



# Klasifikasi Penyakit Jantung Menggunakan Algoritma Analisis Diskriminan Linier

Ibnu Rashad<sup>a,\*</sup>, R Rizal Isnanto<sup>b</sup>, Catur Edi Widodo<sup>c</sup>

<sup>a</sup>Magister Sistem Informasi, Sekolah Pascasarjana Universitas Diponegoro

<sup>b</sup>Teknik Sistem Komputer, Fakultas Teknik Universitas Diponegoro

<sup>c</sup>Departemen Fisika, Fakultas Sains dan Matematika Universitas Diponegoro

*Naskah Diterima :01 Juni 2022; Diterima Publikasi :31 Mei 2023*

*DOI: 10.21456/vol13iss1pp29-36*

## Abstract

The number of documented deaf people continues to increase. To communicate with each other, the deaf use sign language. The problem arises when Muslims with hearing impairment or deafness need to recite the Al-Quran. Muslims recite Al-Quran using their voice, but for the deaf, there are no available means to do the reciting. Thus, learning hijaiyah letters using finger gestures is considered important to develop. In this study, we use the recognition of hijaiyah letters based on pictures as the learning model. The real-time-based recognition then uses the learning model. This study uses 4 CNN pre-trained models, namely MnetV2, VGG16, ResNet50, and Xception. The learning process shows that MnetV2, VGG16, and Xception reach the accuracy limit of 99.85% in 2, 3, and 11 s, respectively. Meanwhile, ResNet50 cannot reach the accuracy limit after processing 100 s. ResNet50 achieves 82.12% accuracy. The testing process shows that MnetV2, VGG16, and ResNet50 achieve 100% precision, recall, f1-score, and accuracy. ResNet50 shows figures 81.55%, 86.04%, 82.04%, and 82.58%. The implementing process of the learning outcomes from MnetV2 shows good performance for recognizing finger shapes in real-time.

**Keywords** : Coronary Heart Disease; Machine Learning; *Linear Discriminant Analysis*

## Abstrak

Penyakit jantung Koroner Ischaemic menjadi penyebab kematian utama global. Penyakit ini hanya dapat didiagnosis dengan berkonsultasi langsung ke dokter ahli jantung dengan biaya yang cenderung tidak sedikit. Di lain pihak membutuhkan suatu sistem untuk mendeteksi penyakit jantung untuk seorang pasien dengan biaya yang minim, dari pengembangan suatu teknologi, khususnya di bidang kecerdasan buatan. Terdapat metode yang dapat mendeteksi penyakit jantung secara otomatis yaitu menggunakan machine learning diantaranya adalah *Linear Discriminant Analysis*. Pada penelitian ini algoritma *Linear Discriminant Analysis* diterapkan untuk mengklasifikasikan penyakit jantung. Dataset yang digunakan berasal dari *UCI machine learning repository*. Penelitian ini melakukan dua kondisi percobaan yakni, penyakit jantung pada penderita atau tidak, klasifikasi penyakit jantung lain ditentukan berdasarkan dari 5 tahapan. Hasil yang didapatkan membuktikan klasifikasi LDA dengan menggunakan 2 kelas lebih baik dibandingkan 5 kelas. Penerapan algoritma LDA dalam pengklasifikasian penyakit jantung dengan 2 lebel digunakan sebagai tujuan atau hasil. Dari hasil yang didapat, ketepatan nilai yang didapat 0,82, nilai pengulangan 0,81, nilai f1 memiliki nilai 0,81, dengan ketepatan 81,22%. hasil dari penerapan algoritma LDA untuk mengkategorikan penyakit jantung dalam 5 tahapan dapat dijadikan hasil akhir atau tujuan. Berdasarkan ini, akurasi nilai adalah 0,56, nilai recall pengulangan adalah 0,59, nilai f1 adalah 0,56, dengan ketepatan presisi mencapai 59,38%.

**Kata Kunci** : Penyakit Jantung Koroner; Machine learning; *Linear Discriminant Analysis*

## 1. Pendahuluan

LDA dipilih karena merupakan salah satu metode yang dipakai untuk pengenalan pola pada perhitungan statistika dengan cara menemukan proyeksi linear dari data yang akan memaksimalkan jarak antar-kelas dan meminimalkan jarak data yang memiliki kesamaan. Metode ini akan menjadi dasar klasifikasi dari data yang ada (Hana *et al.*, 2020). Semakin berkembangnya teknologi dan informasi pada bidang kecerdasan buatan, diperkenalkanlah teknik machine learning membantu meningkatkan pendeteksian

secara otomatis. Metode yang sering digunakan dalam machine learning adalah Salah satunya prediksi *Linear Discriminant Analysis* dan *Linear Discriminant Quadratic* sehingga pendeteksian penyakit jantung dapat di deteksi sedini mungkin. Dengan bantuan sistem pendeteksi, kemungkinan kesalahan diagnosa yang dilakukan oleh para ahli dapat dihindari, dan data medis dapat diperiksa sedini mungkin. Dan data medis dapat diperiksa dalam rentang waktu yang singkat dan lebih tepat dan akurat. (Ghaderyan *et al.*, 2014).

\*) Penulis korespondensi: [ibnurashad04@gmail.com](mailto:ibnurashad04@gmail.com)

Penyakit jantung merupakan suatu istilah umum dari semua gangguan yang penyebabnya berasal dari hati. Penyakit Cardiovascular (CDV) menjadi penyebab utama kasus kematian di dunia dibandingkan penyakit lainnya. Penyakit jantung adalah penyakit tidak menular yang menyebabkan kasus kematian paling besar. Berdasarkan laporan World Health Organization (WHO) tahun 2000, yang dapat membuat seseorang tidak dapat mengenali gejala awalnya sehingga awal pengobatan menjadi sulit.

Tujuan dari penelitian ini adalah untuk membangun Sistem Informasi berbasis Website menggunakan metode LDA (*Linear Discriminant Analysis*) dalam memprediksi penyakit jantung. Sistem ini dapat memberikan informasi secara akurat dan optimal untuk pengenalan penyakit jantung berdasarkan kategori dari penyakit jantung berisiko tinggi (menderita penyakit jantung) hingga penyakit jantung berkategori risiko rendah (tidak menderita penyakit jantung). Selain itu dapat meminimalkan adanya kesalahan pada tahap diagnosis awal atau peringatan awal yang sesuai dengan kebutuhan dan tepat sasaran sesuai dengan penyakit yang di derita pasien, sehingga bisa mengurangi kematian manusia dari penyakit jantung.

Masih kurangnya kesadaran akan pola hidup sehat dan kurangnya informasi mengenai penyakit jantung koroner proses untuk mendeteksi penyakit jantung dapat dilakukan secara manual, yaitu dengan berkonsultasi langsung dengan ahli penyakit jantung dan membawa beberapa hasil tes di laboratorium dilanjutkan lagi dengan berkonsultasi Kembali ke ahli penyakit jantung. Langkah ini relatif mengeluarkan banyak biaya. Dengan resiko kematian yang besar, maka dibutuhkan suatu sistem yang dapat mendeteksi penyakit jantung pada pasien yang akurat dan berbiaya rendah (Mai *et al.*, 2018).

Berbagai studi prediksi telah dilakukan untuk mengetahui tingkat keakurasian yang dihasilkan oleh tiap – tiap metode dan faktor – faktor yang mempengaruhi. Penelitian terkait penerapan metode prediksi analisis diskriminan linear dilakukan oleh Sulistio. Metode analisis diskriminan linier digunakan dalam modul prediksi yang memisahkan kelompok siswa ke dalam kelas masing – masing (lulus tepat waktu atau terlambat) dan selanjutnya menunjukkan hasil untuk pengguna. Data yang digunakan untuk input adalah nilai mata pelajaran masing – masing siswa. Mata pelajaran yang digunakan pada aplikasi dipilih oleh kepala departemen dengan dipilih 22 mata pelajaran, untuk keperluan evaluasi digunakan metode validasi silang dengan rata – rata keakurasian 97% (Sulistio, 2017).

Penelitian sebelumnya tentang akurasi metode analisis diskriminan linier yang disempurnakan oleh penelitian Budiman yang berjudul Deteksi Jenis Autisme Pada Anak Usia Dini Menggunakan Analisis Diskriminan Linier. Penelitiannya menggunakan data

latih dari 75% data yang digunakan dan 25% sisanya sebagai data uji untuk menguji model sistem yang dihasilkan dari penggunaan, jenis autisme pada anak usia dini. Dari Hasil penelitiannya metode analisis diskriminan linier menunjukkan akurasi 88%, tetapi akurasinya dapat berubah ketika data latih ditambahkan ke dataset (Budiman dan Afirianto, 2017).

Metode prediksi yang sering digunakan antara lain algoritma Logistic Regression (LR), *Linear Discriminant Analysis* (LDA), Quadratic Discriminant Analysis, dan K- Nearest Neighbor (KNN). Pada penelitian ini, rumusan masalah dalam mengembangkan metode klasifikasi yaitu menggunakan LDA (*Linear Discriminant Analysis*) untuk kasus penyakit jantung pada penerapannya diarahkan pada kategori risiko tinggi (bagi penderita) hingga kategori risiko rendah (bukan). Metode analisis diskriminan linier digunakan untuk pengenalan pola dan pembelajaran mesin dalam menemukan kombinasi linier yang mencirikan atau memisahkan dua objek atau lebih atau peristiwa. Perpaduan yang diperoleh dapat dipakai sebagai klasifikasi linier atau biasa digunakan sebagai proses pengurangan dimensi sebelum mengklasifikasikan, sehingga dapat dihasilkan nilai yang akurasinya lebih akurat dan presisi.

## 2. Kerangka Teori

### 2.1. Klasifikasi Statistik

Klasifikasi statistik atau sering juga disebut sebagai klasifikasi adalah sebuah cara atau metode untuk mengidentifikasi kemana sebuah objek data dikelompokkan. Sebagai contoh adalah klasifikasi email ke dalam kategori spam atau bukan spam, seseorang yang termasuk ke dalam kategori terdeteksi memiliki penyakit jantung atau tidak, dan sebagainya. Klasifikasi juga bisa digunakan untuk membangun sebuah model prediktif untuk memprediksi objek baru. Klasifikasi statistik mempelajari fungsi antara karakteristik sesuatu (misalnya variabel *independent*) dan keanggotannya (misalnya variabel *output*) melalui suatu proses pembelajaran ‘terawasi’ (*supervised learning process*) pada kedua jenis variabel (input dan output) disajikan ke algoritma (Verma *et al.*, 2021). Dalam klasifikasi statistik, biasanya tiap observasi dari individual atau objek baru akan dianalisa ke dalam suatu set variabel yang bisa dijelaskan (*explanatory variabel*) atau fitur. Nilai dari fitur ini bisa berupa kategori (seperti golongan darah A, B, AB, atau O), urutan/ordinal (seperti besar, kecil, atau sedang) atau berupa angka (bilangan bulat atau real). Ada juga classifier yang bekerja dengan melakukan perbandingan similaritas observasi lalu dan sekarang (Verma *et al.*, 2021).

Metodologi dua langkah yang paling umum dari prediksi jenis *classification* adalah model pelatihan (*training*) dan model pengujian (*testing*). Pada model

pelatihan sekumpulan data masukan termasuk berbagai label kelas yang aktual digunakan. Setelah suatu ‘model’ dilatih, model tersebut di tes terhadap sampel data yang tersisa untuk penilaian akurasi dan pada akhirnya diimplementasikan untuk penggunaan riil yang digunakan untuk memprediksi kelas – kelas dari data baru (label kelas tidak diketahui) (Caulkins *et al.*, 2006).

Dalam berbagai masalah klasifikasi yang dijadikan sumber utama untuk memprediksi akurasi adalah *confusion matrix* (matrik klasifikasi). Tabel 1 di bawah ini adalah *confusion matrix* pada klasifikasi dengan dua kelas.

Tabel 1. Tabel Prediksi Kelas

Prediksi kelas	Kebenaran kelas	
	Positif	Negatif
Positif	Kebenaran positif (TP)	Negatif palsu (FP)
Negatif	Positif palsu (FN)	Kebenaran negatif (TN)

*Confusion matrix* tersebut terdiri dari *true positive* (TP), *false negative* (FN), *false positive* (FP) dan *true negative* (TN). Pada TP merupakan data positif yang terdeteksi benar. FN merupakan kebalikan dari true positive, sehingga data positif, namun terdeteksi sebagai data negatif. Untuk FP sendiri, merupakan data negatif namun terdeteksi sebagai data positif, sedangkan TP merupakan data negatif yang terdeteksi dengan benar.

## 2.2. Algoritma Klasifikasi

Algoritma klasifikasi merupakan algoritma yang banyak dipakai untuk keperluan klasifikasi pada bidang ilmu machine learning. Algoritma klasifikasi biasanya digunakan sebagai metode perhitungan untuk melakukan prediksi data baru dengan karakteristik data lama. Algoritma klasifikasi juga biasanya merujuk kepada rumus atau fungsi matematika yang juga mempunyai kemampuan prediksi dengan mengklasifikasi data baru ke dalam kelas tertentu. Beberapa contoh Algoritma klasifikasi adalah *linear classifier*, *support vector machine*, *quadratic classifier*, *kernel estimation*, *boosting*, *decision tree*, *neural networks*, dan *learning vector quantization*. Semua algoritma klasifikasi mempunyai keuntungan dan kerugian masing-masing, tergantung dari macam-macam faktor seperti tipe data set, jumlah kelas dan lain-lain (Dutta *et al.*, 2020).

## 2.3. Klasifikasi Linear

Pada klasifikasi statistik, target atau objektif dari klasifikasi adalah menentukan atau mengidentifikasi ke mana sebuah objek dimasukkan dalam beberapa grup atau kelas. Hal ini dilakukan dengan menggunakan kombinasi linear dari beberapa karakteristik atau fitur sebuah objek yang diolah menjadi fungsi prediksi linear untuk setiap

kelas yang ada. Setelah itu objek baru akan masuk ke kelas tertentu berdasarkan hasil skor dengan cara memasukkan karakteristik objek baru ke dalam fungsi predictor. Nilai fungsi yang didapat dari proses klasifikasi ini akan menentukan kelas dari objek setelah diinterpretasi. Salah satu keunggulan dari klasifikasi linear adalah kecepatannya dalam melakukan klasifikasi terutama untuk objek yang hanya mempunyai dua kelas. Klasifikasi linear bisa dicari dengan menggunakan model generatif dan model diskriminatif, yang membedakan model generatif dan diskriminatif adalah prosedur pelatihan yang digunakan untuk mencari fungsi diskriminan dan bagaimana hasil skor dari fungsi diskriminan tersebut diinterpretasikan. Contoh algoritma yang digunakan dalam model diskriminatif adalah metode *Linear Discriminant Analysis* (Trajdos dan Burduk, 2021).

## 2.4. Linear Discriminant Analysis (LDA)

*Linear Discriminant Analysis*, disingkat LDA adalah generalisasi diskriminan linear fisher, yaitu sebuah metode yang digunakan dalam ilmu statistika, pengenalan pola dan pembelajaran mesin untuk mencari kombinasi linear fitur yang menjadi ciri atau yang memisahkan dua atau beberapa objek atau peristiwa. Kombinasi yang diperoleh dapat dijadikan pengklasifikasi linear atau biasanya digunakan untuk proses reduksi dimensionalitas sebelum pengklasifikasian. Tujuan dari *Linear Discriminant Analysis* (LDA) adalah mengklasifikasikan objek ke dalam beberapa kelas berdasarkan ciri yang menggambarkan objek tersebut, di dalam *Linear Discriminant Analysis* objek mempunyai dua variabel yaitu variabel kelas/terikat (dependent variabel) dan variabel atribut/bebas (independent variabel), variabel terikat mempunyai ikatan dengan variabel bebas yang menggambarkan variabel tersebut (Wei dan Tan, 2021).

Variabel bebas nantinya akan digunakan untuk menentukan kombinasi linear dari objek tersebut. LDA bekerja dengan menggunakan analisa matriks penyebaran yang bertujuan menemukan proyeksi optimal sehingga dapat memproyeksikan data input pada ruang dengan dimensi yang lebih kecil pada semua pola (*pattern*) dapat dipisahkan semaksimal mungkin. Dalam LDA variabel dependent merupakan kelas dari objek yang biasanya mempunyai nilai nominal/nama dari kelas dan variabel independent merupakan fitur yang menggambarkan objek tersebut biasanya bernilai skalar. Sebelum melakukan prediksi, LDA membutuhkan fase pelatihan untuk menentukan fungsi diskriminan. Fase pelatihan ini membutuhkan objek yang telah terklasifikasi beserta sejumlah variabel ciri/independen variabel. Menurut Wei dan Tan (2021, langkah-langkah dalam menghitung tingkat pelatihan dalam LDA dimulai dengan mengelompokkan pelatihan data ke dalam matriks sejumlah kelasnya dinotasikan dengan  $X_i$ , dengan  $i$  merupakan jumlah kelas. Setelah mengelompokkan

pelatihan data dalam matriks, berikutnya perlu menghitung rata-rata matriks dari tiap kelas ( $\mu_i$ ). Kemudian menghitung nilai rata-rata global dari keseluruhan data ( $\mu$ ) matriks, setelah didapatkan nilai rata-rata global berikutnya perlu menghitung data terkoreksi rerata (mean corrected data/X 0) dengan mengurangkan setiap nilai pada X dengan nilai rata-rata global ( $\mu$ ).

### 2.5. Correlation Coefficient

*Correlation coefficient* adalah nilai yang digunakan untuk mengukur korelasi antara dua pasang set data. Nilai dari korelasi ini bisa dikategorikan menjadi korelasi positif, korelasi negatif dan tanpa korelasi (Wei dan Tan, 2021). Korelasi positif terjadi jika satu variabel naik, maka variabel lainnya juga akan ikut naik, dan begitu pun sebaliknya. Korelasi negatif terjadi jika satu variabel naik, maka variabel lainnya turun, dan begitupun sebaliknya. Sedangkan keadaan tidak adanya korelasi terjadi jika kenaikan variabel tidak mempengaruhi variabel lainnya sama sekali. dalam matriks sejumlah kelasnya dinotasikan dengan  $X_i$ , dengan  $i$  merupakan jumlah kelas. Setelah mengelompokkan pelatihan data dalam matriks, berikutnya perlu menghitung rata-rata matriks dari tiap kelas ( $\mu_i$ ). Kemudian menghitung nilai rata-rata global dari keseluruhan data ( $\mu$ ) matriks, setelah didapatkan nilai rata-rata global berikutnya perlu menghitung data terkoreksi rerata (mean corrected data/X 0) dengan mengurangkan setiap nilai pada X dengan nilai rata-rata global ( $\mu$ ).

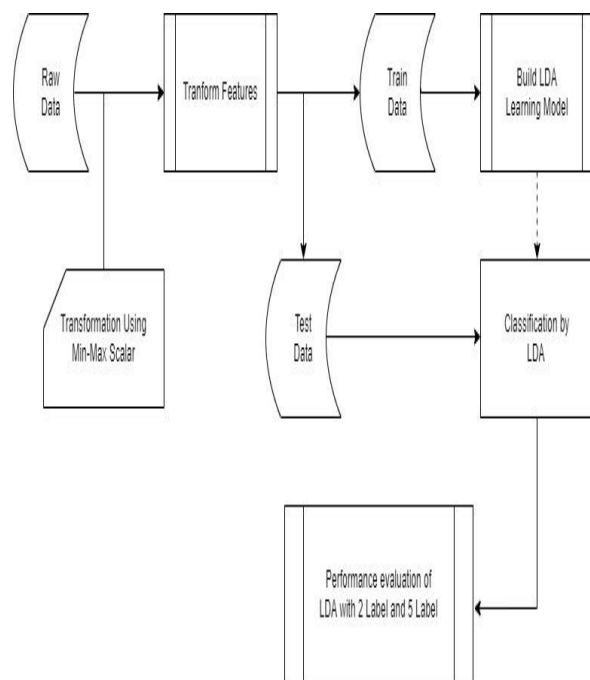
## 3. Metode

Dalam penelitian ini dibutuhkan bahan dan alat untuk menunjang dari proses penelitian yang dilakukan. Perancangan sistem penelitian ini menggunakan alat untuk mengimplementasikan metode *Linear Discriminant Analysis* dengan menggunakan bahasa Python 3.9.0 dan pustaka (*library*) *Python, software database mysql, connector, json* dan *visual studio code version 1.52.1 (user setup)*.

Data yang digunakan pada penelitian ini sebagai basis data awal dari pasien, yaitu terdiri dari jenis kelamin, umur, tinggi badan, berat badan, menderita diabetes atau tidak, jumlah rokok yang dikonsumsi setiap hari dan besar tekanan darah. Data tersebut akan dijadikan unsur paling utama dalam membangun sistem prediksi penyakit jantung. Dataset yang berupa basis data awal pasien tersebut akan disimpan dalam format *.csv (format text editor)*. Proses pengambilan data menggunakan dataset <https://archive.ics.uci.edu/ml/datasets/heart+disease>.

Nilai keluaran atau target dari penelitian ini menentukan adanya penyakit jantung pada pasien. Nilai keluarannya merupakan nilai integrasi dari 0 (tidak menderita penyakit jantung) hingga 4 nilai integrasi. Dengan dataset tersebut diupayakan untuk mendiagnosis penyakit jantung menjadi 4 tingkatan (stadium 1,2,3, dan 4) atau tidak menderita penyakit jantung (0). Secara keseluruhan hasil kombinasi 4 dataset yang terdiri dari *Cleveland, Hungary, Switzerland, dan the VA Long Beach* berjumlah 916 data pasien. Pada penelitian ini data dibagi menjadi 70% data latih berjumlah 687 pasien dan 30% data uji berjumlah 229 pasien.

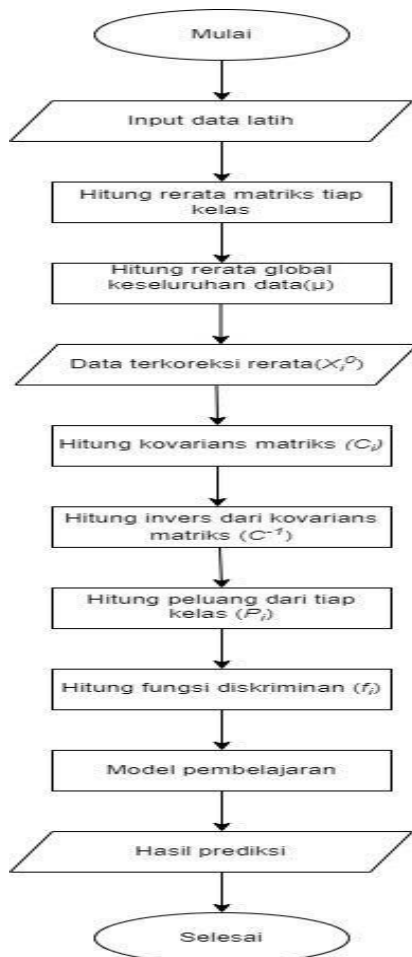
Untuk lebih jelasnya, berikut ini adalah prosedur kerja yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1.



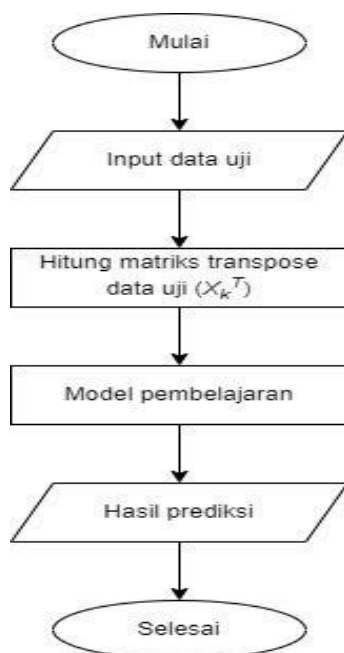
Gambar 1. Prosedur metode

### 3.1. Bagan Alir Metode Penelitian

Pada Penelitian ini, akan dibangun sistem menggunakan metode *Linear Discriminant Analysis* yang mampu memberikan hasil analisis terhadap prediksi penyakit jantung dengan digunakannya dataset yang diperoleh dari website *kaggle.com*. Berikut bagan alir dari sistem yang dibangun menggunakan Algoritma LDA dalam memprediksi penyakit Jantung. Tahapan dari alur metode penelitian ini dapat dilihat pada Gambar 2 dan 3.



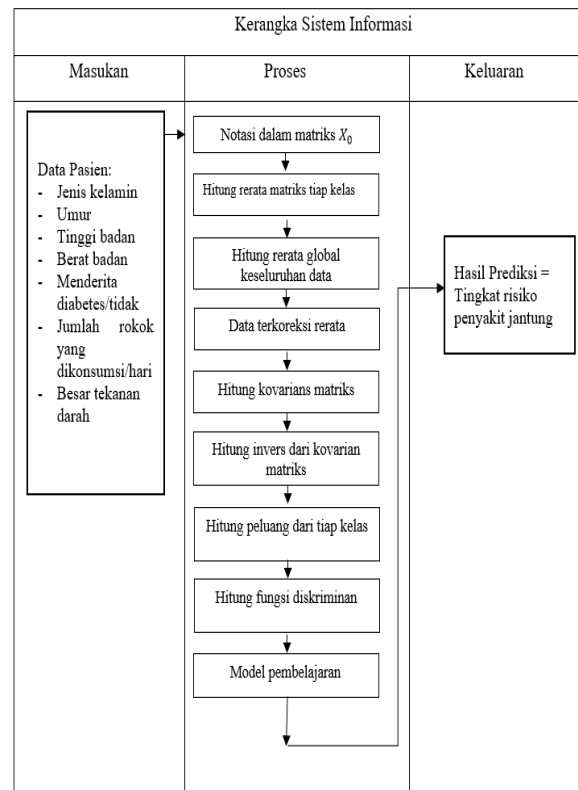
Gambar 2. Bagan alir data latih *Linear Discriminant Analysis*



Gambar 3. Bagan alir data uji algoritma *Linear Discriminant Analysis*

### 3.2. Kerangka Sistem Informasi

Berdasarkan pada penelitian ini, maka kerangka sistem informasi yang digunakan dengan tujuan untuk mengilustrasikan setiap proses dari sistem yang dibangun, mulai dari awal proses hingga akhir proses. Kerangka sistem diilustrasikan pada Gambar 4.



Gambar 4. Kerangka Sistem Informasi

## 4. Hasil dan Pembahasan

Hasil penelitian berupa klasifikasi *Linear Discriminant Analysis* (LDA) berbasis aspek sehingga didapatkan aspek kategori dari setiap prediksi risiko penyakit jantung dengan menggunakan bahasa pemrograman Python.

Klasifikasi *Linear Discriminant Analysis* yang dikembangkan ditranslansikan ke dalam bahasa pemrograman python dan menghasilkan sistem informasi untuk menyelesaikan permasalahan dalam klasifikasi risiko penyakit jantung. Untuk hasil evaluasi akan ditampilkan dalam bentuk matriks konfusi serta tabel yang berisi nilai akurasi, precision dan recall untuk masing-masing kategori. Pada perangkat lunak sistem informasi juga akan ditampilkan data ulasan dan aspek kategori dari data yang digunakan untuk melakukan pengujian sistem.

Perancangan sistem adalah tahapan dalam menjelaskan tentang langkah-langkah perancangan sistem yang akan dikerjakan untuk sistem prediksi metode *Linear Discriminant Analysis* (LDA) dalam identifikasi tingkat risiko terkena penyakit jantung. Sistem yang akan dibangun adalah sebuah sistem

untuk mendeteksi tingkat risiko seseorang untuk terkena penyakit jantung. Sistem ini secara umum dibangun untuk sistem pendeteksi atau sistem klasifikasi tingkat risiko seseorang terkena penyakit jantung berdasarkan pada pemeriksaan darah, kolesterol, gula darah, dan detak jantung.

Data yang dipakai untuk sistem ini adalah data set sebanyak 916 data Microsoft Excel dengan format (.xls) yang terdapat pada halaman lampiran 1. Pada penelitian ini data dibagi menjadi 70% data latih berjumlah 687 pasien dan 30% data uji berjumlah 229 pasien. Data yang sudah ada kemudian dilakukan pelatihan dengan metode LDA. Proses pelatihan LDA dilakukan dengan tahapan-tahapan sesuai dengan persamaan Algoritma LDA. Dalam proses pelatihan harus ditentukan bobot awal, data latih, dan data uji yang diambil dari data set yang sudah ada. Data yang diambil harus mewakili masing-masing kelas atau kelompok. Hasil dari proses pelatihan kemudian akan menghasilkan bobot baru yang akan digunakan untuk proses pengujian. Bobot yang digunakan untuk proses pengujian adalah bobot terakhir yang didapatkan dari proses pelatihan.

Selanjutnya proses pengujian dilakukan dengan menggunakan data uji yang diambil dari beberapa data set yang mewakili masing-masing kelas. Proses selanjutnya adalah mengambil bobot terakhir dari proses pelatihan. Sistem ini memiliki informasi berupa kelompok tingkat risiko yaitu kelompok risiko tinggi, risiko rendah dan tidak menderita penyakit jantung. Hasil keluaran sistem ini berupa informasi identifikasi tingkat risiko terkena penyakit jantung sehingga akan diketahui di tingkat kelompok manakah seseorang tersebut berada. Sistem ini berguna untuk membantu pihak laboratorium dan pihak pasien untuk cepat mengetahui di kelas atau kelompok manakah tingkat terkena penyakit jantung. Analisis Diskriminan merupakan dasar utama pada sistem pengenalan pola dengan memaksimalkan keragaman kelas, sifat dari *Linear Discriminant Analysis* ialah distribusi normal dan homoskedastis, yaitu suatu keadaan dari error bersifat konstan atau dapat dikatakan identik.

Algoritma klasifikasi pastinya berusaha untuk menghasilkan model yang menghasilkan akurasi yang baik. Kinerja model dari algoritma klasifikasi ditentukan pada saat model dihadapkan pada data testing, karena rata-rata model yang dipakai dapat memprediksi dengan benar pada semua data yang menjadi data pelatihnannya. Pengujian perbandingan data latih dan data uji digunakan untuk mengetahui pengaruh masing masing jumlah data terhadap tingkat akurasi yang akan dihasilkan dari sistem yang dibuat. Percobaan yang dilakukan untuk penelitian ini adalah sebanyak 1 kali percobaan.

Berdasarkan data yang dikumpulkan untuk penelitian ini dapat diklasifikasikan pada data uji. Pengujian aplikasi untuk mendeteksi risiko adanya penyakit jantung dapat dilihat pada Tabel 2.

Tabel 2. Klasifikasi Data Uji

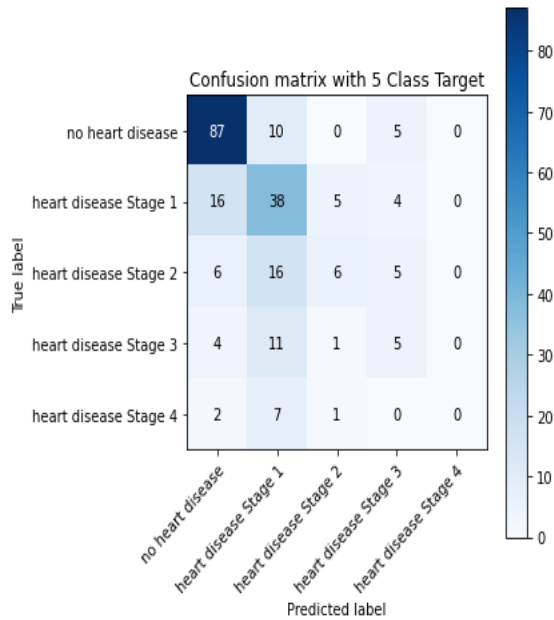
No	Aktual	Prediksi	Jumlah
1	Tidak ada penyakit jantung	Tidak ada penyakit jantung	87
		Penyakit jantung stadium 1	10
		Penyakit jantung stadium 2	0
		Penyakit jantung stadium 3	5
		Penyakit jantung stadium 4	0
2	Penyakit jantung stadium 1	Tidak ada penyakit jantung	16
		Penyakit jantung stadium 1	38
		Penyakit jantung stadium 2	5
		Penyakit jantung stadium 3	4
		Penyakit jantung stadium 4	0
3	Penyakit jantung stadium 2	Tidak ada penyakit jantung	6
		Penyakit jantung stadium 1	16
		Penyakit jantung stadium 2	6
		Penyakit jantung stadium 3	5
		Penyakit jantung stadium 4	0
4	Penyakit jantung stadium 3	Tidak ada penyakit jantung	4
		Penyakit jantung stadium 1	11
		Penyakit jantung stadium 2	1
		Penyakit jantung stadium 3	5
		Penyakit jantung stadium 4	0
5	Penyakit jantung stadium 4	Tidak ada penyakit jantung	2
		Penyakit jantung stadium 1	7
		Penyakit jantung stadium 2	1
		Penyakit jantung stadium 3	0
		Penyakit jantung stadium 4	0

#### 4.2 Performa LDA dengan 5 Target Label

Tabel 3 di atas merupakan hasil kinerja dari algoritma LDA dalam mengklasifikasikan penyakit jantung dengan 5 label yang digunakan sebagai target atau *output*. Berdasarkan hasil tersebut diperoleh nilai presisi 0,56, nilai pengulangan 0,59, nilai f1 0,56, dengan akurasi 59,38%. *Confusion matrix* berikut adalah apa yang didapatkan pada penyakit jantung klasik dengan LDA pada 5 target atau *output*.

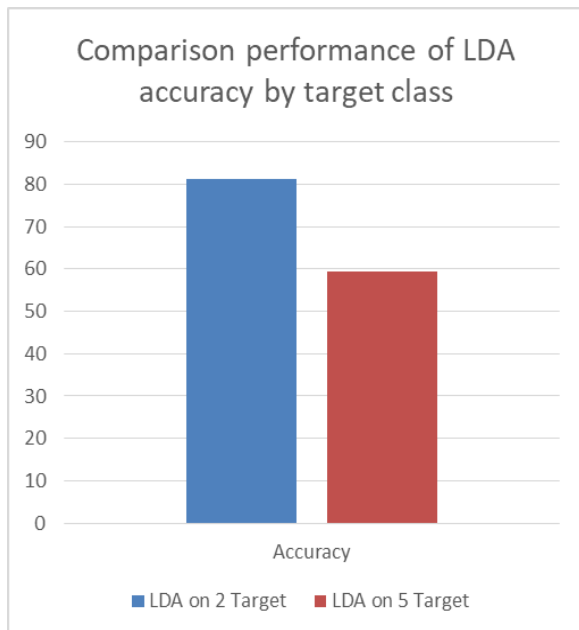
Tabel 3. Hasil performa LDA pada 5 target label

	Precision	Recall	F1-Score
<i>No Heart Disease</i>	0.76	0.85	0.80
<i>Heart Disease stage 1</i>	0.46	0.60	0.52
<i>Heart Disease stage 2</i>	0.26	0.18	0.26
<i>Heart Disease stage 3</i>	0.26	0.24	0.26
<i>Heart Disease stage 4</i>	0	0	0
<i>Weighted Average Accuracy = 59.38 %</i>	0.56	0.59	0.56

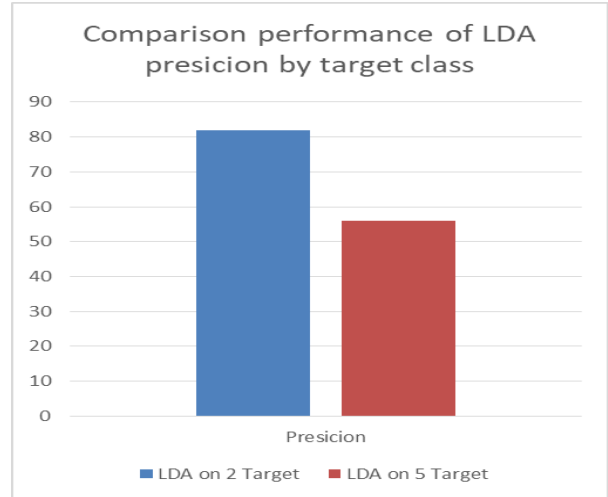


Gambar 5. Hasil *Confusion matrix* LDA dengan 5 target label

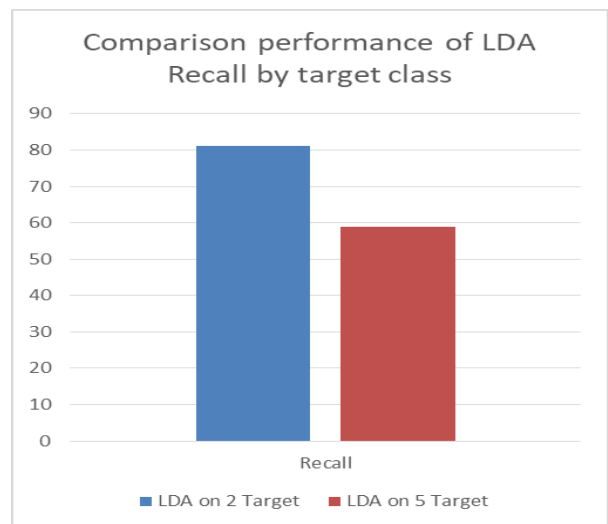
Penelitian ini melakukan pengujian dengan menggunakan Analisis Diskriminan Linier untuk mengklasifikasikan penyakit jantung. Dilakukan pula dengan dua kondisi nilai output, yang pertama dengan output 2 target atau label dan yang kedua dengan output 5 target atau label. Tujuannya adalah untuk menganalisis kinerja algoritma analisis diskriminan linier dalam mengklasifikasikannya.



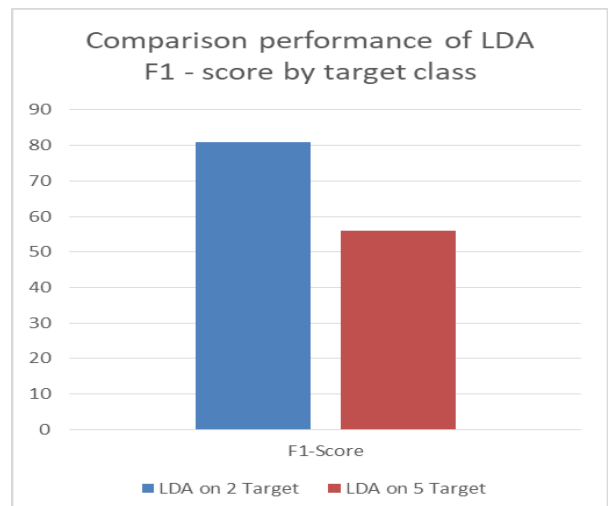
Gambar 6. Perbandingan akurasi LDA berdasarkan kelas target



Gambar 7. Perbandingan presisi LDA berdasarkan kelas target



Gambar 8. Perbandingan *recall* LDA berdasarkan kelas target



Gambar 9. Perbandingan *f1-score* LDA berdasarkan kelas target

## 5. Kesimpulan

Berdasarkan hasil penelitian, dapat kita simpulkan dari penelitian ini bahwa kinerja classifier dengan 2 kelas LDA lebih baik dari pada 5 kelas. Kinerja algoritma LDA dalam mengklasifikasikan penyakit jantung menggunakan target atau dua label yang digunakan sebagai hasil. Berdasarkan hasil tersebut didapatkan nilai akurasi 0.82, nilai pengulangan 0.81, nilai skor f1 0.81, dan akurasi 81,22%. Hasil performansi algoritma LDA dalam klasifikasi penyakit jantung menggunakan 5 penanda yang digunakan sebagai target atau keluaran. Berdasarkan hasil ini, nilai akurasi yang diperoleh adalah 0,56, nilai pengulangan adalah 0,59, dan nilai skor F1 0,56 adalah 59,38%. Penelitian ini masih pada tahap pengembangan dengan membandingkan berbagai metode klasifikasi lainnya.

### Daftar Pustaka

- Budiman, E., Santoso, E., Afirianto, T., 2017. Pendeteksi jenis autisme pada anak usia dini menggunakan metode *linear discriminant analysis* (LDA). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 1(7), 583–592.
- Dhuhita, W, Dutta, D., Sil, J., Dutta, P., 2020. A bi-phased multi objective genetic algorithm based classifier. 146
- Ghaderyan, P., Abbasi, A., Sedaaghi, M. H., 2014. An efficient seizure prediction method using KNN-based undersampling and linear frequency measures. *Journal of Neuroscience Methods* 232, 134–142.
- Hana, F. M., 2020. Perbandingan algoritma neural network dengan linier discriminant analysis (LDA) pada klasifikasi penyakit diabetes. 1, 1541–1541.
- Handayani, I., 2019. Application of k-nearest neighbor algorithm on classification of disk hernia and spondylolisthesis in vertebral column. *Indonesian Journal of Information Systems* 2(1), 57-66.
- Mai, H., Pham, T. T., Nguyen, D. N., Dutkiewicz, E., 2018. Non-laboratory-based risk factors for automated heart disease detection. *International Symposium on Medical Information and Communication Technology (ISMICT)* 1–6.
- Mutawalli, L., Zaen, M. T. A., Bagye, W., 2019. Klasifikasi teks sosial media twitter menggunakan support vector machine (studi kasus penusukan wiranto). *Jurnal Informatika dan Rekayasa Elektronik* 2(2), 43-51.
- Purnama, A., 2020. Edukasi dapat meningkatkan kualitas hidup pasien yang terdiagnosa penyakit jantung koroner. *Jurnal Kesehatan Indonesia* 10(2), 66–71.
- Sulistio, S., 2017. Aplikasi prediksi kelulusan mahasiswa dengan metode *Linear Discriminant Analysis*. *Computation : Journal of Computer Science and Information Systems* 1(1), 58.
- Trajdos, P., Burduk, R., 2021. Linear classifier combination via multiple potential functions. *Pattern Recognition* 111, 107681.
- Wei, J., Chen, H., 2020. Determining the number of factors in approximate factor models by twice K-fold cross validation. *Economics Letters*, 191, 109149.