh terhadap penyakit jantung koroner. . Pada pria, pasien jantung yang menderita hiperkolesterolemia, hipertensi dan diabetes mempunyai peluang 92,54% terkena penyakit jantung koroner..

2. “Deteksi Penyakit Jantung Menggunakan Metode Klasifikasi Decision Tree dan Regresi Logistik” dilakukan oleh (Bukhari et al., 2023)

Penelitian ini bertujuan membandingkan kedua metode klasifikasi tersebut untuk mendeteksi adanya penyakit jantung berdasarkan beberapa indikator. Data yang digunakan adalah data penyakit jantung yang dikeluarkan oleh University of California, Irvine (UCI) Machine Learning Repository. Berdasarkan hasil yang diperoleh, model decision tree yang terbentuk menempatkan variabel thal (tipe detak jantung pasien) sebagai simpul akar, dikarenakan nilai entropy yang paling tinggi. Model decision tree memiliki akurasi terhadap data uji sebesar 75%. Sementara itu, model regresi logistik menempatkan variabel sex, cp_3, slope_1, ca, dan thal_2 sebagai variabelvariabel yang berpengaruh nyata. Model regresi

logistik memiliki akurasi terhadap data uji sebesar 87%. Dari akurasi dari kedua model tersebut, regresi logistik lebih akurat untuk mendeteksi adanya penyakit jantung dibandingkan model decision tree.

3. “Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression” dilakukan oleh (Junifer Pangaribuan et al., 2021)

Penelitian ini menggunakan algoritma regresi logistik, yang menggunakan fungsi logistik untuk menghasilkan output berupa biner (0 atau 1) untuk klasifikasi. Setelah dilakukan eksperimen, algoritma regresi logistik menunjukkan keunggulan yang berbeda-beda dibandingkan dengan metode lain berdasarkan analisis confusion matrix. Pada data training, regresi logistik memiliki sensitivitas tertinggi yaitu 88.54% dibandingkan dengan metode lainnya. Sedangkan pada data testing, regresi logistik memiliki kekhususan tertinggi yaitu 87.50% dibandingkan dengan metode lainnya.

4. “Klasifikasi Kondisi Pasien Serangan Jantung yang Berujung Pada Aritmia Ventikular Berdasarkan Dispersi QT Menggunakan Regresi Logistik” dilakukan oleh (Patmonobo, 2022)

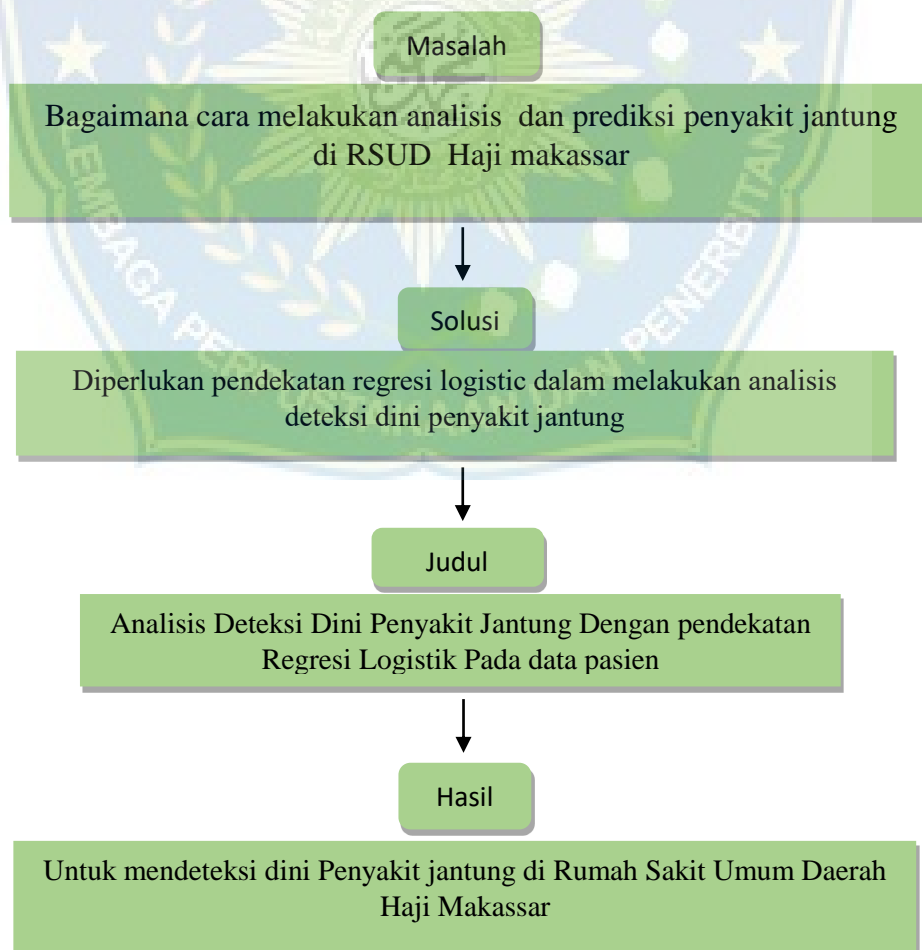
Penelitian ini menggunakan algoritma regresi logistik yang menggunakan fungsi logistik untuk menghasilkan hasil biner (0 atau 1) untuk klasifikasi. Setelah melakukan percobaan, algoritma regresi logistik menunjukkan berbagai keunggulan dibandingkan metode lain berdasarkan analisis matriks konfusi. Pada data latih, regresi logistik memiliki sensitivitas paling tinggi yaitu 88,54% dibandingkan metode lainnya. Sedangkan regresi logistik mempunyai spesifisitas tertinggi pada data uji yaitu 87,50% dibandingkan metode lainnya..

5. “Studi Komparatif Model Klasifikasi Kerentanan Penyakit Jantung Menggunakan Algoritma Machine Learning” dilakukan oleh (Lestari & Sumarlinda, 2023)

Tujuan dari penelitian ini adalah untuk melakukan analisis komparatif model klasifikasi menggunakan algoritma pembelajaran mesin yang berbeda untuk mengidentifikasi kerentanan terhadap penyakit jantung. Dataset yang digunakan diambil dari repositori pembelajaran mesin UCI dengan jumlah 300 data untuk

pelatihan dan 100 data untuk pengujian. Parameter klasifikasinya adalah umur, jenis kelamin, tekanan darah sistolik, kolesterol, denyut jantung maksimal, depresi ST, kemiringan segmen ST dan cap jantung. Model klasifikasi dibangun menggunakan algoritma Naive Bayes, K-Nearest Neighbor (KNN), Decide Tree, Random Forest, Backpropagation, Logistic Regression dan Support Vector Machine (SVM). Hasil pengukuran akurasi menunjukkan Naive Bayes (79.00%), KNN (63.00%), Decision Tree (66.00%), Random Forest (77.00%), Backpropagation (80.00%), Logistic Regression (81.00%) dan SVM (80.00%) (%)). Berdasarkan analisis perbandingan parameter akurasi, presisi, recall dan skor F1, model klasifikasi dengan algoritma regresi logistik dan propjagasi balik memberikan kinerja terbaik.

C. Kerangka Pikir



Gambar 1. Diagram Kerangka Pikir

BAB III

METODE PENELITIAN

A. Tempat dan Waktu Penelitian

1. Tempat Penelitian

Tempat penelitian merupakan lokasi dan objek yang akan digunakan pada suatu penelitian. Lokasi pengambilan data pasien pada penelitian ini dilakukan di Rumah Sakit Umum Daerah (RSUD) Haji Makassar di Pintu keluar rs haji, Jl. Dg. Ngeppe, Balang Baru, Kec. Tamalate, Kota Makassar, Sulawesi Selatan 90122.

2. Waktu Penelitian

Jadwal penelitian yang akan dilaksanakan dimulai pada bulan April sampai bulan agustus sampai semua proses pengumpulan data selesai.

B. Alat dan Bahan

1. Kebutuhan Hardware (perangkat keras)

- a. Laptop HP EliteBook x360 1030 G4
- b. RAM 16 GB

2. Kebutuhan Software (Perangkat Lunak)

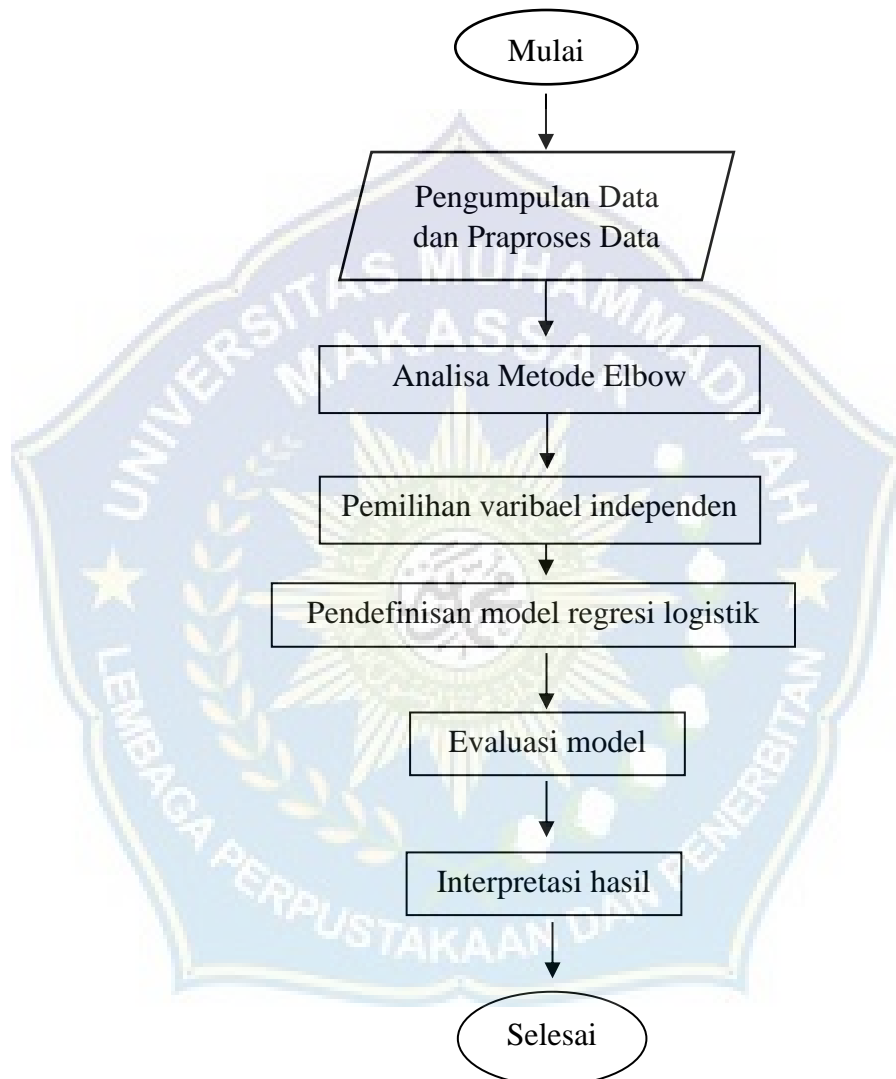
- a. Visual code
- b. Jupiter
- c. Microsoft Excel
- d. python

C. Perancangan Sistem

Perancangan sistem merupakan tahap penting dalam pengembangan sistem yang melibatkan pemodelan dan konstruksi sistem untuk memenuhi kebutuhan khusus dan mengatasi masalah yang diidentifikasi, terutama dalam konteks penelitian ini. Flowchart membantu dalam mengorganisir informasi, memvisualisasikan hubungan antarproses, dan menyediakan panduan yang jelas

bagi pengembang dan pengguna sistem. Dengan menggunakan flowchart, perancangan sistem dapat dilakukan secara terstruktur dan efisien.

1. Flowcart Perancangan Sistem



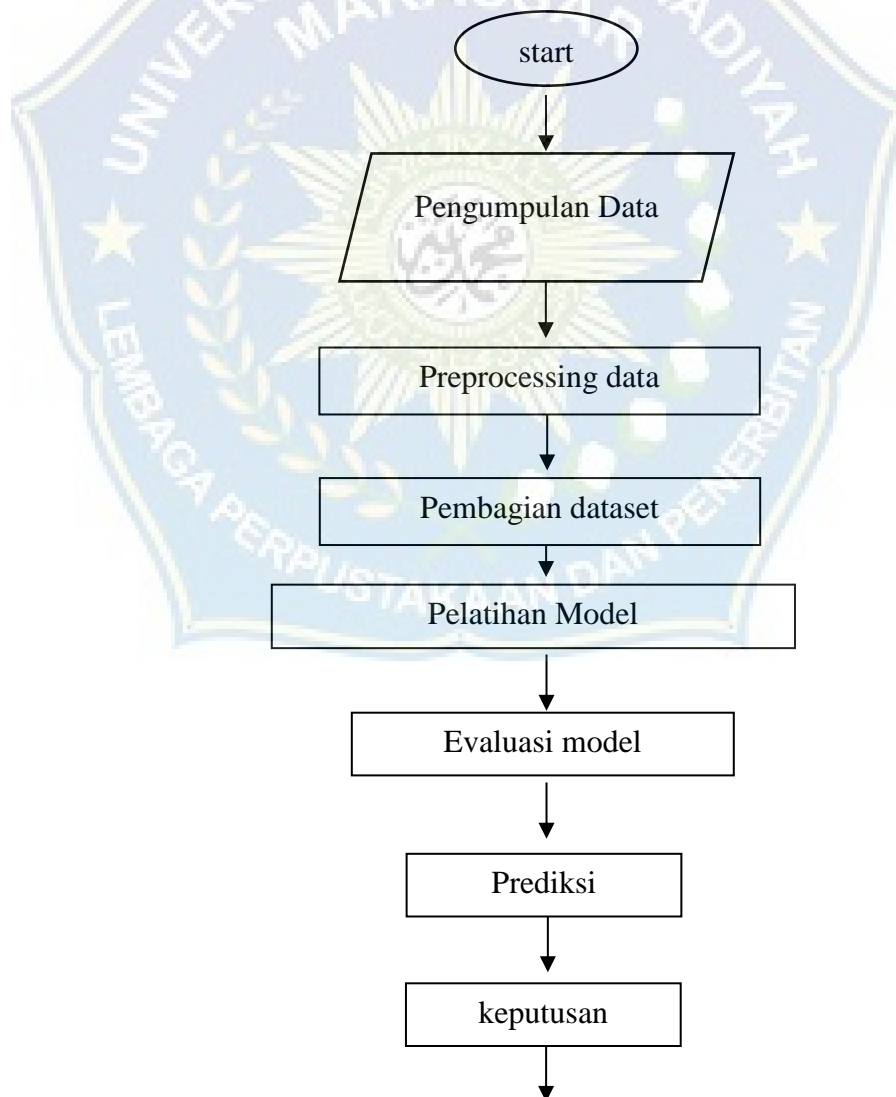
Gambar 2. Teknik Perancangan Sistem

Penjelasan gambar 2. Teknik Perancangan

1. Mulai: Memulai proses pembuatan model.
2. Pengumpulan Data: Mengumpulkan data pasien dan variabel yang diperlukan. Praproses Data: Membersihkan dan mempersiapkan data untuk analisis.

3. Melakukan clustering untuk mengelompokkan data yang lebih akurat
4. Pemilihan Variabel Independen: Memilih variabel yang paling berpengaruh terhadap penyakit jantung.
5. Pendefinisian Model Regresi Logistik: Menentukan model regresi logistik dengan variabel independen yang dipilih.
6. Evaluasi Model: Mengevaluasi kinerja model menggunakan data uji
7. Interpretasi Hasil: Menginterpretasikan koefisien model untuk memahami faktor risiko penyakit jantung.
8. Selesai: Mengakhiri proses pembuatan model dan implementasi

Adapun flowchart dari algoritma Regresi Logistik Sebagai Berikut:



end

Berikut penjelasan lebih lanjut mengenai Gambar 3. Flowchart regresi logistik untuk deteksi penyakit jantung.

1. Start

Deskripsi: Langkah awal memulai proses deteksi penyakit jantung menggunakan model regresi logistik. Pada tahap ini, kita merencanakan proses dan mengidentifikasi tujuan analisis.

2. Pengumpulan Data

Deskripsi: Data pasien dikumpulkan dari berbagai sumber, seperti rekam medis rumah sakit, survei kesehatan, atau dataset publik yang berisi fitur relevan terkait dengan penyakit jantung. Fitur yang sering digunakan termasuk usia, jenis kelamin, tekanan darah, kolesterol, dll.

Tujuan: Mengumpulkan data yang cukup untuk melatih dan menguji model.

3. Preprocessing Data

Data Cleaning: Pada langkah ini, data dibersihkan dari nilai yang hilang (missing values), outliers, atau data yang tidak relevan. Data yang tidak lengkap atau rusak perlu diatasi agar tidak memengaruhi hasil analisis.

Normalisasi/Standardisasi: Fitur numerik yang memiliki skala berbeda dinormalisasi atau distandarisasi untuk membuat model lebih stabil dan efisien. Misalnya, usia dan tekanan darah mungkin memiliki rentang yang berbeda, sehingga mereka perlu dinormalisasi.

4. Pembagian Dataset

Deskripsi: Data yang telah diproses dibagi menjadi dua subset utama:

Training Set (Data Pelatihan): Bagian data yang digunakan untuk melatih model regresi logistik.

Testing Set (Data Pengujian): Bagian data yang digunakan untuk menguji model setelah pelatihan, sehingga kita dapat mengevaluasi kinerjanya. Proporsi umum pembagian data adalah 80% untuk pelatihan dan 20% untuk pengujian, meskipun ini bisa bervariasi.

Tujuan: Memastikan bahwa model tidak hanya belajar dari data pelatihan, tetapi juga dapat menghasilkan prediksi yang baik pada data baru (data pengujian).

5. Pelatihan Model (Training)

Deskripsi: Pada tahap ini, model regresi logistik dilatih menggunakan data pelatihan. Regresi logistik bekerja dengan mencoba menemukan hubungan antara fitur input (X) dan output biner atau kategori (Y). Misalnya, model akan belajar bagaimana usia, tekanan darah, dan faktor lainnya memengaruhi kemungkinan seseorang terkena penyakit jantung.

Tujuan: Membangun model prediktif yang dapat mengidentifikasi risiko penyakit jantung berdasarkan data fitur.

6. Evaluasi Model

Deskripsi: Setelah model dilatih, kinerjanya diuji menggunakan data pengujian.

Metode evaluasi meliputi:

Akurasi: Persentase prediksi yang benar dari keseluruhan prediksi.

Precision: Jumlah prediksi positif benar dari semua prediksi positif yang dibuat.

Recall (Sensitivity): Kemampuan model untuk mendeteksi semua kasus positif yang sebenarnya.

AUC (Area Under Curve): Mengukur kemampuan model untuk memisahkan kelas dengan baik, di mana nilai AUC mendekati 1 menunjukkan kinerja yang sangat baik.

Tujuan: Mengevaluasi seberapa baik model memprediksi risiko penyakit jantung pada data yang tidak terlihat sebelumnya (data pengujian).

7. Prediksi

Deskripsi: Setelah model dievaluasi, model digunakan untuk memprediksi risiko penyakit jantung pada pasien baru. Regresi logistik menghasilkan probabilitas antara 0 dan 1. Misalnya, jika probabilitasnya lebih besar dari ambang batas tertentu (misalnya 0.5), maka pasien dikategorikan memiliki risiko tinggi terkena penyakit jantung.

Tujuan: Membuat prediksi risiko penyakit jantung berdasarkan fitur pasien.

8. Keputusan

Deskripsi: Berdasarkan hasil prediksi, pasien diklasifikasikan ke dalam kategori risiko tertentu (misalnya risiko rendah, tinggi, atau sangat tinggi). Keputusan ini bisa digunakan oleh dokter untuk mengambil langkah lebih lanjut dalam diagnosis dan pengobatan.

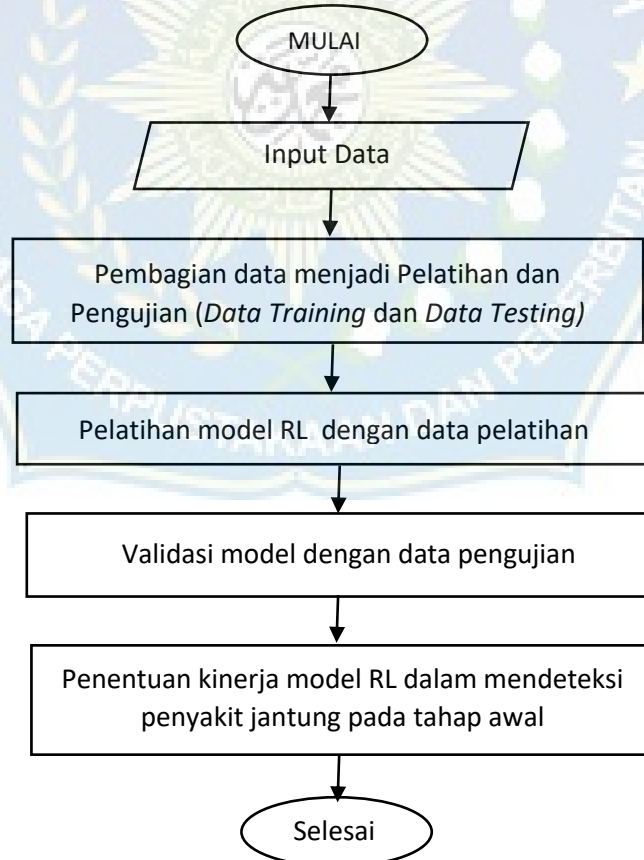
Tujuan: Memberikan hasil yang bisa digunakan untuk pengambilan keputusan klinis.

9. End

Deskripsi: Tahap akhir di mana proses deteksi selesai. Hasilnya kemudian bisa diserahkan kepada dokter atau digunakan dalam sistem untuk tindakan lanjutan seperti pengobatan atau perencanaan perawatan.

Tujuan: Menyelesaikan alur proses dan menyediakan hasil yang bermanfaat bagi pengambil keputusan.

D. Teknik Pengujian



Gambar 4. *Flowchart* Teknik pengujian

Penjelasan gambar 4. *Flowchart* Teknik Pengujian

1. Tahap Awal: Langkah pertama dalam pengujian sistem, menandakan dimulainya proses evaluasi.
2. Masukan Data: Informasi pasien yang diperlukan untuk analisis dimasukkan ke dalam sistem.
3. Pembagian Data Latih dan Uji: Informasi pasien dipisahkan menjadi dua kelompok terpisah, yaitu data latih untuk melatih model dan data uji untuk menguji performanya.
4. Pelatihan Model Regresi Logistik dengan Data Latih: Model regresi logistik dilatih menggunakan data latih untuk mengajarkan model cara mengklasifikasikan data dengan akurat.
5. Validasi Model dengan Data Uji: Model regresi logistik yang telah dilatih divalidasi menggunakan data uji untuk mengevaluasi seberapa baik model dapat memprediksi hasil yang benar pada data yang belum pernah dilihat sebelumnya.
6. Evaluasi Hasil: Setelah validasi, performa model regresi logistik dievaluasi termasuk akurasi, presisi, recall, dan metrik evaluasi lainnya untuk menilai kemampuan model dalam mendeteksi penyakit jantung pada tahap awal.
7. Penilaian Performa Model Regresi Logistik dalam Deteksi Penyakit Jantung pada Tahap Awal: Langkah terakhir adalah menentukan secara akurat kinerja model regresi logistik dalam mendeteksi penyakit jantung pada tahap awal berdasarkan hasil evaluasi menggunakan data uji.

E. Teknik Analisis Data

deteksi dini penyakit jantung menggunakan regresi logistik pada dataset pasien, proses analisis data melibatkan beberapa langkah. Langkah-langkah tersebut meliputi pengumpulan informasi pasien, pra-pemrosesan data, pembagian data menjadi data latih dan data uji, pelatihan model regresi logistik, validasi dengan data uji, evaluasi performa model, hingga analisis hasil evaluasi untuk menentukan keandalan model dalam mendeteksi penyakit jantung pada tahap awal penelitian.





BAB IV

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang diperoleh mencakup 640 data pasien penyakit jantung, yang diambil dari catatan rekam medis pasien RSUD Haji Makassar, yang

bersumber dari poli jantung dan pembuluh darah .data tersebut konsultasi dari Dr.Husnul Khatimah,S.Ked dokter koas RS haji Makassar dan Dr.ANDI MUHAMMAD REIS R SAIBY, Sp.JP sebagai dokter spesialis jantung RS Haji Makassar, sehingga mendapatkan data sebagai berikut :

Tabel 1. Data mentah

No.	No. RM	NOOPEN	Nama Pasien	JK	Ruang Akhir	Diagnosa / ICD 10
1	282954	2106220023	Mr.S	L	Poli Jantung dan Pembuluh Darah	Belum Koding
2	284021	2107070033	Mr .IR	L	Poli Jantung dan Pembuluh Darah	Belum Koding
3	284660	2107260008	Mr.R	L	Poli Jantung dan Pembuluh Darah	Z02.3-[Examination for recruitment to armed forces()]
4	280215	2109060040	Mr. RU	L	Poli Jantung dan Pembuluh Darah	Belum Koding
5	283316	2109080016	Mr.MU	L	Poli Jantung dan Pembuluh Darah	Belum Koding
.....
3518	279552	2407150037	Ny. IR	P	Poli Jantung dan Pembuluh Darah	Belum Koding
3519	202896	2407150055	Mr. AM	L	Poli Jantung dan Pembuluh Darah	I50-[Heart failure(CHF)];Z09.8-[Follow-up examination after other treatment for other conditions(kontrol pemeriksaan)];I25-[Chronic ischaemic heart disease(CAD)]
3520	303741	2407150059	Ny. E	P	Poli Jantung dan Pembuluh Darah	Belum Koding
3521	322650	2407150076	Ny. RA	P	Poli Jantung dan Pembuluh Darah	Belum Koding

3522	227838	2407150089	Mr. H	L	Poli Jantung dan Pembuluh Darah	Belum Koding
------	--------	------------	-------	---	-------	---	--------------

B. Proprocessing Data

1. Data Cleaning Dan Feature Selection: Menghapus atau memperbaiki data yang hilang, tidak valid, atau anomali. Serta Memilih fitur-fitur yang paling relevan untuk analisis, seperti usia, glukosa, ureum, kreatinin , sgot , sgpt , Tekanan Darah.

Tabel 2. Data olahan

N O	NO.R M	NOPEN	NAM A PASIE N	Pendengar an	Penciuma n	Bicara	Pernafasa n
1	285519	22071900 85	Mr. DA	Normal	Normal	Norm al	Gangguan
2	289769	22030200 59	Ny. HA	Normal	Normal	Norm al	Normal
3	289983	22030100 82	Ny. FA	Normal	Normal	Norm al	Normal
4	289940	22040100 41	Ny. NU	Normal	Normal	Norm al	Gangguan
5	152363	24011501 14	Ny. SA	Normal	Normal	Norm al	Normal
....
63 6	320603	24061401 27	Mr. TA	Normal	Normal	Norm al	Normal
63 7	320975	24062101 33	Mr. MU	Normal	Normal	Norm al	Normal
63 8	302598	24062500 47	Mr. RA	Normal	Normal	Norm al	Normal
63 9	318110	24062500 79	Mr. AB	Normal	Normal	Norm al	Normal
64 0	295262	24050300 22	Ny. EM	Normal	Normal	Norm al	Normal

2. Membagi dua hasil tekanan darah pasien yaitu : *systolic dan Diastolic*, Tekanan sistolik adalah angka pertama atau yang lebih tinggi dalam pengukuran tekanan darah. Ini mengukur tekanan di arteri ketika jantung berkontraksi dan memompa darah ke seluruh tubuh. Tekanan diastolik adalah angka kedua atau yang lebih rendah dalam pengukuran tekanan darah. Ini mengukur tekanan di arteri ketika jantung sedang dalam fase relaksasi (diastole), yaitu ketika jantung mengisi kembali dengan darah.

Tabel 3. hasil *systolic dan diastolic*

N O	NO. RM	NOPE N	NA MA PASI EN	Pendeng aran	Penciu man	Bicar a	Pernaf asan	Syst olic	Diast olic
1	28551 9	220719 0085	Mr. DA	Normal	Normal	Nor mal	Ganggu an	119	65
2	28976 9	220302 0059	Ny. HA	Normal	Normal	Nor mal	Normal	130	90
3	28998 3	220301 0082	Ny. FA	Normal	Normal	Nor mal	Normal	150	45
4	28994 0	220401 0041	Ny. NU	Normal	Normal	Nor mal	Ganggu an	125	77
5	15236 3	240115 0114	Ny. SA	Normal	Normal	Nor mal	Normal	130	60
...		
63 6	32060 3	240614 0127	Mr. TA	Normal	Normal	Nor mal	Normal	116	80
63 7	32097 5	240621 0133	Mr. MU	Normal	Normal	Nor mal	Normal	143	91
63 8	30259 8	240625 0047	Mr. RA	Normal	Normal	Nor mal	Normal	155	87
63 9	31811 0	240625 0079	Mr. AB	Normal	Normal	Nor mal	Normal	160	70
64 0	29526 2	240503 0022	Ny. EM	Normal	Normal	Nor mal	Normal	129	78

C. Implementasi Metode Elbow

Clustering adalah teknik analisis data yang digunakan untuk mengelompokkan sekumpulan objek atau data berdasarkan kemiripan mereka. Tujuan dari clustering adalah untuk memisahkan data ke dalam grup (atau kluster) yang anggotanya memiliki karakteristik yang mirip satu sama lain dan berbeda dari anggota di kluster lain.

Pada clustering ini menggunakan metode Elbow Adapun Langkah-langkahnya sebagai berikut:

1. Persiapan data

Pengumpulan Data: Kumpulkan semua data yang relevan dari pasien, termasuk atribut-atribut seperti tekanan darah, kadar kolesterol, usia, jenis kelamin, dll.

Preprocessing: Bersihkan data dari nilai yang hilang atau anomali dan lakukan normalisasi atau standarisasi data jika variabel memiliki skala yang berbeda-beda.

2. Standarisasi Fitur

Gunakan StandardScaler atau metode standarisasi lainnya untuk memastikan bahwa semua fitur memiliki rata-rata nol dan standar deviasi satu. Ini penting terutama jika skala fitur berbeda.

3. Menerapkan K-Means Clustering dengan Berbagai Jumlah Kluster

Rentang Kluster: Tentukan rentang jumlah kluster yang akan diuji, misalnya dari 1 hingga 10.

Penerapan K-Means: Terapkan algoritma K-Means untuk setiap jumlah kluster dalam rentang tersebut. Pada setiap iterasi, hitung inertia, yaitu jumlah kuadrat jarak dari setiap titik data ke centroid kluster terdekat.

4. Menghitung Inertia (Sum of Squared Distances)

Inertia adalah ukuran seberapa kompak kluster yang terbentuk. Semakin kecil inertia, semakin baik kluster yang terbentuk. Namun, seiring bertambahnya jumlah kluster, inertia selalu berkurang.

5. Memplot Inertia terhadap Jumlah Kluster

Buat plot dengan sumbu X menunjukkan jumlah kluster dan sumbu Y menunjukkan inertia. Cari titik di mana penurunan inertia mulai melambat secara signifikan, membentuk bentuk "siku" pada grafik.

6. Menentukan jumlah Kluster optimal

Identifikasi titik "siku" pada grafik. Jumlah kluster pada titik ini dianggap optimal karena menambah lebih banyak kluster setelah titik ini memberikan pengurangan inertia yang marginal.

Berikut cara clustering menggunakan metode Elbow pada pemrograman python

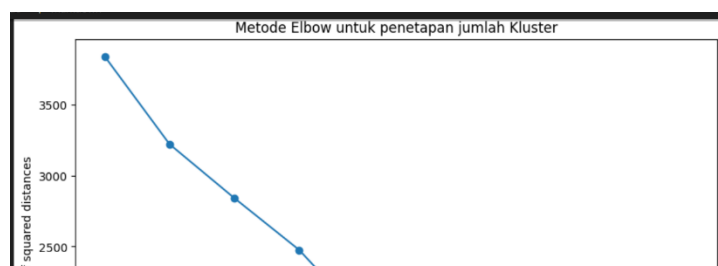
```
# penggunaan standar skalar
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data_cluster)

# Determine the optimal number of clusters using the
Elbow Method
sse = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(data_scaled)
    sse.append(kmeans.inertia_)

# Plot the SSE for each number of clusters
plt.figure(figsize=(10, 6))
plt.plot(range(1, 11), sse, marker='o')
plt.xlabel('Jumlah Kluster')
plt.ylabel('Sum of squared distances')
plt.title('Metode Elbow untuk penetapan jumlah
Kluster')
plt.show()

# Based on the Elbow Method plot, determine the optimal
number of clusters (e.g., 3)
optimal_clusters = 3
kmeans = KMeans(n_clusters=optimal_clusters,
random_state=42)
data['Kluster'] = kmeans.fit_predict(data_scaled)

# Display the cluster assignments
data[['NO', 'NAMA PASIEN', 'Cluster']]
```



Gambar 4. Hasil metode Elbow

Pada gambar tersebut, diperlihatkan hasil plot dari Metode Elbow untuk menentukan jumlah kluster yang optimal pada data pasien jantung. Pada sumbu horizontal (X), terdapat jumlah kluster yang diuji, mulai dari 1 hingga 10. Sedangkan pada sumbu vertikal (Y), terdapat nilai inerti, yang merupakan jumlah kuadrat dari jarak antara titik-titik data dengan centroid klusternya.

D. Menentukan Hasil Clustering

Dalam penetapan cluster setiap pasien penyakit jantung setelah melakukan analisis sebelumnya menggunakan metode elbow dapat diperoleh hasil cluster ada 3 jumlah clustering yang optimal. Setelah itu file hasil clustering disimpan kedalam file excel. Hasil klustering tersebut akan diinisialisasi menggunakan algoritma k-means.

Implementasi dalam inisialisasi algoritma k-means

1. Menentukan Jumlah Kluster Optimal

```
optimal_clusters = 3
```

Pada bagian ini, variabel `optimal_clusters` diatur dengan nilai 3. Nilai ini ditentukan berdasarkan analisis yang telah dilakukan sebelumnya menggunakan Metode Elbow, yang menunjukkan bahwa jumlah kluster optimal adalah 3.

2. Inisialisasi dan Pelaksanaan Algoritma K-Means

```
kmeans = KMeans(n_clusters=optimal_clusters,
random_state=42)
data['Cluster'] = kmeans.fit_predict(data_scaled)
```

- a. KMeans(n_clusters=optimal_clusters, random_state=42): Algoritma K-Means diinisialisasi dengan parameter n_clusters=3, yang artinya data akan dikelompokkan menjadi tiga kluster. Parameter random_state=42 digunakan untuk memastikan bahwa hasil klasterisasi konsisten setiap kali algoritma dijalankan.
 - b. fit_predict(data_scaled): Metode fit_predict diterapkan pada data yang telah diskalakan (data_scaled). Metode ini melakukan dua hal: menyesuaikan (fit) model K-Means dengan data dan memprediksi kluster untuk setiap sampel data. Hasilnya adalah array berisi label kluster untuk setiap titik data.
 - c. data['Cluster'] = ...: Label kluster yang dihasilkan kemudian disimpan dalam kolom baru bernama 'Cluster' di dalam dataframe data. Ini menunjukkan kluster mana yang telah ditetapkan untuk setiap titik data.
3. Menyimpan Hasil Klasterisasi ke File Excel

```
# Save the results to an Excel file
output_path = 'Hasil Kluster.xlsx'
data.to_excel(output_path, index=False)
```

- a. output_path = 'Hasil Kluster.xlsx': Menentukan jalur dan nama file output untuk menyimpan hasil klasterisasi.
- b. data.to_excel(output_path, index=False): Metode to_excel digunakan untuk menyimpan dataframe data ke dalam file Excel dengan nama 'Hasil Kluster.xlsx'. Parameter index=False digunakan untuk menghindari penyertaan indeks dataframe sebagai kolom dalam file Excel.

Setelah hasil clustering disimpan dalam file excel data yang telah diklasterisasi dengan terlebih dahulu mengidentifikasi kolom yang non-numerik, kemudian memisahkan data numerik untuk analisis lebih lanjut. Langkah selanjutnya adalah

menghitung rata-rata untuk setiap klaster berdasarkan fitur numerik yang tersedia, yang menghasilkan centroid untuk masing-masing klaster.

Menampilkan hasil clustering pada data pasien penyakit jantung

```

code: 1. markom
Glukosa          int64
Ureum           int64
Kreatinin       float64
SGOT            int64
SGPT            float64
Tekanan darah   object
Penglihatan     object
Pendengaran     object
Penciuman       object
Bicara          object
Pernafasan      object
Systolic        float64
...
Cluster
0      0.809608  28.980447  28.614972  130.299076  76.870787  0.0
1      1.400187  30.809701  29.212687  169.498881  92.531835  1.0
2      1.269286  155.642857  103.785714  135.714286  73.214286  2.0
  
```

Gambar 5. Hasil clustering

Dari hasil klasterisasi ini, kita dapat melihat bagaimana data dalam setiap klaster memiliki karakteristik yang berbeda berdasarkan fitur yang diukur (misalnya, Systolic, Diastolic).

- **Klaster 0:** Mempunyai rata-rata tekanan darah sistolik dan diastolik yang lebih rendah dibandingkan dengan klaster lain, yang mungkin menunjukkan pasien dengan tekanan darah lebih normal atau rendah.
- **Klaster 1:** Menunjukkan rata-rata tekanan darah yang lebih tinggi, mungkin menunjukkan pasien dengan hipertensi.
- **Klaster 2:** Memiliki nilai rata-rata tertentu untuk beberapa fitur, dan juga tekanan darah yang berbeda dengan klaster lainnya.

Adapun hasil dari cluster pada pasien jantung yang telah dimasukkan ke dalam file excel sebagai berikut

NO	NAMA	USI A	GLU KOS A	URE UM	KREA TININ	SG OT	SGP T	T.DA RAH	SYSTO LIC	DIAST OLIC	CLUST ER
----	------	----------	-----------------	-----------	---------------	----------	----------	-------------	--------------	---------------	-------------

1	Mr. DA	73	135	21	0.57	23	12	119/ 65	119	65	0
2	Ny. HA	51	93	18	0.76	17	15	130/ 90	130	90	0
3	Ny. FA	74	269	21	0.71	95	83	150/ 45	150	45	2
4	Ny. NU	59	124	25	0.45	54	63	125/ 77	125	77	0
5	Ny. SA		129	22	0.25	31	28	130/ 60	130	60	0
....
636	Mr. TA	25	169	25	0.73	17	13	116/ 80	116	80	0
637	Mr. MU	47	184	47	1.25	16	21	143/ 91	143	91	1
638	Mr. RA	35	183	35	0.59	28	43	155/ 87	155	87	0
639	Mr. AB	69	152	69	1.12	13	15	160. 70	160.7		1
640	Ny. EM	50	168	50	0.53	42	18	129/ 78	129	78	0

Tabel 4. Hasil clustering

E . Implementasi Model Regresi Logistik

Regresi logistik adalah teknik analisis statistik yang digunakan untuk memodelkan hubungan antara variabel independen (prediktor) atau variabel biner. Secara umum, regresi logistik digunakan ketika variabel dependen memiliki dua komponen (seperti “ya” atau “tidak”, “lulus” atau “gagal”).

Dalam melakukan permodelan regresi logistic menggunakan scikirt learn ada beberapa tahapan yang akan dilakukan sehingga mendapatkan hasil yang maksimal yaitu

1. Normalisasi data : memastikan bahwa semua fitur berkontribusi secara seimbang pada model dan tidak mendominasi model karena skala yang berbeda.
2. Pelatihan model regresi logistik : Model ini akan memprediksi apakah seorang pasien memiliki penyakit jantung berdasarkan fitur-fitur seperti usia, tekanan darah, kadar kolesterol, dll.
3. Evaluasi model : mengevaluasi kinerjanya menggunakan data pengujian.

IMPLEMENTASI

1. Normalisasi data : StandardScaler adalah alat untuk melakukan standardisasi fitur, yang mengubah nilai-nilai fitur dalam dataset sehingga memiliki rata-rata nol dan deviasi standar satu.

```
# Normalisasi data
scaler = StandardScaler()
X = scaler.fit_transform(X)
```

2. Pelatihan model regresi logistik : menginisialisasi dan melatih model regresi logistik menggunakan pustaka scikit-learn

```
# Menginisialisasi dan melatih model regresi
logistik
model = LogisticRegression(max_iter=2000) #
Meningkatkan max_iter
model.fit(X_train, y_train)
```

3. Evaluasi model : untuk mengevaluasi performa model yang telah dilatih dengan menggunakan metrik akurasi dan laporan klasifikasi.

```
# Mengevaluasi model
akurasi = accuracy_score(y_test, y_pred)
laporan = classification_report(y_test, y_pred)
```

HASIL PREDIKSI

Hasil prediksi menunjukkan bahwa model *regresi logistik* dapat memisahkan kelas dengan baik. Laporan langsung dan klasifikasi memberikan gambaran

umum tentang kinerja model. Hasil ini penting dalam mengidentifikasi pasien dengan gagal jantung yang signifikan sehingga intervensi yang tepat dapat diberikan.

Tabel 5. Hasil Akurasi

akurasi :				
0.984375				
laporan				
klasifikasi :				
	precision	recall	f1-score	support
0	0.974684	1	0.987179	77
1	1	0.979592	0.989691	49
2	1	0.5	0.666667	2
accuracy	0.984375	0.984375	0.984375	0.984375
macro avg	0.991561	0.826531	0.881179	128
weighted avg	0.984771	0.984375	0.983133	128

Penjelasan mengenai metrik-metrik dalam laporan klasifikasi:

Model Regresi logistic: memiliki akurasi sebesar 98.44%. Artinya, dari 128 sampel data pengujian, model berhasil memprediksi 126 sampel dengan benar.

Laporan Klasifikasi

menunjukkan hasil evaluasi model regresi logistik dengan metrik precision, recall, f1-score, dan support untuk tiga kelas berbeda (0, 1, dan 2). Berikut penjelasan mengenai hasil tersebut:

Akurasi

- **Akurasi: 0.984375 (98.44%)**

- Akurasi ini menunjukkan bahwa model mampu memprediksi dengan benar 98.44% dari total sampel. Dari 128 sampel, model membuat sekitar 126 prediksi yang benar. Ini adalah tingkat akurasi yang sangat tinggi, menunjukkan bahwa model regresi logistik bekerja sangat baik pada data ini.

Evaluasi Per Kelas

Kelas 0

- **Precision: 0.974684**
 - Precision sebesar 97.47% berarti dari semua prediksi yang dilakukan untuk kelas 0, sekitar 97.47% benar-benar merupakan kelas 0.
- **Recall: 1.00**
 - Recall sebesar 100% berarti model berhasil mendeteksi semua sampel yang benar-benar termasuk dalam kelas 0.
- **F1-score: 0.987179**
 - F1-score, harmonik rata-rata dari precision dan recall, sebesar 98.72% menunjukkan bahwa model bekerja sangat baik dalam mendeteksi kelas 0.
- **Support: 77**
 - Ini menunjukkan bahwa terdapat 77 sampel sebenarnya yang termasuk dalam kelas 0.

Kelas 1

- **Precision: 1.00**
 - Precision sebesar 100% berarti semua prediksi yang dilakukan untuk kelas 1 benar-benar merupakan kelas 1.
- **Recall: 0.979592**

- Recall sebesar 97.96% berarti model mampu mendeteksi hampir semua sampel yang benar-benar termasuk dalam kelas 1, dengan hanya sedikit yang terlewatkan.
- **F1-score: 0.989691**
 - F1-score sebesar 98.97% menunjukkan bahwa model memiliki kinerja yang sangat baik dalam mendeteksi kelas 1.
- **Support: 49**
 - Ini menunjukkan bahwa terdapat 49 sampel sebenarnya yang termasuk dalam kelas 1.

Kelas 2

- **Precision: 1.00**
 - Precision sebesar 100% berarti semua prediksi yang dilakukan untuk kelas 2 benar-benar merupakan kelas 2.
- **Recall: 0.50**
 - Recall sebesar 50% berarti model hanya mampu mendeteksi setengah dari sampel yang benar-benar termasuk dalam kelas 2. Ini menunjukkan adanya kelemahan model dalam mendeteksi kelas ini.
- **F1-score: 0.666667**
 - F1-score sebesar 66.67% menunjukkan bahwa meskipun precision tinggi, recall yang rendah menyebabkan F1-score yang lebih rendah untuk kelas ini.
- **Support: 2**
 - Ini menunjukkan bahwa hanya ada 2 sampel sebenarnya yang termasuk dalam kelas 2, yang mungkin menyebabkan ketidakstabilan dalam metrik evaluasi untuk kelas ini.

Rata-Rata

- **Macro avg (Rata-rata makro):**

- **Precision: 0.991561**
- **Recall: 0.826531**
- **F1-score: 0.881179**
- Ini adalah rata-rata dari precision, recall, dan F1-score dari semua kelas tanpa mempertimbangkan jumlah sampel per kelas. Rata-rata ini memberikan gambaran umum tentang kinerja model di setiap kelas, tetapi tidak mempertimbangkan ketidakseimbangan kelas.
- **Weighted avg (Rata-rata berbobot):**
 - **Precision: 0.984771**
 - **Recall: 0.984375**
 - **F1-score: 0.983133**
 - Ini adalah rata-rata dari precision, recall, dan F1-score yang mempertimbangkan jumlah sampel di setiap kelas. Ini lebih representatif dari performa keseluruhan model ketika ada ketidakseimbangan kelas.

Selanjutnya adalah Penginputan Hasil, Model diinput kedalam file excel

```
# Mengevaluasi model
akurasi = accuracy_score(y_test, y_pred)
laporan = classification_report(y_test, y_pred,
output_dict=True)

# Konversi laporan klasifikasi menjadi DataFrame
laporan_df = pd.DataFrame(laporan).transpose()

# Membuat DataFrame untuk data uji beserta prediksinya
df_test = df.loc[y_test.index].copy()
df_test['Prediksi Kluster'] = y_pred

# Ekspor hasil ke file Excel
with
pd.ExcelWriter('Hasil_Validasi_Kluster_Test.xlsx',
engine='openpyxl') as writer:
    df_test.to_excel(writer, sheet_name='Data Uji dan
Prediksi', index=False)
```

```

    laporan_df.to_excel(writer, sheet_name='Laporan
Klasifikasi')
    # Menambahkan akurasi sebagai lembar baru
    pd.DataFrame({'Akurasi':
[akurasi]}).to_excel(writer, sheet_name='Akurasi',
index=False)

print("Hasil validasi dan data uji telah diekspor ke
file Excel 'Hasil_Validasi_Kluster_Test.xlsx'.")

```

hasil prediksi telah diekspor ke file

Tabel 6. Prediksi Cluster

NO	NAMA	USI A	GLU KOS A	UREU M	KREAT ININ	SG OT	SG PT	T.DARA H	SYS TOL IC	DIA ST OLI C	CL US TER	PREDIKSI CLUSTER
1	Ny .HA	78	158	24	0.69	15	16	170/82	170	82	0	rendah
2	Ny. JU	72	113	26	0.59	23	45	149/107	149	107	1	tinggi
3	Mr.NG	69	143	24	0.7	28	15	146/108	146	108	1	tinggi
4	Mr. ME	81	165	45	1.77	51	18	207/135	207	135	1	tinggi
5	Mr. FA	56	109	52	1.65	24	29	138/105	138	105	1	tinggi
...
124	Ny. JA	67	144	10	0.37	24	38	135/85	135	85	0	rendah
125	Mr. MU	57	111	15	0.88	37	25	108/68	108	68	0	rendah
126	Ny. TA	77	139	64	1.89	33	26	159/70	159	70	1	tinggi
127	Ny. NU	72	143	18	0.87	10	14	164/62	164	62	0	rendah
128	Mr.AJ	53	95	15	1.6	23	42	176/101	176	101	1	tinggi

Model regresi logistik yang diterapkan menunjukkan skor recall dan f1 terbaik dan terakurat (98,44%), hampir sempurna untuk ketiga kelas. Hal ini menunjukkan bahwa model sangat efektif dalam membedakan kelas-kelas pada kumpulan data ini, dengan kesalahan prediksi yang sangat rendah. Model ini dapat dianggap andal untuk digunakan dalam konteks yang mirip dengan data eksperimen, seperti prediksi berdasarkan kondisi yang tersedia..

F. Pengujian Sistem

1. Pengujian Sistem Clustering Elbow

Pengujian sistem clustering Elbow dengan melihat hasil kluster pada kolom kluster dimana nilai 0 itu mempresentasikan resiko rendah penyakit jantung sedangkan nilai 1 mempresentasikan resiko tinggi penyakit jantung, dan nilai 2 mempresentasikan resiko sangat tinggi penyakit jantung.

Tabel 7. clustering

NO	NAMA	USI A	GLU KOS A	URE UM	KREAT ININ	SG OT	SGPT	T.DA RAH	SYSTO LIC	DIAST OLIC	CLUST ER
1	Mr. DA	73	135	21	0.57	23	12	119/ 65	119	65	0
2	Ny. HA	51	93	18	0.76	17	15	130/ 90	130	90	0
3	Ny. FA	74	269	21	0.71	95	83	150/ 45	150	45	2
4	Ny. NU	59	124	25	0.45	54	63	125/ 77	125	77	0
5	Ny. SA		129	22	0.25	31	28	130/ 60	130	60	0
636	Mr. TA	25	169	25	0.73	17	13	116/ 80	116	80	0
637	Mr. MU	47	184	47	1.25	16	21	143/ 91	143	91	1
638	Mr. RA	35	183	35	0.59	28	43	155/ 87	155	87	0
639	Mr. AB	69	152	69	1.12	13	15	160. 70	160.7		1
640	Ny. EM	50	168	50	0.53	42	18	129/ 78	129	78	0

Selanjutnya mencari jumlah data kluster beresiko tinggi dan beresiko rendah dari total 640 data pasien deteksi dini.

Tabel 8. Jumlah prediksi

NO	NAMA	USI A	GLU KOS A	UREU M	KREAT ININ	SG OT	SG PT	T.DARA H	SYS TOL IC	DIA ST OLI C	CLUSTE R
1	Mr. DA	73	135	21	0.57	23	12	119/65	119	65	0
2	Ny. HA	51	93	18	0.76	17	15	130/90	130	90	0
3	Ny. FA	74	269	21	0.71	95	83	150/45	150	45	2
4	Ny. NU	59	124	25	0.45	54	63	125/77	125	77	0
5	Ny. SA		129	22	0.25	31	28	130/60	130	60	0
.....
636	Mr. TA	25	169	25	0.73	17	13	116/80	116	80	0
637	Mr. MU	47	184	47	1.25	16	21	143/91	143	91	1
638	Mr. RA	35	183	35	0.59	28	43	155/87	155	87	0
639	Mr. AB	69	152	69	1.12	13	15	160.70	160		1
640	Ny. EM	50	168	50	0.53	42	18	129/78	129	78	0
jumlah predikis 0											358
jumlah prediksi 1											268
jumlah prediksi 2											14

Berdasarkan data prediksi klaster, dapat disimpulkan bahwa mayoritas individu, sebanyak 358 orang, masuk ke dalam Klaster 0, yang menunjukkan kondisi kesehatan yang relatif stabil atau normal. Sebanyak 268 individu masuk ke dalam Klaster 1, yang kemungkinan memiliki peningkatan risiko kesehatan dibandingkan dengan Klaster 0. Sementara itu, hanya 14 individu yang termasuk dalam Klaster 2, Tahap Analisis ini berguna dalam mengidentifikasi terkait pasien penyakit jantung. Hal ini telah sesuai dengan yang dijelaskan oleh narasumber yaitu dr.Husnul Khatimah S, ked Selaku dokter Rumah Sakit Umum Daerah (RSUD) Haji Makassar mengatakan bahwa dapat dilihat dari hasil pemeriksaan laboratorium yang menunjukkan bahwa pasien yang mengalami peningkatan kadar ureum dan kreatinin, mengindikasikan adanya

kerusakan fungsi ginjal.serta hipertensi yang dapat menyebabkan penyumbatan *aterosklerosis dan glukosa* dapat menyebabkan diabetes yang juga dapat meningkatkan aterosklerosis. Selain itu ,kadar SGOT dan SGPT yang tinggi mengindikasikan gangguan pada hati dan jantung pasien.

2. Pengujian sistem klasifikasi dan Prediksi Regresi Logistik

Pengujian Sistem klasifikasi dan prediksi *Regresi Logistik* Dilakukan dengan cara manual Menghitung regresi logistik secara manual memerlukan beberapa langkah matematis yang melibatkan penggunaan fungsi logistik

Menghitung manual dalam metode regresi logistic

Asumsi Baru:

- Dari 77 pasien dengan prediksi kategori rendah, 76 pasien sebenarnya rendah (True Negative, TN) dan 1 pasien sebenarnya tinggi (False Negative, FN).
- Dari 49 pasien dengan prediksi kategori tinggi, 48 pasien sebenarnya tinggi (True Positive, TP) dan 1 pasien sebenarnya sangat tinggi (False Positive, FP).
- Dari 2 pasien dengan prediksi kategori sangat tinggi, 2 pasien benar-benar berada di kategori ini (True Positive, TP).

Matriks Kontingensi Berdasarkan Asumsi:

Tabel 9.laporan klasifikasi

	Prediksi Rendah (0)	Prediksi Tinggi (1)	Prediksi Sangat Tinggi (2)
Sebenarnya Rendah (0)	76 (TN)	0	0

Sebenarnya Tinggi	1 (FN)	48 (TP)	1 (FP)
(1)			
Sebenarnya Sangat Tinggi (2)	0	0	2 (TP)

Menghitung Metode Evaluasi

1. **Akurasi:**

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{76+48+2}{128} = \frac{126}{128} \approx 0.9844 \text{ atau } 98.44\%$$

2. Presisi (precision) untuk kategori tinggi

$$\text{presisi} = \frac{TP}{TP+FP} = \frac{48}{48+1} = \frac{48}{49} \approx 0.98 \text{ atau } 98\%$$

3. Recall (sensivity) untuk kategori tinggi

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{48}{48+1} = \frac{48}{49} \approx 0.98 \text{ atau } 98\%$$

4. F1-score untuk kategori tinggi

$$\text{F1-Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{0.98 \times 0.98}{0.98 + 0.98} \approx 0.98 \text{ atau } 98\%$$

5. Presisi dan Recall untuk kategori sangat tinggi:

Karena hanya ada dua prediksi dan semuanya benar:

$$\text{Presisi}_{\text{sangat tinggi}} = \text{Recall}_{\text{sangat tinggi}} = \frac{2}{2} = 1.0 \text{ atau } 100\%$$

Dengan akurasi 99.22%, model ini menunjukkan kinerja yang sangat baik:

- **Akurasi:** 98.44%
- **Presisi:** 98%
- **Recall:** 98.0%
- **F1-Score:** 98%

Hasil prediksi model klasifikasi menunjukkan akurasi yang tinggi dengan 126 prediksi benar dari seluruh peserta uji. Ini menghasilkan persentase prediksi

akurat sebesar 98.44%. Data tersebut menunjukkan bahwa model klasifikasi yang digunakan mampu memprediksi hasil paling akurat dengan akurasi hampir 100%. Bagian prediksi dibagi menjadi tiga bagian yaitu “Rendah”, “Tinggi”, dan “Sangat Tinggi” yang menampilkan hasil evaluasi masing-masing peserta sesuai kriteria yang telah ditetapkan. Keberhasilan ini menunjukkan potensi besar model klasifikasi untuk digunakan dalam berbagai situasi yang memerlukan akurasi prediksi tinggi



BAB V

KESIMPULAN DAN SARAN

A. Kesimpulan

1. Penelitian ini menggunakan metode clustering dan regresi logistik untuk menganalisis data 640 pasien jantung di RSUD Haji Makassar. Dari hasil analisis cluster menggunakan metode Elbow dan K-Means ditemukan tiga cluster optimal yang dapat mengklasifikasikan data pasien. Kelompok-kelompok ini memiliki karakteristik yang berbeda-beda: kelompok 0 mencakup pasien dengan risiko penyakit jantung rendah, kelompok 1 mencakup pasien dengan risiko penyakit jantung tinggi. Dan kelompok 2 mencakup pasien dengan risiko penyakit jantung sangat tinggi
2. Dalam model regresi logistik menunjukkan tingkat akurasi yang tinggi yaitu 98.44% . hal ini menunjukkan bahwa metode regresi logistik sangat akurat dalam mendeteksi penyakit jantung

B. Saran

Dari kesimpulan yang telah dijelaskan Adapun saran dari penelitian ini yaitu:

Penelitian ini dapat diperluas dengan menggunakan metode machine learning lain, seperti Random Forest atau Neural Networks, untuk membandingkan hasil prediksi dan akurasi yang lebih baik dalam klasifikasi risiko penyakit jantung.

DAFTAR PUSTAKA

- Ahmad, I., Samsugi, S., & Irawan, Y. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Teknoinfo*, 16(1), 46. <http://portaldata.org/index.php/portaldata/article/view/107>
- Amalia, F. S., Setiawansyah, S., & ... (2021). Analisis Data Penjualan Handphone Dan Elektronik Menggunakan Algoritma Apriori (Studi Kasus: Cv Rey Gasendra). ... *Journal of Telematics and ...*, 2(1), 1–6. <https://ejurnal.teknokrat.ac.id/index.php/telefortech/article/view/1810>
- Bukhari, F., Nurdiati, S., Najib, M. K., & Amalia, R. N. (2023). Deteksi Penyakit Jantung Menggunakan Metode Klasifikasi Decision Tree dan Regresi Logistik. *Sains, Aplikasi, Komputasi Dan Teknologi Informasi*, 5(1), 41. <https://doi.org/10.30872/jsakti.v5i1.10780>
- Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 219. <https://doi.org/10.30865/jurikom.v8i6.3655>
- Disemadi, H. S., & Pardede, T. S. (2021). Problematika Pemberian Sanksi Terhadap Penolakan Vaksinasi Covid-19: Suatu Kajian Perspektif HAM. *Jurnal Supremasi*, 11(2020), 107–119. <https://doi.org/10.35457/supremasi.v11i2.1442>
- Faiza, I. M., Gunawan, G., & Andriani, W. (2022). Tinjauan Pustaka Sistematis: Penerapan Metode Machine Learning untuk Deteksi Bencana Banjir. *Jurnal Minfo Polgan*, 11(2), 59–63. <https://doi.org/10.33395/jmp.v11i2.11657>
- Finamore, P. da S., Kós, R. S., Corrêa, J. C. F., D, Collange Grecco, L. A., De Freitas, T. B., Satie, J., Bagne, E., Oliveira, C. S. C. S., De Souza, D. R., Rezende, F. L., Duarte, N. de A. C. A. C. D. A. C., Grecco, L. A. C. A. C., Oliveira, C. S. C. S., Batista, K. G., Lopes, P. de O. B., Serradilha, S. M., Souza, G. A. F. de, Bella, G. P., ... Dodson, J. (2021). No Title. *Journal of Chemical Information and Modeling*, 53(February), 2021. <https://doi.org/10.1080/09638288.2019.1595750><https://doi.org/10.1080/17518423.2017.1368728><http://dx.doi.org/10.1080/17518423.2017.1368728><https://doi.org/10.1016/j.ridd.2020.103766><https://doi.org/10.1080/02640414.2019.1689076>
- Hayadi, B. H., & Damanik, A. R. (2022). Pendekatan Machine Learning Menggunakan Algoritma C4 . 5 Berbasis Pso Dalam Analisa. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(3).

- Huda, N. (2022). Analisis Dan Prediksi Penyakit Jantung Menggunakan Algoritma K-Means Clustering Pada Rumah Sakit Umum Daerah Prabumulih. ... *Algoritma K-Means Clustering Pada ...*, 197–207. [http://eprints.binadarma.ac.id/12822/%0Ahttp://eprints.binadarma.ac.id/12822/1/jurnal trisca.pdf](http://eprints.binadarma.ac.id/12822/%0Ahttp://eprints.binadarma.ac.id/12822/1/jurnal_trisca.pdf)
- Junifer Pangaribuan, J., Tanjaya, H., & Kenichi, K. (2021). Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression. *Journal Information System Development (ISD)*, 06(02), 1–10.
- Kasron, Susilawati, Susilawati, & Subrono, W. (2022). Penguatan Peran Kader Kesehatan dalam Deteksi Dini Resiko Penyakit Jantung di Kelurahan Tambakreja Kabupaten Cilacap. *Jurnal Abdi Mercusuar*, 2(1), 73–79.
- Lestari, W., & Sumarlinda, S. (2023). Studi Komparatif Model Klasifikasi Kerentanan Penyakit Jantung Menggunakan Algoritma Machine Learning. *SATIN - Sains Dan Teknologi Informasi*, 9(1), 107–115. <https://doi.org/10.33372/stn.v9i1.918>
- Maharany Shandra Ayu Hapsary, Sawitri Subiyanto, & Hana Sugiastu Firdaus. (2021). Analisis Prediksi Perubahan Penggunaan Lahan dengan Pendekatan Artificial Neural Network dan Regresi Logistik di Kota Balikpapan. *Jurnal Geodesi Undip*, 10(2), 88–97.
- Mestika, J. C., Selan, M. O., & Qadafi, M. I. (2022). Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python. *BIIKMA : Buletin Ilmiah Ilmu Komputer Dan Multimedia*, 99(99), 216–219.
- Nafisah Nurul Hakim. (2020). Implementasi Machine Learning pada Sistem Prediksi Kejadian dan Lokasi Patah Rel Kereta Api di Indonesia. *Jurnal Sistem Cerdas*, 3(1), 25–35. <https://doi.org/10.37396/jsc.v3i1.58>
- Nurmasani, A., & Pristyanto, Y. (2021). Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class. *Pseudocode*, 8(1), 21–26. <https://doi.org/10.33369/pseudocode.8.1.21-26>
- Patmonobo, Y. W. (2022). Klasifikasi Kondisi Pasien Serangan Jantung yang berujung Pada Aritmia Ventrikular Berdasarkan Dispersi QT Menggunakan Regresi Logistik. *Jurnal UIN Syarif Hidayatullah*.
- Patunduk, K. W., Hidayat, R., Avini, A., Sumarni, S., Pratiwi, A., & Harbianti, H. (2022). PEMODELAN PASIEN COVID-19 DI KOTA PALOPO DENGAN REGRESI LOGISTIK (Studi Perbandingan Regresi Logistik dan Analisis Survival). *Proximal: Jurnal Penelitian Matematika Dan Pendidikan Matematika*, 5(2), 260–269. <https://doi.org/10.30605/proximal.v5i2.1963>

- Rabbani, H. H. A., Jamaluddin, A., & Solehudin, A. (2023). Sistem Pakar Diagnosa Awal Penyakit Jantung Menggunakan Metode Forward Chaining Dan Certainty Factor Berbasis Website. *INFOTECH Journal*, 9(2), 442–451. <https://doi.org/10.31949/infotech.v9i2.6401>
- Rahmat, A., Syafiih, M., & Faid, M. (2023). Implementasi Klasifikasi Potensi Penyakit Jantung Dengan Menggunakan Metode C4.5 Berbasis Website (Studi Kasus Kaggle.Com). *INFOTECH Journal*, 9(2), 393–400. <https://doi.org/10.31949/infotech.v9i2.6295>
- Ryfa, D. A., Hidayat, N., & Santoso, E. (2022). Klasifikasi Tingkat Resiko Serangan Penyakit Jantung menggunakan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(10), 4701–4707. <http://j-ptiik.ub.ac.id>
- Salam, Ayati, S., Muhamad, J., & Sulistyawati, A. (2023). Gambaran Pengetahuan Pasien Tentang Perawatan Kacamata Di Optik King Rancaekek Kabupaten Bandung 2023. *Jurnal Alat Optik*, 2018, 1–13.
- Salsabya, A. A., & Wulandari, S. P. (2023). Permodelan Regresi Logistik Biner terhadap Analisis Penderita Penyakit Jantung Koroner Di RSUD Dr SOEGIRI Lamongan. *Jurnal Sains Dan Seni ITS*, 12(1). <https://doi.org/10.12962/j23373520.v12i1.107942>
- Syahfitri, N., Budianita, E., Nazir, A., & Afrianty, I. (2023). Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(3), 1668–1675. <https://doi.org/10.30865/klik.v4i3.1525>
- Tyasnurita, R., & Pamungkas, A. Y. M. (2020). Deteksi Diabetik Retinopati menggunakan Regresi Logistik. *ILKOM Jurnal Ilmiah*, 12(2), 130–135. <https://doi.org/10.33096/ilkom.v12i2.578.130-135>



Lampiran 1. Source code

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Memuat Data
data = pd.read_excel('DATA JANTUNG RM(3).xlsx')
data.info()
data.head()
# Kolom Patokan Kluster
columns = ['Glukosa', 'Ureum', 'Kreatinin', 'SGOT',
           'SGPT', 'Tekanan darah']

# Pembersihan kolom tekanan darah dan bagi dua ke dalam
dua kolom
data['Tekanan darah'] = data['Tekanan
darah'].str.replace(r'\D', '/')
data[['Systolic', 'Diastolic']] = data['Tekanan
darah'].str.split('/', expand=True).astype(float)
columns.extend(['Systolic', 'Diastolic'])

# masukkan Systolic dan Diastolic ke dalam kolom
inputan
columns = ['Ureum', 'Kreatinin', 'SGOT', 'SGPT',
           'Systolic', 'Diastolic']
data_cluster = data[columns]

# penanganan nilai kosong
data_cluster = data_cluster.fillna(data_cluster.mean())

# penggunaan standar skalar
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data_cluster)

# Determine the optimal number of clusters using the
Elbow Method
sse = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(data_scaled)
```

```

    sse.append(kmeans.inertia_)

# Plot the SSE for each number of clusters
plt.figure(figsize=(10, 6))
plt.plot(range(1, 11), sse, marker='o')
plt.xlabel('Jumlah Kluster')
plt.ylabel('Sum of squared distances')
plt.title('Metode Elbow untuk penetapan jumlah
Kluster')
plt.show()

# Based on the Elbow Method plot, determine the optimal
number of clusters (e.g., 3)
optimal_clusters = 3
kmeans = KMeans(n_clusters=optimal_clusters,
random_state=42)
data['Cluster'] = kmeans.fit_predict(data_scaled)

# Display the cluster assignments
data[['NO', 'NAMA PASIEN', 'Cluster']]
# Based on the Elbow Method plot, determine the optimal
number of clusters (e.g., 3)
optimal_clusters = 3
kmeans = KMeans(n_clusters=optimal_clusters,
random_state=42)
data['Cluster'] = kmeans.fit_predict(data_scaled)

# Save the results to an Excel file
output_path = 'Hasil Kluster.xlsx'
data.to_excel(output_path, index=False)
# Load the data
file_path = 'Hasil Kluster.xlsx'
df = pd.read_excel(file_path)

# Print data types to identify non-numeric columns
print(df.dtypes)

# Convert columns to numeric if needed (example column
names)
# df['Column1'] = pd.to_numeric(df['Column1'],
errors='coerce')
# df['Column2'] = pd.to_numeric(df['Column2'],
errors='coerce')

# Filter out non-numeric columns

```

```

numeric_data = df.select_dtypes(include=[np.number])

# Perform groupby and calculate the mean
centroids = numeric_data.groupby(df['Cluster']).mean()

# Display the centroids
print(centroids)

# Load the data
file_path = 'Hasil Kluster.xlsx'
df = pd.read_excel(file_path, sheet_name='Sheet1')
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report,
accuracy_score
from sklearn.preprocessing import StandardScaler
import openpyxl

# Memastikan bahwa kolom memiliki tipe data numerik
df['Usia'] = pd.to_numeric(df['Usia'], errors='coerce')
df['Ureum'] = pd.to_numeric(df['Ureum'],
errors='coerce')
df['Kreatinin'] = pd.to_numeric(df['Kreatinin'],
errors='coerce')
df['SGOT'] = pd.to_numeric(df['SGOT'], errors='coerce')
df['SGPT'] = pd.to_numeric(df['SGPT'], errors='coerce')
df['Glukosa'] = pd.to_numeric(df['Glukosa'],
errors='coerce')
df['Systolic'] = pd.to_numeric(df['Systolic'],
errors='coerce')
df['Diastolic'] = pd.to_numeric(df['Diastolic'],
errors='coerce')
df['Cluster'] = pd.to_numeric(df['Cluster'],
errors='coerce')

# Mendefinisikan fitur input dan target
fitur = ['Usia', 'Ureum', 'Kreatinin', 'SGOT', 'SGPT',
'Glukosa', 'Systolic', 'Diastolic']
target = 'Cluster'

# Mengekstraksi fitur input dan variabel target
X = df[fitur]

```

```

y = df[target]

# Menangani nilai yang hilang (jika ada) dengan
mengisinya dengan rata-rata kolom
X = X.fillna(X.mean())

# Normalisasi data
scaler = StandardScaler()
X = scaler.fit_transform(X)

# Membagi data menjadi set pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(X,
y, test_size=0.2, random_state=42)

# Menginisialisasi dan melatih model regresi logistik
model = LogisticRegression(max_iter=2000) #
Meningkatkan max_iter
model.fit(X_train, y_train)
# Membuat prediksi pada set pengujian
y_pred = model.predict(X_test)

# Mengevaluasi model
akurasi = accuracy_score(y_test, y_pred)
laporan = classification_report(y_test, y_pred)

print("Akurasi:", akurasi)
print("Laporan Klasifikasi:\n", laporan)

# Mengevaluasi model
akurasi = accuracy_score(y_test, y_pred)
laporan = classification_report(y_test, y_pred,
output_dict=True)

# Konversi laporan klasifikasi menjadi DataFrame
laporan_df = pd.DataFrame(laporan).transpose()

# Membuat DataFrame untuk data uji beserta prediksinya
df_test = df.loc[y_test.index].copy()
df_test['Prediksi Kluster'] = y_pred

# Mapping prediksi kluster ke label deskriptif
label_mapping = {0: 'rendah', 1: 'tinggi', 2: 'sangat
tinggi'}

```

```

df_test['Prediksi Kluster'] = df_test['Prediksi
Kluster'].map(label_mapping)

# Ekspor hasil ke file Excel
with pd.ExcelWriter('Hasil_Validasi_Kluster_Test.xlsx',
engine='openpyxl') as writer:
    df_test.to_excel(writer, sheet_name='Data Uji dan
Prediksi', index=False)
    laporan_df.to_excel(writer, sheet_name='Laporan
Klasifikasi')
    # Menambahkan akurasi sebagai lembar baru
    pd.DataFrame({'Akurasi':
[akurasi]}).to_excel(writer, sheet_name='Akurasi',
index=False)

print("Hasil validasi dan data uji telah diekspor ke
file Excel 'Hasil_Validasi_Kluster_Test.xlsx'.")

```



Lampiran 2.Surat Permohonan Penelitian

SURAT PERMOHONAN PENELITIAN

ZSR

Hal : Permohonan Surat Penelitian

Kepada Yth,

Ketua Program Studi Informatika

Di

Tempat

Assalamu Alaikum Warahmatullahi Wabarakatuh

Sehubungan dengan akan dilaksanakannya Penelitian yang akan dilaksanakan di Rumah Sakit Umum Daerah (RSUD) Haji Makassar oleh mahasiswa Fakultas Teknik Program Studi Informatika. Adapun Mahasiswa yang bersangkutan adalah sebagai berikut :

No	Nama	Nim
1	Iklashul Amal	105841105220

Maka dengan ini kami memohon dibuatkan surat pengantar atau pengajuan Penelitian pada Instansi dibawah ini.

Judul Skripsi : ANALISIS DETEKSI DINI PENYAKIT KRONIS DENGAN PENDEKATAN REGRESI LOGISTIK MENGGUNAKAN DATA PASIEN

Nama Instansi : Rumah Sakit Umum Daerah (RSUD) Haji Makassar

Alamat : Jl. Dg. Ngeppe No. 14 Tamalate Makassar

Demikian surat permohonan kami ajukan, atas dukungan dan kerjasamanya kami haturkan terima kasih.

Billahi Fii Sabilihq, Fastabiqul Khairat

Walaikumsalam Warahmatullahi Wabarakatuh

Makassar, 19 Zulkaidah 1445 H

28 Mei 2024 M

Pemohon

Iklashul Amal
105841105220

Lampiran 3 surat pengantar LP3M Unismuh Makassar



UNIVERSITAS MUHAMMADIYAH MAKASSAR
FAKULTAS TEKNIK
PROGRAM STUDI INFORMATIKA



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Nomor : 382/05/C.4-VI/VI/45/2024
Lamp. : -
Hal : **Pengantar Penelitian**

Makassar, 06 Dzulhijjah 1445 H
13 Juni 2024 M

Kepada yang Terhormat,
Ketua LP3M Unismuh Makassar
Di -
Tempat

Assalamu 'Alaikum Warahmatullahi Wabarakatuh

Dengan Rahmat Allah SWT, Semoga aktivitas kita bernilai ibadah di Sisi - Nya. Dalam rangka penyelesaian Tugas Sarjana / Tugas Akhir Mahasiswa pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar dengan judul: **Analisis Deteksi Dini Penyakit Kronis dengan Pendekatan Regresi Logistik Menggunakan Data Pasien**”, Sehubungan hal tersebut, maka kami meminta kesedian Bapak/Ibu agar kiranya berkenan membantu perihal surat tersebut. Bersama ini kami sampaikan mahasiswa(i):

No.	Stambuk	Nama
1.	105 84 11052 20	Ikhlasul Amal

Demikian surat kami atas perhatian dan kerja samanya kami haturkan banyak terima kasih.

Jazakumullah Khaeran Katsiran

Wassalamu 'Alaikum warahmatullah Wabarakatuh



Muhyiddin A. M. Havat, S.Kom., MT.
NBM. 150 4577

Tembusan: Kepada Yang Terhormat,
1 *Dekan Fakultas Teknik*
2 *Arsip*

Lampiran 4 .surat permohonan izin penelitian PTSP sul-sel



MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH MAKASSAR

LEMBAGA PENELITIAN PENGEMBANGAN DAN PENGABDIAN KEPADA MASYARAKAT

Jl. Sultan Alauddin No. 259 Telp.866972 Fax (0411)865588 Makassar 90221 e-mail :lp3m@unismuh.ac.id

nomor : 4497/05/C.4-VIII/VI/1445/2024

24 June 2024 M

jumlah : 1 (satu) Rangkap Proposal

18 Dzulhijjah 1445

jenis : Permohonan Izin Penelitian

Kepada Yth,

Bapak Gubernur Prov. Sul-Sel

Cq. Kepala Dinas Penanaman Modal & PTSP Provinsi Sulawesi Selatan

di -

Makassar

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Berdasarkan surat Dekan Teknik Universitas Muhammadiyah Makassar, nomor: 382/05/C.4-VI/VI/45/2024 tanggal 13 Juni 2024, menerangkan bahwa mahasiswa tersebut di bawah ini :

Nama : IKHLASHUL AMAL

No. Stambuk : 10584 1105220

Fakultas : Teknik

Jurusan : Informatika

Pekerjaan : Mahasiswa

Bermaksud melaksanakan penelitian/pengumpulan data dalam rangka penulisan Skripsi dengan judul :

"ANALISIS DETEKSI DINI PENYAKIT KRONIS DENGAN PENDEKATAN REGRESI LOGISTIK MENGGUNAKAN DATA PASIEN"

Yang akan dilaksanakan dari tanggal 27 Juni 2024 s/d 27 Agustus 2024.

Sehubungan dengan maksud di atas, kiranya Mahasiswa tersebut diberikan izin untuk melakukan penelitian sesuai ketentuan yang berlaku.

Demikian, atas perhatian dan kerjasamanya diucapkan Jazakumullahu khaeran

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Ketua LP3M,



M. Arief Muhsin, M.Pd.

NBM 1127761

06-24

Lampiran 5. Surat izin penelitian RSUD Haji makassar



**PEMERINTAH PROVINSI SULAWESI SELATAN
DINAS PENANAMAN MODAL DAN PELAYANAN TERPADU SATU PINTU**

Jl. Bougenville No.5 Telp. (0411) 441077 Fax. (0411) 448936
Website : <http://simap-new.sulselprov.go.id> Email : ptsp@sulselprov.go.id
Makassar 90231

Nomor : **16157/S.01/PTSP/2024** Kepada Yth.
Lampiran : - Direktur RSUD Haji Kota Makassar
Perihal : **Izin penelitian**

di-
Tempat

Berdasarkan surat Ketua LP3M UNISMUH Makassar Nomor : 4497/05/C.4-VIII/VI/1445/2024 tanggal 24 Juni 2024 perihal tersebut diatas, mahasiswa/peneliti dibawah ini:

Nama : **IKHLASHUL AMAL**
Nomor Pokok : 105841105220
Program Studi : informatika
Pekerjaan/Lembaga : Mahasiswa (S1)
Alamat : Jl. Slt Alauddin No. 259 Makassar

Bermaksud untuk melakukan penelitian di daerah/kantor saudara dalam rangka menyusun SKRIPSI, dengan judul :

" ANALISIS DETEKSI DINI PENYAKIT KRONIS DENGAN PENDEKATAN REGRESI LOGISTIK PADA DATA PASIEN "

Yang akan dilaksanakan dari : Tgl. **27 Juni s/d 27 Agustus 2024**

Sehubungan dengan hal tersebut diatas, pada prinsipnya kami **menyetujui** kegiatan dimaksud dengan ketentuan yang tertera di belakang surat izin penelitian.

Demikian Surat Keterangan ini diberikan agar dipergunakan sebagaimana mestinya.

Diterbitkan di Makassar
Pada Tanggal 25 Juni 2024

**KEPALA DINAS PENANAMAN MODAL DAN PELAYANAN TERPADU
SATU PINTU PROVINSI SULAWESI SELATAN**



ASRUL SANI, S.H., M.Si.
Pangkat : PEMBINA TINGKAT
I Nip : 19750321 200312 1 008

Tembusan Yth

1. Ketua LP3M UNISMUH Makassar di Makassar;
2. *Pertinggal.*

KETENTUAN PEMEGANG IZIN PENELITIAN :

1. Sebelum dan sesudah melaksanakan kegiatan, kepada yang bersangkutan melapor kepada Bupati/Walikota C q. Kepala Bappelitbangda Prov. Sulsel, apabila kegiatan dilaksanakan di Kab/Kota
2. Penelitian tidak menyimpang dari izin yang diberikan
3. Mentaati semua peraturan perundang-undangan yang berlaku dan mengindahkan adat istiadat setempat
4. Menyerahkan 1 (satu) eksamplar hardcopy dan softcopy kepada Gubernur Sulsel. Cq. Kepala Badan Perencanaan Pembangunan Penelitian dan Pengembangan Daerah Prov. Sulsel
5. Surat izin akan dicabut kembali dan dinyatakan tidak berlaku apabila ternyata pemegang surat izin ini tidak mentaati ketentuan tersebut diatas.

REGISTRASI ONLINE IZIN PENELITIAN DI WEBSITE :
<https://izin-penelitian.sulselprov.go.id>





MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH MAKASSAR
UPT PERPUSTAKAAN DAN PENERBITAN

Alamat kantor: Jl.Sultan Alauddin NO.259 Makassar 90221 Tlp.(0411) 866972,881593, Fax.(0411) 865588

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

SURAT KETERANGAN BEBAS PLAGIAT

UPT Perpustakaan dan Penerbitan Universitas Muhammadiyah Makassar,
Menerangkan bahwa mahasiswa yang tersebut namanya di bawah ini:

Nama : Ikhlashul Amal
Nim : 105841105220
Program Studi : Teknik Informatika

Dengan nilai:

No	Bab	Nilai	Ambang Batas
1	Bab 1	10 %	10 %
2	Bab 2	21 %	25 %
3	Bab 3	10 %	10 %
4	Bab 4	0 %	10 %
5	Bab 5	0 %	5 %

Dinyatakan telah lulus cek plagiat yang diadakan oleh UPT- Perpustakaan dan Penerbitan Universitas Muhammadiyah Makassar Menggunakan Aplikasi Turnitin.

Demikian surat keterangan ini diberikan kepada yang bersangkutan untuk dipergunakan
seperlunya.

Makassar, 16 Agustus 2024
Mengetahui,

Kepala UPT- Perpustakaan dan Penerbitan,



Jl. Sultan Alauddin no 259 makassar 90222
Telepon (0411)866972,881 593, fax (0411)865 588
Website: www.library.unismuh.ac.id
E-mail : perpustakaan@unismuh.ac.id

BAB I ikhlashul amal 105841105220

ORIGINALITY REPORT

10%

SIMILARITY INDEX

10%

INTERNET SOURCES

5%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

1	www.kemkes.go.id Internet Source	3%
2	repository.unama.ac.id Internet Source	2%
3	core.ac.uk Internet Source	2%
4	eprints2.undip.ac.id Internet Source	2%
5	rama.unimal.ac.id Internet Source	2%

Exclude quotes Off

Exclude matches < 2%

Exclude bibliography Off

BAB II ikhlashul amal 105841105220

ORIGINALITY REPORT

21 %

SIMILARITY INDEX

19 %

INTERNET SOURCES

8 %

PUBLICATIONS

3 %

STUDENT PAPERS

PRIMARY SOURCES

1	Krisna Wansi Patunduk, Rahmat Hidayat, Avini Avini, Sumarni Sumarni, Ananda Pratiwi, Harbianti Harbianti. "PEMODELAN PASIEN COVID-19 DI KOTA PALOPO DENGAN REGRESI LOGISTIK (Studi Perbandingan Regresi Logistik dan Analisis Survival)", Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika, 2022 Publication	4 %
2	jurnal.stmik-amik-riau.ac.id Internet Source	4 %
3	ejurnal.its.ac.id Internet Source	3 %
4	www.honestdocs.id Internet Source	2 %
5	pdfs.semanticscholar.org Internet Source	2 %
6	eprints.binadarma.ac.id Internet Source	2 %

repository.unhas.ac.id

BAB III ikhlashul amal 105841105220

ORIGINALITY REPORT

10%

SIMILARITY INDEX

8%

INTERNET SOURCES

4%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1	eprints.uns.ac.id Internet Source	3%
2	karinov.co.id Internet Source	3%
3	www.scilit.net Internet Source	2%
4	Davito Rasendriya Rizqullah Putra, Reyhan Adi Saputra. "IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK (CNN) UNTUK MENDETEKSI PENGGUNAAN MASKER PADA GAMBAR", Jurnal Informatika dan Teknik Elektro Terapan, 2023 Publication	2%

Exclude quotes Off

Exclude bibliography Off

Exclude matches < 2%

BAB IV ikhlashul amal 105841105220

ORIGINALITY REPORT

0%

SIMILARITY INDEX

2%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

Exclude quotes Off

Exclude bibliography Off

Exclude matches < 2%



BAB V ikhlashul amal 105841105220

ORIGINALITY REPORT

0%

SIMILARITY INDEX

0%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

Exclude quotes Off

Exclude bibliography Off

Exclude matches < 2%

