

Pengurangan Dimensi dengan Metode Linear Discriminant Analist (LDA)

Winarnie^{1*}, Kusri², Anggit Dwi Hartanto³

^{1,2,3}Program Studi Magister Teknik Informatika, Universitas Amikom Yogyakarta

*winarnie@students.amikom.ac.id

Abstrak

Tujuan penelitian ini adalah untuk mengurangi dimensi dari dataset yang mempengaruhi pada prediksi kanker payudara.. Data yang dipakai pada penelitaian merupakan data yang sangat banyak atau disebut data yang berdimensi tinggi. Penggunaan algoritma klasifikasi memiliki kelemahan jika digunakan pada data berdimensi tinggi., sehingga diperlukan metode yang tepat dalam mereduksi dimensi atau variabel yang digunakan. Ada beberapa metode yang dapat digunakan dalam mereduksi dimensi. Pada penelitian ini menggunakan metode *linear discriminant analist* (LDA). LDA merupakan algoritma *supervised machine learning* yang digunakan sebagai pengklasifikasikan data kedalam beberapa kelas, menggunakan teknik linier untuk menentukan dari kumpulan variabel linier yang terbaik dalam memisahkan kelas data. LDA digunakan untuk mengurangi variabel dataset yang digunakan dengan mempertahankan informasi yang penting untuk proses klasifikasi. Metode yang digunakan pada penelitian adalah menggunakan LDA pada pemrosesan data dan selanjutnya menggunakan model regresi logistik untuk proses klasifikasi. Kesimpulan yang didapat pada penelitian ini adalah LDA dapat mengatasi masalah klasifikasi multiklas. Hasil yang didapat adalah 16 kasus salah dari total 455 kasus sehingga didapat hasil 0.035% kesalahan klasifikasi

Kata kunci : Kanker Payudara, Klasifikasi, LDA, Reduksi variabel, Regresi Logistik

Abstract

The purpose of this study is to reduce the dimensions of the dataset that affect the prediction of breast cancer. The data used in research is very much data or is called high-dimensional data. The use of classification algorithms has weaknesses when used on high-dimensional data, so an appropriate method is needed to reduce the dimensions or variables used. There are several methods that can be used to reduce dimensions. In this study using the method of linear discriminant analysis (LDA). LDA is a supervised machine learning algorithm that is used to classify data into several classes, using a linear technique to determine the best set of linear variables to unify class data. LDA is used to reduce the dataset variables used by retaining information that is important for the classification process. The method used in this research is using LDA in data processing and then using a logistic regression model for the classification process. The conclusion obtained in this study is that LDA can overcome the problem of multiclass classification. The results obtained were 16 wrong cases out of a total of 455 cases so that the results obtained were 0.035% misclassification.

Keywords : *Breast Cancer, Classification, LDA, Reduction of variables, Logistic Regression.*

1. Pendahuluan

Kanker payudara merupakan kanker yang umum pada wanita. Kanker payudara penyebab kedua kematian kanker pada wanita setelah kanker

paru-paru. Deteksi dini pencegahan kanker payudara merupakan solusi dari masalah tersebut. Kanker payudara penyebab 22% kematian oleh kanker pada wanita. [1]. Algoritma

klasifikasi merupakan algoritma *supervised machine learning*. Klasifikasi merupakan teknik mengelompokkan data kedalam klas data yang telah ditentukan sebelumnya. [2] Ada beberapa algoritma yang termasuk dalam algoritma klasifikasi diantaranya adalah : *support vector machine*, *decision tree*, *naïve bayes*, *k-nn*, *random forest*, dan *logistic regrestion*. [3] Regresi logistik merupakan algoritma klasifikasi yang sudah umum digunakan untuk menentukan hubungan antara dua faktor data dan memprediksi nilai dari salah satu faktor berdasarkan faktor yang lain. [4] Regresi logistik memiliki kelemahan ketika digunakan pada data yang variabelnya sangat banyak atau disebut data berdimensi tinggi [5], sehingga diperlukan teknik pengurangan dimensi untuk mengatasi masalah yang terjadi pada regresi logistik. Beberapa teknik pengurangan dimensi yaitu *principal component analist (PCA)* [6], *linear discriminant analist (LDA)*[7], dan *lasso* [8]. Menggunakan dataset kanker payudara yang diambil dari *kaggle repository* yang ada maka penelitian ini dilakukan untuk mengetahui prediksi jenis kanker, payudara apakah jenis kanker jinak atau ganas dengan menggunakan teknik pengurangan dimensi *linear discriminant analist (LDA)* untuk mengatasi masalah pada algoritma regresi logistik sehingga mendapatkan hasil akurasi yang tinggi

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Ada beberapa penelitian yang telah dilakukan berhubungan dengan reduksi dimensi untuk mendukung dari penelitian yang akan dilakukan, beberapa penelitian tersebut diantaranya :

1. Penelitian untuk mencari referensi publikasi ilmiah yang relevan dengan melihat kemiripan dengan dokumen lain pada database dan dilihat presentasi kemiripannya, dengan menggunakan teknik frekuensi itemset mining (FIM) menganalisa keberhasilan reduksi dimensi dengan melihat hasil akurasi yang didapat. Penelitian ini menghasilkan 80.17% reduksi dimensi dengan menggunakan FIM.[9]. Penelitian ini dilakukan oleh Dini.N dkk
2. Penelitian tentang klasifikasi penyakit diabetes menggunakan algoritma K-NN dan metode reduksi dimensi *principal component analist (PCA)* dilakukan oleh Joko, dkk[10]. Hasil dari penelitian ini mendapatkan 3 komponen dari 8 variabel dataset. Hasil akurasi terbaik sebesar 77.92% pada $k = 7$. Tahap – tahap pada penelitian ini dimulai dengan normalisasi data dengan z-score kemudian melakukan reduksi dimensi dengan membentuk matrik kovarian, menghitung nilai eigen, vector eigen, dan membentuk komponen utama dari hasil vector eigen

- selanjutnya menghitung akursi dengan menggunakan algoritma K-NN.
- Fitrianiingsih, dkk[11] melakukan penelitian mereduksi variabel yang mempengaruhi kegunaan ginjal pada tikus. Penelitian ini menggunakan metode principal component analist untuk mereduksi dimensi dari data yang dipakai yaitu fungsi ginjal tikus, menghasilkan 3 komponen utama yang benar-benar mempengaruhi dataset dari 8 variabel yang ada.
 - Penelitian yang dilakukan oleh Dian Ad, dkk[12] dengan mengidentifikasi wajah menurut jenis kelamin dari gambar kartu mahasiswa dari sebuah perguruan tinggi. Metode yang digunakan adalah principal component analist sebagai analisa data kemudian diterapkan pada dua model algoritma yaitu *support vector machine* (SVM), dan *artificial neural network* (ANN). Hasil kinerja akurasi yang didapat SVM sebesar 77.5%, dan ANN sebesar 78.10%.
 - Penelitian tentang ekstrasi warna dengan hue dan saturasi pada tingkat kematangan buah nanas dilakukan oleh Rahmat. D, dkk [13], menggunakan metode hue dan saturasi untuk mendapatkan informasi warna sebagai identifikasi kemudian di proyeksi pada ruang dimensi yang lebih kecil untuk pengenalan pola dengan menggunakan metode linear

discriminant analist (LDA) dan didapat performa akurasi sebesar 83%.

- Penelitian selanjutnya dilakukan oleh [14] yang melakukan penelitian untuk mendapatkan fungsi dikriminan dengan metode WMOM menggunakan data diabetes. Hasil penelitian yang diperoleh sebesar 16.67% kesalahan klasifikasi

2.2. Landasan Teori

1. Klasifikasi

Klasifikasi dalam mesin pembelajaran merupakan proses memprediksi kelas berdasarkan data yang sudah dilatih sebelumnya. Beberapa metode klasifikasi diantaranya : *naives bayes*, *support vector machine*, *decision tree*, dan *logistic regresi*. [15]. Dengan memanfaatkan teknik Pembelajaran Mesin pada klasifikasi sangat membantu. Beberapa Contohnya ketika berhadapan dengan multidimensi dan fitur berkelanjutan, jaringan saraf tiruan dan SVM cenderung berkinerja lebih andal. Selain itu, ketika berhadapan dengan fitur diskrit atau kategorikal, sistem berbasis logika cenderung berkinerja lebih baik. Untuk mencapai akurasi prediksi maksimum, besar ukuran sampel diperlukan untuk model jaringan saraf dan SVM; namun jaringan *Naive Bayesian* (NB) mungkin membutuhkan dataset yang relatif kecil ,SVM adalah algoritma biner; oleh karena itu, *errorcorrecting* output coding (ECOC) harus digunakan. yaitu, pendekatan pengkodean

keluaran dapat mengurangi masalah multi-kelas ke satu set dari beberapa masalah klasifikasi biner. Apalagi kebanyakan pohon keputusan algoritma tidak dapat bekerja dengan tepat ketika ada kendala yang membutuhkan partisi diagonal. Dengan demikian, daerah yang dihasilkan setelah partisi semuanya adalah hyperrectangles. Ketika multikolinieritas adalah sekarang, JST dan SVM beroperasi dengan baik, dan ada korelasi nonlinear antara input dan fitur keluaran.[16]

2. Linear Discriminant Analist (LDA)

Tujuan utama dari teknik reduksi dimensi adalah untuk mengurangi dimensi dengan menghilangkan redundan dan dependen fitur dengan mengubah fitur dari ruang dimensi yang lebih tinggi, ke ruang dengan dimensi yang lebih rendah. Ada dua pendekatan utama dari teknik reduksi dimensi, yaitu tanpa pengawasan dan pengawasan. Dalam pendekatan tanpa pengawasan, tidak ada kebutuhan untuk pelabelan kelas data. Sedangkan dalam pendekatan terawasi, teknik reduksi dimensi mempertimbangkan label kelas[17].

Linear Discriminant Analist (LDA) merupakan salah satu metode untuk mereduksi dimensi. LDA termasuk dalam teknik supervised yang menggunakan kelas – kelas pada dataset yang digunakan. Teknik LDA dikembangkan untuk mengubah fitur ke ruang dimensi yang lebih rendah, yang memaksimalkan rasio varians antar kelas ke varian dalam kelas. Ada dua jenis LDA

untuk menangani kelas: tergantung kelas dan kelas-independen. Dalam LDA yang bergantung pada kelas, satu ruang dimensi bawah yang terpisah dihitung untuk masing-masing kelas untuk memproyeksikan datanya di atasnya sedangkan, di LDA kelas independen, setiap kelas akan dianggap sebagai a memisahkan kelas terhadap kelas lainnya .. Di dalam jenis, hanya ada satu ruang dimensi yang lebih rendah untuk semua kelas untuk memproyeksikan data mereka di atasnya [18]. Tujuan dari teknik LDA adalah memproyeksikan matriks data asli ke ruang dimensi yang lebih rendah. [19].

3. Regresi Logistic (LR)

Regresi logistic merupakan algoritma yang menggambarkan hubungan antara variabel independent dan variabel dependent[20]. Variabel dependen memiliki dua katagori yang secara umum digambarkan dengan angka 0 dan 1[21]. Regresi logistic digunakan untuk menjelaskan hubungan antara X dan Y yang tidak linier [22]. Hubungan tersebut dapat dilihat pada rumus dibawah

$$p(x) = \frac{1}{1 + e^{-\alpha - \beta x}}$$

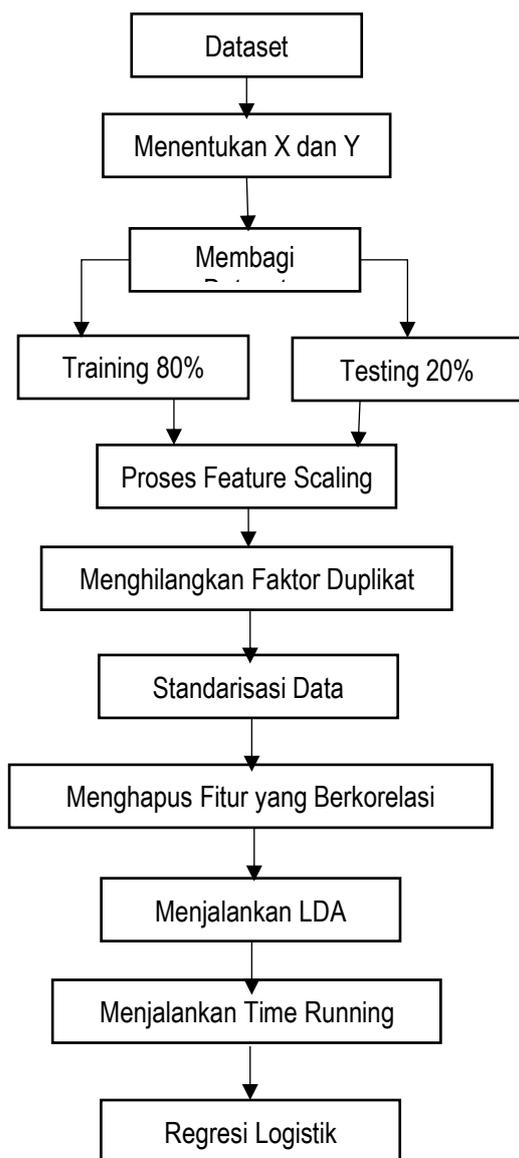
β, α : Dua parameter model yang akan dievaluasi, sedangkan $\beta > 0$ Mereka tidak dibatasi.

$p(x)$: Probabilitas respons dan spesifik antara(1,0) .

(x) : Variabel penjelas di mana $(-\infty < x < \infty)$.

3. Metode Penelitian

Pada penelitian ini menggunakan metode *linear discriminant analyst* (LDA) sebagai analisa data kemudian untuk pengujian performa model menggunakan algoritma regresi logistic dan konfosisus matrik.



Gambar 1. Tahapan Penelitian

Keterangan :

- a Menentukan Nilai X dan Y Untuk menentukan nilai X dan Y kita menentukan terlebih dahulu

variabel dependen dan independen atau nilai X dan Y. Nilai Y yang digunakan adalah variabel target yang memiliki dua klas katagori yaitu 0 dan 1.

- b Membagi dataset menjadi data training dan data testing. Tahap setelah selanjutnya adalah membagi data. Data training digunakan sebagai data pembelajaran dari model untuk dapat memprediksi variabel dependen. Data testing berfungsi sebagai pelengkap dari data training yang digunakan untuk menguji model algoritma apakah berhasil dalam memprediksi variabel dependen berdasarkan variabel independen.
- c. Menghilangkan faktor duplikat. Tahap selanjutnya adalah menghilangkan faktor duplikat dengan menghilangkan fitur konstanta dan quasi konstanta dengan menggunakan ambang batas 0.01 yang berarti memiliki 99% kemiripan antar data.
- d. Menstandari data dengan skala yang sama. Penstandarisasi berfungsi untuk menormalkan rentang variabel independen untuk menghindari hasil yang bias. Skala standarisasi dilakukan saat data tidak terdistribusi secara normal.
- e. Menghapus fitur yang berkorelasi. Menghapus fitur yang berkorelasi untuk mendapatkan variabel yang benar-benar sangat dibutuhkan atau untuk mengurangi dimensidari data.

- f. Menjalankan LDA. Pada penelitian ini menggunakan komponen 1 karena jumlah klas maksimal pada dataset yang digunakan adalah 1. LDA mengekstrksi beberapa komponen untuk memaksimalkan kelas yang terpisah.
- g. Menjalankan Time running. Pada tahap menjalankan tool untuk time running menggunakan bahasa pemrograman python , pada proses ini akan terlihat berapa banyak waktu yang dihabiskan dalam menjalankan program ketika menggunakan proses LDA dan tanpa menggunakan LDA.
- h. Menjalankan performa evaluasi dengan algoritma regresi logistik. Setelah menjalankan LDA sebagai reduksi dimensi atau variabel dari data yang kita pakai selanjutnya akan diuji performa dari model yang telah dibuat dengan menggunakan algoritma regresi logistik dengan cara menggunakan konfosis matrik sehingga diketahui nilai akurasi, F1 score, recall, presicion, dan support dari keluaran algoritma regresi logistik menggunakan LDA dan tanpa menggunakan LDA.

4. Hasil dan Pembahasan

4.1. Hasil

1) Dataset

Penelitian ini menggunakan platform *google colab* dan bahasa pemrograman python. Dataset yang digunakan adalah dataset *breast*

cancer yang diambil dari *kaggle repository* <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>. Dataset terdiri dari 31 atribut dan 569 record. Rincian 5 atribut atas pada dataset dapat dilihat pada tabel 1.

Tabel 1. Rincian Dataset

No.	Atribut	Katagori
1	target	Y (dependen)
2	radius_mean	X (Independen)
3	texture_mean	X (Independen)
4	perimeter_mean	X (Independen)
5	area_mean	X (Independen)

2) Hasil dari Menentukan Nilai X dan Y

```
X = data.drop('target', axis = 1)
y = data['target']
X.shape, y.shape

((569, 30), (569,))
```

Gambar 2 Nilai X dan Y

Dataset awal terdapat 30 atribut dan 569 record. Seperti terlihat pada gambar 2. Nilai Y yang digunakan adalah variabel target yang memiliki dua klas katagori yaitu 0 dan 1. Untuk angka 0 menunjukkan diagnosa kanker ganas dan angka 1 sebagai diagnosa kanker jinak. Variabel X adalah semua variabel selain dari variabel target.

3) Hasil Menghapus fitur yang berkorelasi .

Hasil dari menghilangkan fitur yang berkorelasi adalah data menjadi 455 record dan 4 atribut untuk data training dan 114 record dan 4 atribut untuk data testing. Seperti terlihat pada gambar 3.

```
X_train_uncorr = X_train_unique.drop(labels=corr_features, axis = 1)
X_test_uncorr = X_test_unique.drop(labels = corr_features, axis = 1)
X_train_uncorr.shape, X_test_uncorr.shape

((455, 4), (114, 4))
```

Gambar 3 Fitur Berkorelasi

4. Hasil dari pengurangan dimensi menggunakan LDA .

Data yang dihasilkan setelah menggunakan LDA adalah sebanyak 455 record dan 1 atribut untuk data training dan 114 record dan 1 atribut untuk data testing, seperti terlihat pada gambar 4.

```
X_train_lda.shape, X_test_lda.shape

((455, 1), (114, 1))
```

Gambar 4 Hasil Reduksi LDA

5. Hasil menjalankan Time Running

Dari hasil percobaan yang dilakukan menjalankan time running pada proses setelah LDA dan sebelum LDA tampak pada gambar 5 dan gambar 6. Gambar 5 adalah hasil menggunakan LDA dan gambar 6 hasil tanpa menggunakan LDA.

```
%%time
run_LogisticRegression(X_train_lda, X_test_lda, y_train, y_test)

Accuracy on test set:
0.9122807017543859
CPU times: user 10.5 ms, sys: 0 ns, total: 10.5 ms
Wall time: 13 ms
```

Gambar 5 Waktu LDA

Waktu yang digunakan untuk menjalankan model logistik regresi dengan menggunakan LDA sebesar 13 ms, terlihat pada gambar 5.

```
%%time
run_LogisticRegression(X_train, X_test, y_train, y_test)

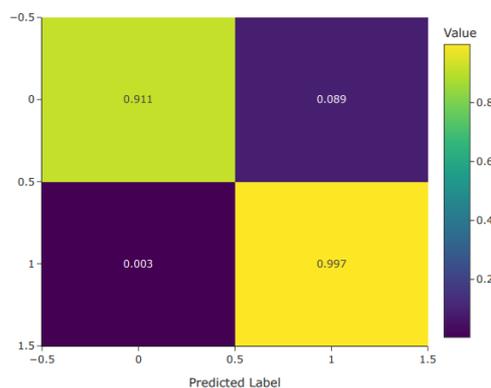
Accuracy on test set:
0.9473684210526315
CPU times: user 54.5 ms, sys: 72.1 ms, total: 127 ms
Wall time: 95 ms
```

Gambar 6 Waktu Tanpa LDA

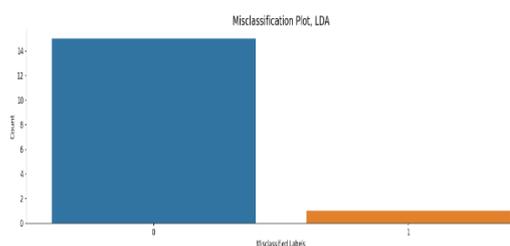
Waktu yang digunakan untuk menjalankan model logistik regresi tanpa LDA adalah selama 95 ms seperti terlihat pada gambar 6.

6. Hasil Prediksi dengan LDA

Dengan menggunakan konfousius matrik, hasil prediksi tampak pada gambar 7. Nilai yang didapat true positive (TP) sebesar 0.911, true negative (TN) sebesar 0.003, false positive (FP) sebesar 0.089, dan false negative (FN) sebesar 0.997.



Gambar 2. Hasil Prediksi LDA



Gambar 3. Nilai Kesalahan Klasifikasi LDA

Kesalahan klasifikasi pada proses LDA terdapat 16 kasus dari 455 kasus yang ada, seperti terlihat pada gambar 8.

7. Hasil Prediksi dengan Regresi Logistik.

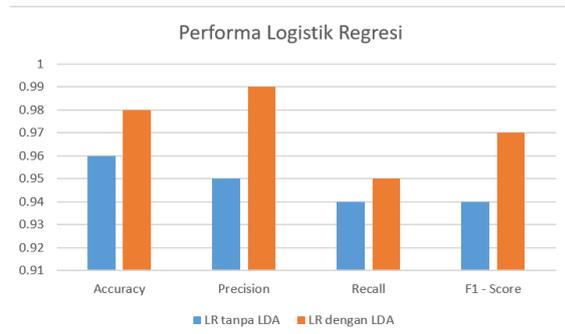
Performa dari model diukur dengan algoritma logistik dengan menggunakan konfosis matrik. Hasil yang didapat berupa nilai akurasi, presicion, recall, dan f1_score. Hasil dari evaluasi model regresi logistick setelah melalui proses reduksi dimensi dengan menggunakan LDA dapat dilihat pada gambar 9. Gambar 9 menunjukkan nilai akurasi sebesar 98%.

```
print(classification_report(y_train, LRtrain_preds))
```

	precision	recall	f1-score	support
0	0.99	0.95	0.97	169
1	0.97	0.99	0.98	286
accuracy			0.98	455
macro avg	0.98	0.97	0.98	455
weighted avg	0.98	0.98	0.98	455

Gambar 9 Performa LR dengan LDA

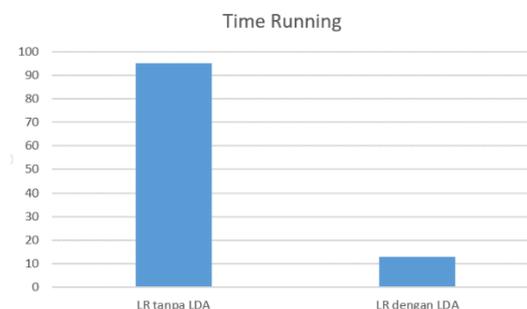
2. Pembahasan



Gambar 4. Perbandingan Performa LR

Gambar 10 merupakan tampilan hasil yang didapat model logistik regresi menggunakan LDA maupun tanpa menggunakan LDA . Hasil yang didapat berupa nilai akurasi , precision, recall, dan

f1-score. Nilai masing – masing dapat dilihat pada gambar 10



Gambar 5. Performa Waktu Tempuh LR

Gambar 11 menampilkan perbandingan waktu tempuh untuk menjalankan model logistik regresi.

5. Kesimpulan

Penelitian tentang reduksi dimensi menggunakan LDA pada model algoritma logistik regresi dengan menggunakan bahasa pemrograman python didapat performa menggunakan LDA, nilai akurasi sebesar 98%, presicion sebesar 99%, recall sebesar 95%, dan f1 – score sebesar 97%. Untuk nilai performa tanpa menggunakan LDA didapat nilai akurasi sebesar 96%, presicion sebesar 95%, recall sebesar 94%, dan f1 – score sebesar 94%. *Time running* menjalankan program pada logistik regresi dengan LDA selama 13 ms dan tanpa LDA selama 95 ms.

Dari hasil penelitian yang didapat maka dapat disimpulkan bahwa *linear discriminant analist* (LDA) dapat meningkat performa kinerja dari algoritma logistik regresi dalam menangani dataset yang memiliki dimensi tinggi

6. Daftar Pustaka

- [1] S. Ketut, "Kanker payudara: Diagnostik, Faktor Risiko dan Stadium," *Ganesha Med. J.*, vol. 2, no. 1, pp. 2–7, 2022.
- [2] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTI/K/article/view/141/156>
- [3] P. Xu, "Review on Studies of Machine Learning Algorithms," *J. Phys. Conf. Ser.*, vol. 1187, no. 5, 2019, doi: 10.1088/1742-6596/1187/5/052103.
- [4] Erlin, Yulvia Nora Marlim, Junadhi, Laili Suryati, and Nova Agustina, "Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 11, no. 2, pp. 88–96, 2022, doi: 10.22146/jnteti.v11i2.3586.
- [5] F. Dewanta and M. Abdillah, "Klasifikasi Beban Listrik dengan Machine Learning Menggunakan Metode K- Nearest Neighbor," vol. 5, no. 2, pp. 163–172.
- [6] D. Nugraha, P. Novantara, and A. Muhammad, "Implementasi Algoritma PCA pada Pengenalan Pola Tanda Tangan dengan Menggunakan Bahasa Pemrograman Phyton," *JEJARING J. Teknol. dan Manaj. Inform.*, vol. 6, no. 1, pp. 7–12, 2021, doi: 10.25134/jejaring.v6i1.6734.
- [7] R. Kosasih, "Penggunaan Metode Linear Discriminant Analysis Untuk Pengenalan Wajah Dengan Membandingkan Banyaknya Data Latih," *J. Ilm. Teknol. dan Rekayasa*, vol. 26, no. 1, pp. 25–34, 2021, doi: 10.35760/tr.2021.v26i1.3520.
- [8] N. Chintalapudi *et al.*, "LASSO Regression Modeling on Prediction of Medical Terms among Seafarers' Health Documents Using Tidy Text Mining," *Bioengineering*, vol. 9, no. 3, pp. 1–14, 2022, doi: 10.3390/bioengineering9030124.
- [9] D. Nurmalasari and W. Nengsih, "Sistem Rekomendasi Referensi Jurnal Ilmiah dengan Metode Frequent Itemset," vol. 5, no. 2, pp. 176–181, 2021.
- [10] A. Dinanti and J. Purwadi, "Analisis Performa Algoritma K-Nearest Neighbor dan Reduksi Dimensi Menggunakan Principal Component Analysis," vol. 5, no. 1, pp. 155–165, 2023.
- [11] F. Fitrianiingsih and S. Sugiyarto, "Implementasi Analisa Komponen Utama untuk Mereduksi Variabel yang Mempengaruhi Perbaikan pada Fungsi Ginjal Tikus," *J. Ilm. Mat.*, vol. 6, no. 2, p. 62, 2019, doi: 10.26555/konvergensi.v6i2.19549.
- [12] T. Informasi *et al.*, "Jurnal Pengenalan Wajah menggunakan Principle Component Analysis (PCA) dengan Model Algoritma Machine Learning untuk Mengidentifikasi Jenis Kelamin pada Kartu Identitas Mahasiswa Face Recognition using Principle Component Analysis (PCA) with Machine L," vol. 5, pp. 219–224, 2023.
- [13] R. Destriana, D. Numaningsih, and ..., "Implementasi Metode Linear Discriminant Analysis (LDA) Pada Klasifikasi Tingkat Kematangan Buah Nanas," ... *Sci.*, vol. 3, no. 1, pp. 56–63, 2021, doi: 10.47065/bits.v3i1.1007.
- [14] M. S. Tjahaya, Raupong, and G. M. Tinungki, "Analisis Dsikriminan Linear Robust Dengan Metode Winsorized Modified One-Step M-Estimator," *J. Estimasi*, vol. 3, no. 1, pp. 1–13, 2022, doi: 10.20956/ejsa.vi.11302.
- [15] F. A. D. Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, "Metode-metode Klasifikasi," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, p. 134, 2018.
- [16] H. Tan, "Machine Learning Algorithm for Classification," *J. Phys. Conf. Ser.*, vol. 1994, no. 1, 2021, doi: 10.1088/1742-6596/1994/1/012016.
- [17] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A

- detailed tutorial," *AI Commun.*, vol. 30, no. 2, pp. 169–190, 2017, doi: 10.3233/AIC-170729.
- [18] P. Boedeker and N. T. Kearns, "Linear Discriminant Analysis for Prediction of Group Membership: A User-Friendly Primer," *Adv. Methods Pract. Psychol. Sci.*, vol. 2, no. 3, pp. 250–263, 2019, doi: 10.1177/2515245919849378.
- [19] I. Rodrigues, A. Parayil, T. Shetty, and I. Mirza, "Use of Linear Discriminant Analysis (LDA), K Nearest Neighbours (KNN), Decision Tree (CART), Random Forest (RF), Gaussian Naive Bayes (NB), Support Vector Machines (SVM) to Predict Admission for Post Graduation Courses," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3683065.
- [20] G. O. Yusuf, A. K. Jaya, and N. Ilyas, "Pemodelan Regresi Logistik Menggunakan Metode Momen Diperumum," *ESTIMASI J. Stat. Its Appl.*, vol. 1, no. 2, p. 74, 2020, doi: 10.20956/ejsa.v1i2.9304.
- [21] A. A. T. Fernandes, D. B. F. Filho, E. C. da Rocha, and W. da Silva Nascimento, "Read this paper if you want to learn logistic regression," *Rev. Sociol. e Polit.*, vol. 28, no. 74, pp. 1/1-19/19, 2020, doi: 10.1590/1678-987320287406EN.
- [22] Z. A. A. Al-Bairmani and A. A. Ismael, "Using Logistic Regression Model to Study the Most Important Factors Which Affects Diabetes for the Elderly in the City of Hilla / 2019," *J. Phys. Conf. Ser.*, vol. 1818, no. 1, 2021, doi: 10.1088/1742-6596/1818/1/012016.