



Prediksi Risiko Penyakit Jantung Menggunakan Algoritma *Linear Discriminant Analysis*

Syakur¹, Wardianto²

Program Studi Sistem Informasi, Universitas YPPI Rembang¹
Program Studi Informatika, Universitas Nurul Huda²

Jl. Pamotan No.KM.4, Tireman Timur, Tireman, Kec. Rembang, Kabupaten Rembang, Jawa Tengah
59219, Indonesia¹

Kampus A & Kampus B. Jalan Kota Baru, Desa Sukaraja, Kecamatan Buay Madang, Kabupaten
Ogan Komering Ulu Timur, Sumatera – Selatan, Kodepos 32161., Indonesia²
syakurpragen@uyr.ac.id¹, wardianto@unuha.ac.id²

Kata Kunci:

Penyakit jantung koroner, Machine learning, Linear Discriminant Analysis.

ABSTRAK

Penyakit jantung iskemik koroner merupakan penyebab kematian paling umum di seluruh dunia. Diagnosis penyakit ini hanya dapat ditegakkan melalui konsultasi langsung dengan dokter spesialis jantung yang relatif membutuhkan banyak tenaga. Di satu sisi diperlukan suatu sistem untuk mendeteksi penyakit jantung pada pasien dengan biaya minimal, dimulai dari perkembangan teknologi khususnya di bidang kecerdasan buatan. Ada metode yang bisa mendeteksi penyakit jantung secara otomatis yaitu machine learning, termasuk analisis diskriminan linier. Dalam penelitian ini, kami menerapkan algoritma analisis diskriminan linier pada klasifikasi penyakit jantung. Dataset yang digunakan diambil dari UCI Machine Learning Repository. Studi tersebut melakukan dua kondisi eksperimental di mana pasien diminta untuk memutuskan apakah mereka menderita penyakit jantung atau kondisi jantung lainnya berdasarkan skala lima poin. Hasil yang diperoleh membuktikan bahwa klasifikasi LDA dua kelas lebih baik dibandingkan klasifikasi LDA lima kelas. Tujuan atau hasilnya adalah penerapan algoritma LDA untuk mengklasifikasikan penyakit jantung dengan dua label. Dari hasil yang diperoleh diperoleh nilai presisi sebesar 0,82, nilai repeatability sebesar 0,81, nilai f1 sebesar 0,81, dan akurasi sebesar 81,22%. Hasil penerapan algoritma LDA untuk mengklasifikasikan penyakit jantung menjadi lima stadium dapat dijadikan hasil atau tujuan akhir. Berdasarkan hal tersebut diperoleh nilai presisi sebesar 0,56, nilai repeat sebesar 0,59, nilai f1 sebesar 0,56, dan presisi sebesar 59,38%.

Keywords

Coronary heart disease, Machine learning, Linear Discriminant Analysis.

ABSTRACT

Coronary ischemic heart disease is the most common cause of death worldwide. The diagnosis of this disease can only be established through direct consultation with a cardiologist which is relatively labor-intensive. On the one hand, a system is needed to detect heart disease in patients at minimal cost, starting from technological developments, especially in the field of artificial intelligence. There is a method that can detect heart disease automatically, namely machine learning, including linear discriminant analysis. In this study, we applied the linear discriminant analysis algorithm to the classification of heart disease. The dataset used was taken from the UCI Machine Learning Repository. The study conducted two experimental conditions in which patients were asked to decide whether they had heart disease or other heart conditions based on a five-point scale. The results



obtained proved that the two-class LDA classification was better than the five-class LDA classification. The purpose or result is the application of the LDA algorithm to classify heart disease with two labels. From the results obtained, a precision value of 0.82, a repeatability value of 0.81, an f1 value of 0.81, and an accuracy of 81.22% were obtained. The results of applying the LDA algorithm to classify heart disease into five stages can be used as the final result or goal. Based on this, the precision value was obtained as 0.56, the repeat value was 0.59, the f1 value was 0.56, and the precision was 59.38%.

---Jurnal JISTI @2024---

PENDAHULUAN

Organ jantung merupakan organ terpenting dalam tubuh manusia, dan ukurannya kira-kira sebesar kepala tangan manusia normal. Jantung bertindak sebagai pompa, mengedarkan darah ke seluruh tubuh manusia. Penyakit jantung yang umum terjadi memiliki penyebab yang berbeda-beda, tergantung dari faktor pemicunya. Penyakit jantung terjadi ketika aliran darah ke jantung terhambat akibat adanya lemak jahat di dalam tubuh manusia.

Penimbunan dari lemak di dalam arteri jantung ini dikenal dengan atherosklerosis. Atherosklerosis merupakan penyebab utama penyakit jantung koroner. Suatu kondisi sehingga jantung mengalami suatu gangguan atau kerusakan sehingga akan mengakibatkan terganggunya seluruh kinerja organ di dalam tubuh manusia (Junaid dan Kumar, 2020).

Penyakit jantung adalah istilah umum untuk semua jenis gangguan yang mempengaruhi jantung. Penyakit jantung atau Cardiovascular Disease (CDV) merupakan peringkat pertama penyebab kematian di dunia melebihi penyakit lain. Penyakit jantung merupakan penyakit tidak menular yang paling menyebabkan kematian. Pada tahun 2020, Organisasi Kesehatan Dunia (WHO) melaporkan bahwa penyakit jantung membunuh 82 juta orang setiap tahun di seluruh dunia. Prevalensi penyakit jantung cenderung meningkat setiap tahunnya. Diperkirakan akan lebih banyak lagi kematian yang terjadi akibat penyakit jantung di tahun-tahun mendatang (Purnama, 2020). Masih kurangnya kesadaran mengenai pola hidup sehat dan informasi mengenai penyakit jantung koroner sehingga dapat menyebabkan gejala awal tidak diketahui dan sulitnya pengobatan dini.

Saat ini, proses deteksi penyakit jantung bisa dilakukan secara manual. Tidak perlu berkonsultasi langsung dengan dokter spesialis jantung, melakukan beberapa pemeriksaan laboratorium, kemudian mengunjungi dokter spesialis jantung kembali. Tentu saja hal ini membutuhkan biaya yang relatif tinggi. Karena risiko kematian yang sangat tinggi, diperlukan sistem yang dapat mendeteksi penyakit jantung pada individu yang terkena dampak secara akurat dan hemat biaya (Mai et al., 2018).

Semakin berkembangnya teknologi dan informasi terutama di bidang kecerdasan buatan, teknik machine learning diperkenalkan untuk membantu untuk meningkatkan kemampuan pendeteksi otomatis. Salah satunya metode yang banyak digunakan pada machine learning adalah metode prediksi Linear Discriminant Analysis dan Quadratic Discriminant Analysis, sehingga pendeteksian penyakit jantung dapat dideteksi sedini mungkin. Dengan bantuan sistem deteksi, pemeriksaan data medis dapat dilakukan dengan lebih cepat, akurat dan akurat, sehingga menghindari kemungkinan kesalahan dalam diagnosis profesional (Ghaderyan et al., 2014).

Metode prediksi yang umum digunakan meliputi algoritma regresi logistik (LR), analisis diskriminan linier (LDA), analisis diskriminan kuadrat, dan K-nearest neighbour (KNN). Pada penelitian ini diteliti metode klasifikasi yang digunakan untuk mendeteksi penyakit jantung dari kategori risiko tinggi (menderita penyakit jantung) hingga kategori risiko rendah (tidak menderita penyakit jantung), yaitu LDA (analisis diskriminan linier). Untuk mengembangkannya. Penyakit jantung). Teknik analisis diskriminan linier digunakan dalam pengenalan pola dan pembelajaran mesin untuk mencari kombinasi fitur linier yang mencirikan atau memisahkan dua atau lebih objek atau peristiwa. Kombinasi yang dihasilkan dapat digunakan sebagai pengklasifikasi linier.



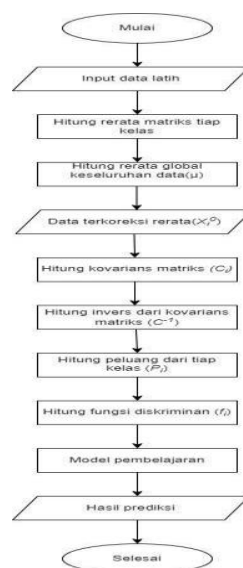
Alternatifnya, nilai presisi biasanya digunakan untuk proses reduksi dimensi sebelum menjadi lebih akurat dan presisi.

METODE PENELITIAN

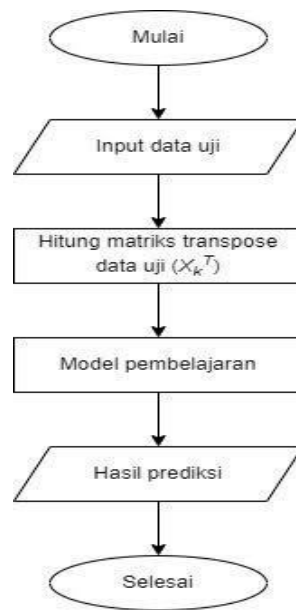
Dalam penelitian ini dibutuhkan bahan dan alat untuk menunjang dari proses penelitian yang dilakukan. Perancangan sistem penelitian ini menggunakan alat untuk mengimplementasikan metode Linear Discriminant Analysis dengan menggunakan bahasa Python 3.9.0 dan pustaka (library) Python, software database 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. Gambar menunjukkan sistem diagnosis penyakit jantung. Bagan Alir Metode Penelitian Pada Penelitian ini, akan dibangun sistem menggunakan metode Linear Diskriminant Analisis 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 2.1.



Gambar 2.1 Bagan alir data latih algoritma Linear Discriminant Analysis

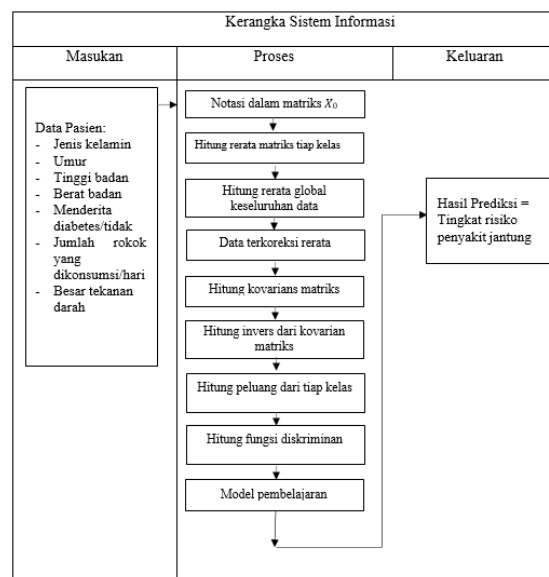


Gambar 2.2 Bagan alir data uji algoritma Linear Discriminant Analysis

Kerangka Sistem Informasi

Dengan bantuan sistem deteksi, pemeriksaan data medis dapat dilakukan dengan lebih cepat, akurat dan akurat, sehingga menghindari kemungkinan kesalahan dalam diagnosis profesional (Ghaderyan et al., 2014).

Berdasarkan penelitian tersebut digunakan kerangka sistem informasi dengan tujuan untuk menunjukkan setiap proses sistem yang dibangun dari awal hingga akhir. Kerangka sistem ditunjukkan pada Gambar 2.3.



Gambar 2.3 Kerangka Sistem Informasi



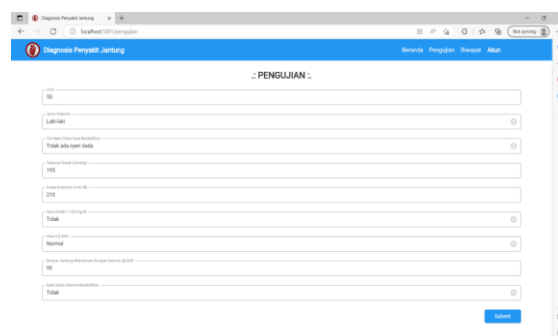
HASIL PENELITIAN DAN PEMBAHASAN

Hasil Penelitian

Klasifikasi Linear Disriminant Analysis (LDA) Pengembangan tersebut diterjemahkan ke dalam bahasa pemrograman Python untuk menciptakan sistem informasi penyelesaian masalah klasifikasi risiko penyakit jantung.

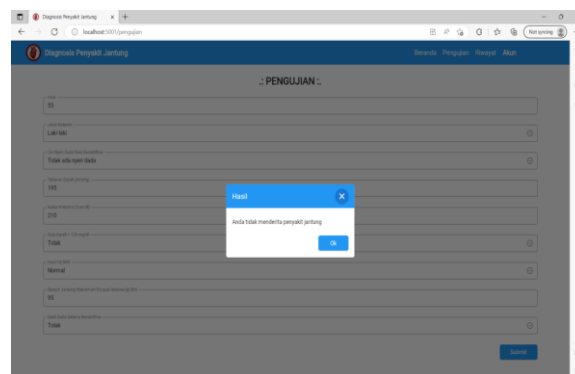
Hasil evaluasi disajikan dalam bentuk tabel yang berisi matriks konfusi serta nilai presisi, presisi, dan recall untuk setiap kategori. Perangkat lunak sistem informasi juga menampilkan data validasi dan aspek kategoris dari data yang digunakan untuk melakukan pengujian sistem.

Dataset ini terbagi menjadi 4 yaitu dataset Cleveland, Hungary, Switzerland, dan the VA Long Beach dan berisi 76 atribut, tetapi pada penelitian ini menggunakan 10 atribut. Secara khusus, dataset Cleveland adalah satu-satunya yang telah digunakan oleh para peneliti machine learning hingga saat ini. Namun penelitian ini menggunakan keempat dataset.

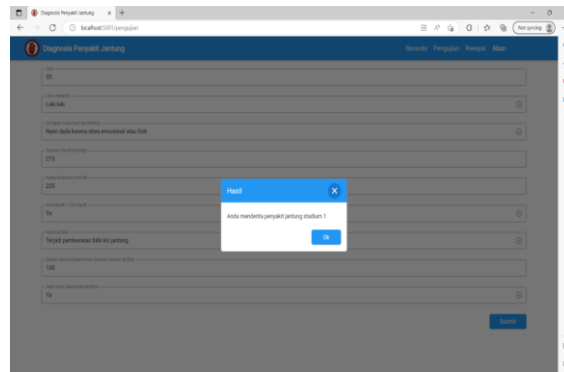


Gambar 3.3 Tampilan penginputan database pengujian

Halaman penginputan ke dalam sistem basis data. Penginputan ini digunakan untuk mengklasifikasi tingkat ukur adanya risiko penyakit jantung. Basis data yang dimasukkan untuk pengujian berupa usia, jenis kelamin, nyeri pada dada yang dirasakan, tekanan darah, kadar kolestrol, gula darah, denyut nadi.

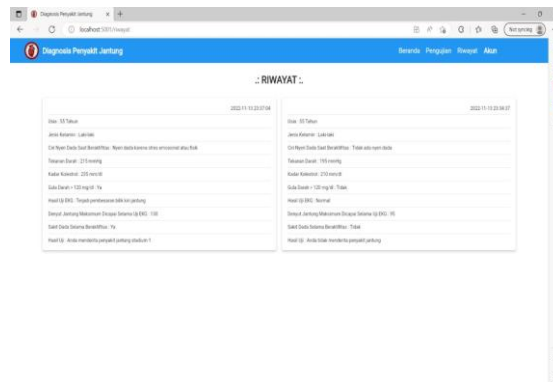


Gambar 3.4 Tampilan hasil pengujian tidak menderita penyakit jantung



Gambar 3.5 Tampilan hasil pengujian menderita penyakit jantung.

Pada halaman hasil tes pengujian setelah basis data diisi oleh pasien maka akan langsung muncul hasil jika menderita dan tidak menderita penyakit jantung



Gambar 3.6 Riwayat pengujian penyakit jantung

menderita penyakit jantung stadium 4

Tahap pertama adalah mengubah data awal ke dalam bentuk matriks dengan x sebagai data masukan dan y adalah nilai keluaran. Tabel 4.1 merupakan kriteria mulai stadium 1 hingga stadium 4.

Tabel 4.1 Kriteria pada stadium 1-4

Kriteria Penyakit Jantung
Stadium I Pasien dengan penyakit jantung tetapi tidak ada pembatasan aktivitas fisik. Aktivitas fisik biasa tidak menyebabkan kelelahan berlebihan, palpitasi, dispnea atau nyeri angina.
Stadium II Pasien dengan penyakit jantung dengan sedikit pembatasan aktivitas fisik. Merasa nyaman saat istirahat. Hasil aktivitas normal menyebabkan fisik kelelahan, palpitasi, dispnea atau nyeri angina.



<p>Stadium III</p> <p>Pasien dengan penyakit jantung yang terdapat pembatasan aktivitas fisik. Merasa nyaman saat istirahat. Aktifitas fisik ringan menyebabkan kelelahan, palpitasi, dispnea atau nyeri angina.</p>
<p>Stadium IV</p> <p>Pasien dengan penyakit jantung yang mengakibatkan ketidakmampuan untuk melakukan aktivitas fisik apapun tanpa ketidaknyamanan. Gejala gagal jantung dapat muncul bahkan pada saat istirahat. Keluhan meningkat saat melakukan aktifitas.</p>

Pembahasan

Implementasi Sistem Data yang dipakai untuk sistem ini adalah dataset 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.

```
Confusion matrix, without normalization
[[87 28]
 [15 99]]
Confusion matrix, without normalization
[[87 10 0 5 0]
 [16 38 5 4 0]
 [ 6 16 6 5 0]
 [ 4 11 1 5 0]
 [ 2 7 1 0 0]]
```

Gambaran Confusion matrix without normalization



Tahapan Linear Discriminant Analysis

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, (Ghwanmeh, 2013). 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 4.5, serta terlampir untuk keseluruhan data uji pada lampiran pada halaman 48.

Tabel 4.1 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



Tahap pertama dalam penelitian ini adalah menyiapkan dataset, selanjutnya di split datanya dibagi menjadi data uji dan data latih dengan presentase 70% untuk data pelatihan dan 30% untuk data pengujian. Selanjutnya dilakukan klasifikasi menggunakan algoritma LDA dan dihitung akurasi menggunakan Confusion Matrix. Dari hasil pengujian menghasilkan performance vector dan tabel Confusion Matrix. Metode confusion matrix dipilih karena mampu memberikan perbandingan hasil klasifikasi model dengan klasifikasi sebenarnya menggunakan 4 kombinasi nilai prediksi yaitu True Positive (TP) menunjukkan, False Positive (FP), False Negative (FN) dan True Negative (TN). Selain itu metode confusion matrix cocok untuk mengukur kinerja dari model klasifikasi yang menghasilkan dua output kelas seperti pada penelitian ini yaitu kelas positif dan negatif (Merawati Ni Luh Putu, 2019). Confusion matrix merupakan salah satu tools penting dalam metode evaluasi yang digunakan pada machine learning yang biasanya memuat dua kategori atau lebih (Manning, 2009). Setiap unsur matriks menunjukkan jumlah contoh data uji untuk kelas sebenarnya yang digambarkan dalam bentuk baris sedangkan kolom menggambarkan kelas yang diprediksi. Untuk melakukan evaluasi model, pada percobaan ini dilakukan dengan iterasi pada dataset sebagai validasi silang untuk menemukan nilai akurasi prediksi yang terbaik

Pada penelitian ini confusion matrix digunakan untuk evaluasi sistem. Hasil evaluasi sistem menggunakan data train sebanyak 916, Hasil percobaan diperoleh sejumlah 87 ulasan diprediksi salah dengan nilai Negatif (True Negative), dan 99 diprediksi benar dengan nilai positif (True Positive). Sedangkan 28 ulasan diprediksi salah dengan nilai Negatif (False Negative) dan 15 ulasan diprediksi salah dengan nilai Positif (False Positive). Hasil matriks konfusi sentimen dapat dilihat pada Tabel 4.2.

Tabel 4.2 Label matriks konfusi

Label	Prediksi label	
	Tidak ada penyakit jantung	Penyakit jantung stadium 1
Kebenaran	87	28
Tidak ada penyakit jantung	15	99
Penyakit Jantung stadium 1		

Hasil dan Analisis Pengujian Perbandingan Keseimbangan Class Data Latih dengan Data Uji

Pengujian perbandingan class data latih dan data uji digunakan untuk mengetahui pengaruh masing masing jumlah data terhadap tingkat akurasi yang akan dihasilkan dari sistem yang dibuat. Jumlah perbandingan class data untuk pengujian data latih (Panjaitan, 2013). Pengujian keseimbangan class data latih dan data uji hanya dilakukan sebanyak 1 kali, dikarenakan banyak nya data latih harus sama antar class nya yaitu sebanyak 230 data latih dengan nilai class 1 sebanyak 82 data, class 2 sebanyak 13 data, class 3 sebanyak 19 dan class 0 sebanyak 115 data dari skenario diatas didapat hasil pengujian dengan nilai akurasi sebesar 59% pada data uji, dan pada data train akurasi 81,22%. Pada hasil pembahasan di atas bisa disimpulkan bahwa ketika pengujian perbandingan keseimbangan class data latih dengan data uji berbeda signifikan, karena rata-rata akurasi adalah 59% pada data uji dengan perbandingan ketidakseimbangan data latih dengan data uji sebesar 115 class 0: 82 class 1: 13 class 2: 19 class 3. Artinya bahwa metode LDA kurang cocok dalam pengujian keseimbangan class data uji dilihat dari nilai akurasi. Ini dikarenakan kurangnya data untuk penyakit jantung stadium 2,3, dan 4 didalam data testing.



Hal ini dapat disimpulkan bahwa pendekatan mencapai tingkat pengakuan yang tinggi dibandingkan dengan pendekatan lain dalam literatur yang diterbitkan. Hal ini terjadinya ketidakakuratan data diagnosis penyakit jantung terhadap data tingkat risiko, sehingga mengakibatkan rendahnya kinerja system (El-Sayed, 2018). Analisis diskriminan linier digunakan dalam penelitian ini untuk mengetahui akurasi klasifikasi dalam prediksi penyakit jantung (El-Sayed, 2018).

KESIMPULAN

Setelah dilakukan pengujian dan analisis dalam penelitian yang sudah dilakukan, maka dapat diambil kesimpulan yaitu Sistem prediksi risiko penyakit jantung dikembangkan menggunakan algoritma klasifikasi analisis diskriminan linier yang melalui proses pelatihan. Pelatihan tahap pertama meliputi penentuan berat badan awal yang diambil sampelnya secara acak dari dataset yang tersedia berupa data rekam medis pasien hasil pemeriksaan darah laboratorium rumah sakit.

Beberapa pengujian yang dilakukan menunjukkan hasil sebagai berikut: Skor F1 sebesar 38,009%, presisi sebesar 75,65% yaitu true positif dibagi true positif ditambah false positif, dan recall 85 yaitu 29% dari true positif dibagi . Negatif palsu ditambahkan ke positif sebenarnya. Dari beberapa pengujian yang dilakukan diperoleh data akurasi tertinggi pada kategori data latih dengan akurasi sebesar 81,22%, sedangkan untuk kategori data uji perbandingan antara data latih dan data uji sebesar 70% : dengan akurasi sebesar 30%. 59% tercapai.

DAFTAR PUSTAKA

- Abufaraj, M., Shariat, S., Moschini, M., Rohrer, F., Papantoniou, K., Devore, E., ... & Schernhammer, E. (2020). The impact of hormones and reproductive factors on the risk of bladder cancer in women: results from the Nurses' Health Study and Nurses' Health Study II. *International journal of epidemiology*, 49(2), 599-607.
- Angrestianingsih, A., Widodo, A. W., dan Furqon, M. T. (2019). Implementasi Metode Linear Discriminant Analysis (LDA) Untuk Klasifikasi Pengambilan Mata Kuliah Pilihan. Vol 3(10), 10337–10343.
- DwiBhaskara, A. (2016). Implementasi Linear Discriminant Analysis (Lda) Untuk Klasifikasi Tingkat Resiko Terkena Penyakit Stroke (Doctoral dissertation, Universitas Brawijaya).
- Budiman, E., Santoso, E., & Afirianto, T. (2017). Pendeteksi Jenis Autis pada Anak Usia Dini Menggunakan Metode Linear Discriminant Analysis (LDA). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, 964X.
- Caulkins, J. P., Ding, W., Duncan, G., Krishnan, R., dan Nyberg, E. (2006). A method for managing access to web pages: Filtering by Statistical Classification (FSC) applied to text. *Decision Support Systems*, Vol 42(1), 144–161.
- Dimas, K. J., Anisa, R., & Sulvianti, I. D. (2020). Perbandingan Quadratic Discriminant Analysis dan Support Vector Machine untuk Klasifikasi Tutupan Lahan di DKI Jakarta. *Xplore: Journal of Statistics*, 9(1), 12-20.
- Dutta, D., Sil, J., dan Dutta, P. (2020). A bi-phased multi-objective genetic algorithm based classifier. Vol 146.
- El-Sayed, R. S. (2018). Linear Discriminant Analysis for An Efficient Diagnosis of Heart Disease via Attribute Filtering Based on GeneticAlgorithm. *journal of computer Department of Math & Computer Science, Faculty of Science, Al- Azhar University, Cairo. Egypt.*



Ghaderyan, P., Abbasi, A., dan Sedaaghi, M. H. (2014). An efficient seizure prediction method using KNN-based undersampling and linear frequency measures. *Journal of Neuroscience Methods*, Vol 232, 134–142.

Ghwanmeh, S., Mohammad, A., & Al-Ibrahim, A. (2013). Innovative artificial neural networks-based decision support system for heart diseases diagnosis. Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Springer

Berlin.

Han, J., Kamber, M., dan Pei, J. (2012). Introduction. In *Data Mining*.

Hana, F. M. (2020). Perbandingan Algoritma Neural Network Dengan Linier Discriminant Analysis (Lda) Pada Klasifikasi Penyakit Diabetes. Vol 1, 1541– 1541.

Jabri, A., Kumar, A., Verghese, E., Alameh, A., Kumar, A., Khan, M. S., ... & Kalra, A. (2021). Meta-analysis of effect of vegetarian diet on ischemic heart disease and all-cause mortality. *American Journal of Preventive Cardiology*, 7, 100182.

Jumeilah, F. S. (2017). Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, Vol 1(1), 19.

Junaid, M. J. A., dan Kumar, R. (2020). Data Science and Its Application in Heart Disease Prediction. *Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020*, 396–400.

Mai, H., Pham, T. T., Nguyen, D. N., dan Dutkiewicz, E. (2018). Non-Laboratory- Based Risk Factors for Automated Heart Disease Detection. *International Symposium on Medical Information and Communication Technology, ISMICT, 2018-March*, 1–6.45

Manning, D. d. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Merawati Ni Luh Putu, A. Z. (2019). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok. *Jurnal Rekayasa Sistem dan Teknologi Informasi*, 1-9.

Mori, K., Nomura, T., Akezaki, Y., Yamamoto, R., dan Iwakura, H. (2020). Impact of Tai Chi Yuttari-exercise on arteriosclerosis and physical function in older people. *Archives of Gerontology and Geriatrics*, 87(January), 104011.

Manning, D. d. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Merawati Ni Luh Putu, A. Z. (2019). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok. *Jurnal Rekayasa Sistem dan Teknologi Informasi*, 1-9.

Panjaitan, H. B. (2013). *Linier Discriminant Analysis Dalam Klasifikasi Data Pada Data Teori Informasi Dengan Metode Cross-Validation*. Skripsi. Bandung: Universitas Telkom.

Praiss, A. M., Huang, Y., St. Clair, C. M., Tergas, A. I., Melamed, A., Khoury-

Collado, F., Hou, J. Y., Hu, J., Hur, C., Hershman, D. L., dan Wright, J. D. (2020). Using machine learning to create prognostic systems for endometrial cancer. *Gynecologic Oncology*, Vol 159(3), 744–750.

Purnama, A. (2020). Edukasi dapat meningkatkan kualitas hidup pasien yang terdiagnosa penyakit jantung koroner. *Jurnal Kesehatan Indonesia*, X(2), 66– 71.

Putria. (2018). Data Mining Using Apriori Algorithm. *Computer Based Information System Journal*, Vol 06(No.1), 29–39.

Roihan, A., Sunarya, P. A., dan Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, Vol 5(1), 75–82.



Sulistio, S. (2017). Aplikasi Prediksi Kelulusan Mahasiswa Dengan Metode Linear Discriminant Analysis. *Computation : Journal of Computer Science and Information Systems*, Vol 1(1), 58.

Trajdos, P., dan Burduk, R. (2021). Linear classifier combination via multiple potential functions. *Pattern Recognition*, Vol 111, 107681.

Verma, S. P., Uscanga-Junco, O. A., & Díaz-González, L. (2021). A statistically coherent robust multidimensional classification scheme for water. *Science of the Total Environment*, 750, 141704.

Wei, J., dan Chen, H. (2020). Determining the number of factors in approximate factor models by twice K-fold cross validation. *Economics Letters*, Vol 191, 109149.

Wei, Y., Gu, K., & Tan, L. (2022). A positioning method for maize seed laser-measurement. *Information Processing in Agriculture*, 9(2), 224-232.

Zhou, X., Lu, P., Zheng, Z., Tolliver, D., & Keramati, A. (2020). Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliability Engineering & System Safety*, 200, 106931.