

Perbandingan Metode Klasifikasi *Learning Vector Quantization* Dengan *Diskriminan Fisher* Pada Data Bunga Iris

Dani Rofianto¹, Oki Arifin², Dewi Kania Widyawati³

^{1,2,3}Jurusan Ekonomi dan Bisnis, Program Studi Manajemen Informatika, Politeknik Negeri Lampung, Bandar Lampung, Indonesia

¹danirofiyanto@polinela.ac.id, ²okiarifin@polinela.ac.id, ³dewi_mi@polinela.ac.id

Abstrak

Klasifikasi merupakan masalah penting dalam dunia ilmu data, dimana model dibangun untuk memprediksi label atau kelas pada suatu data berdasarkan fitur-fitur yang dimilikinya. Pembangunan model dan penggunaannya untuk melakukan klasifikasi adalah pekerjaan utama dalam klasifikasi. Berbagai metode klasifikasi telah dikembangkan, termasuk *Learning Vector Quantization* (LVQ) dan *Diskriminan Fisher*. Penelitian ini bertujuan untuk membandingkan metode klasifikasi LVQ dengan *Diskriminan Fisher* pada data bunga iris untuk menentukan metode yang lebih efektif dan efisien dalam melakukan klasifikasi. Hasil penelitian menunjukkan bahwa kedua metode mampu melakukan klasifikasi dengan akurasi yang tinggi. Namun, terdapat perbedaan dalam kinerja kedua metode tersebut. Metode LVQ menghasilkan akurasi tertinggi sebesar 96.29% dengan waktu proses 4.56 *second*, sedangkan metode *Diskriminan Fisher* mampu menghasilkan akurasi sebesar 100% dengan waktu proses 0.1 *second*. Hal ini menunjukkan bahwa metode *Diskriminan Fisher* lebih unggul dalam melakukan klasifikasi pada data bunga iris. Selain itu, hasil pengukuran waktu eksekusi menunjukkan bahwa metode *Diskriminan Fisher* lebih cepat dalam melakukan klasifikasi dibandingkan dengan metode LVQ.

Kata Kunci: *Learning Vector Quantization*, *Diskriminan Fisher*, Algoritma, Klasifikasi, Dataset Iris

Abstract

Classification is the main problem in data science, where models are built to predict labels or classes in data based on its features. Model building and its use to accomplish classification is the main work in classification. Various classification technique have been developed, including Learning Vector Quantization (LVQ) and Fisher's Discriminant. This study aims to analyze the LVQ classification and Fisher's discriminant on iris data to determine a more effective and efficient technique classification. The results of the study show that both technique are capable of classifying data with adequately high accuracy. However, there are differences in the performance of the two methods. The LVQ method produces the highest accuracy of 96.29% with a processing time of 4.56 seconds, while the Fisher Discriminant method produces an accuracy of 100% with a processing time of only 0.1 second. It shows that the Fisher Discriminant technique is superior in classifying iris data. In addition, the results of the execution time measurement show that the Fisher Discriminant technique is faster in classifying than the LVQ method.

Keywords: Learning Vector Quantization, Diskriminan Fisher, Algorithm, Classification, Iris Dataset

1. PENDAHULUAN

Dalam dunia ilmu data, klasifikasi merupakan salah satu masalah penting yang banyak dihadapi [1], [2]. Klasifikasi dapat digunakan untuk memprediksi label atau kelas pada suatu data berdasarkan fitur-fitur yang dimiliki oleh data tersebut [3], [4]. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan atau klasifikasi atau prediksi pada suatu objek data lain

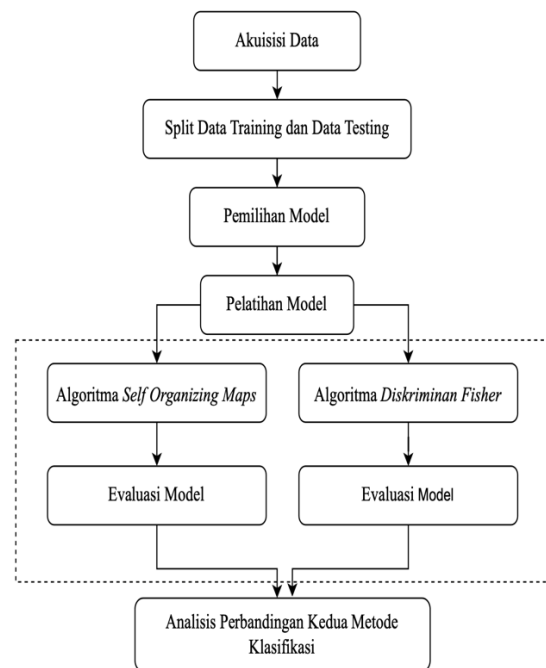
agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya [5], [6]. Terdapat berbagai metode klasifikasi yang telah dikembangkan hingga saat ini, salah satunya adalah *Learning Vector Quantization* (LVQ) dan *Diskriminan Fisher*. Metode klasifikasi LVQ merupakan sebuah jaringan syaraf dengan arsitektur jaringan lapis tunggal umpan-maju (*single layer feed forward*). LVQ dikenal sebagai salah satu metode pembelajaran pada lapisan kompetitif yang terawasi (*supervised*), dimana lapisan kompetitif akan secara otomatis belajar untuk mengklasifikasikan vektor-vektor input [7], [8]. Kelas-kelas yang dihasilkan dari lapisan kompetitif

ini bergantung pada jarak antara vektor-vektor input [9]. LVQ merupakan metode pengelompokan di mana target/kelas dari setiap kelompok sudah ditentukan, tujuan dari algoritma ini adalah mendekati distribusi kelas vektor untuk meminimalkan kesalahan dalam pengklasifikasi [10]. Algoritma *Diskriminan Fisher* (*Fisher's Discriminant Analysis*) adalah sebuah algoritma klasifikasi yang bersifat *supervised*. Artinya, algoritma ini membutuhkan label atau kelas pada data latih sebagai masukan untuk melakukan pembelajaran dan membedakan kelas-kelas pada data uji [11], [12]. Tujuan utama dari algoritma ini adalah untuk menemukan kombinasi linear dari fitur-fitur dalam data yang paling baik membedakan kelas-kelas yang diberikan [13]. Kedua metode ini telah terbukti efektif dalam melakukan klasifikasi pada berbagai jenis data. Namun, setiap metode klasifikasi memiliki kelebihan dan kekurangan masing-masing. Oleh karena itu, diperlukan perbandingan antara metode-metode klasifikasi yang berbeda untuk menentukan metode yang lebih baik dalam melakukan klasifikasi pada suatu data. Salah satu data yang sering digunakan untuk membandingkan metode klasifikasi adalah data bunga iris [14].

Data bunga iris terdiri dari tiga jenis spesies bunga iris, yaitu *iris setosa*, *iris versicolor*, dan *iris virginica*. Setiap spesies bunga iris memiliki ciri-ciri khusus yang membedakannya, seperti panjang dan lebar kelopak bunga serta panjang dan lebar mahkota bunga [15]. Oleh karena itu, data bunga iris sering digunakan sebagai contoh kasus dalam pembelajaran mesin dan pengujian algoritma klasifikasi. Berdasarkan pemaparan dari kedua metode diatas, penulis tertarik melakukan perbandingan antara metode klasifikasi *Learning Vector Quantization* (LVQ) dan *Diskriminan Fisher* pada data bunga iris. Melalui perbandingan ini, diharapkan dapat diketahui metode mana yang lebih efektif dan efisien dalam melakukan klasifikasi pada data bunga iris. Selain itu, penelitian ini juga dapat memberikan informasi yang bermanfaat dalam pengembangan metode-metode klasifikasi baru yang lebih baik di masa depan.

2. METODOLOGI PENELITIAN

Secara garis besar, metode yang digunakan pada penelitian ini yaitu akuisisi data, *split* atau membagi data menjadi data *training* dan *testing*, pemilihan model klasifikasi, pelatihan model, dan evaluasi yang dihasilkan. Alur lengkap dari tahapan metode penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

2.1 Akuisisi Data

Penelitian ini menggunakan objek penelitian berupa dataset bunga iris yang diperoleh dari *UCI Machine Learning Repository*. Iris adalah dataset standar dalam *machine learning* dan statistik yang sering digunakan untuk tujuan pengujian dan pembelajaran. Dataset ini terdiri dari 150 sampel bunga iris yang dikumpulkan dari tiga jenis spesies iris yaitu *iris setosa*, *iris versicolor*, dan *iris virginica*. Setiap sampel memiliki empat *atribut*, yaitu panjang dan lebar kelopak dan mahkota bunga.

2.2 Split Data Training dan Testing

Tahap ini melibatkan pembagian data menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk melatih model klasifikasi, sedangkan data uji digunakan untuk menguji kinerja model.

2.3 Pemilihan Model

Pada tahapan ini memilih algoritma klasifikasi yang sesuai dengan jenis data dan tujuan klasifikasi yang ingin dicapai. Dalam penelitian ini dua metode yang telah ditetapkan yaitu *Learning Vector Quantization* dan *Diskriminan Fisher*.

2.4 Pelatihan Model

Tahapan ini melibatkan proses melatih model klasifikasi menggunakan data latih. Pada tahap ini, model klasifikasi *Learning Vector Quantization* dan *Diskriminan Fisher* dipelajari untuk mengenali pola dan membuat prediksi berdasarkan *atribut* yang diberikan. Selanjutnya dalam pelatihan model ini akan dilakukan beberapa modifikasi parameter

untuk melihat seberapa efektif akurasi yang dihasilkan.

2.5 Evaluasi Model

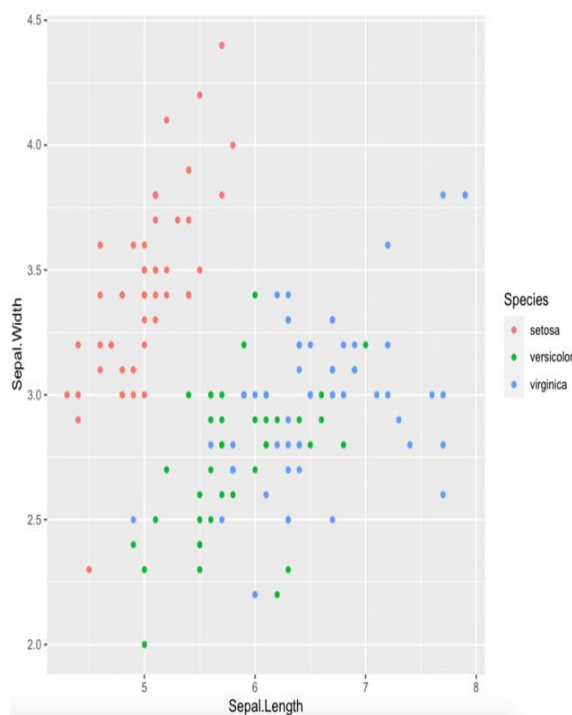
Tahap ini melibatkan evaluasi kinerja model klasifikasi pada masing-masing algoritma menggunakan data uji. Kinerja model dievaluasi dengan menghitung tingkat akurasi menggunakan *confusion matrix*.

2.6 Hasil Perbandingan Kedua Metode

Tahap terakhir yaitu membandingkan hasil testing pada kedua algoritma yaitu *Learning Vector Quantization* dan *Diskriminan Fisher*. Hasil dari eksperimen yang telah dilakukan selengkapnya akan dibahas pada hasil dan pembahasan.

3. HASIL DAN PEMBAHASAN

Dalam bab ini, fokus penelitian adalah membandingkan hasil klasifikasi dari dua algoritma klasifikasi yang berbeda, yaitu *Learning Vector Quantization* dan *Diskriminan Fisher*. Penelitian ini menggunakan data bunga iris sebagai objek penelitian yang diperoleh dari *UCI Machine Learning Repository*. Data bunga iris terdiri dari tiga jenis spesies bunga iris, yaitu *iris setosa*, *iris versicolor*, dan *iris virginica*. Pengolahan data dalam penelitian ini dibantu menggunakan *software R Studio* dengan bahasa pemrograman R. Berikut ini sebaran pola dataset Iris setelah diolah menggunakan R Studio, seperti yang ditampilkan pada gambar 2.



Gambar 2. Sebaran dataset bunga iris

Metode *Learning Vector Quantization* (LVQ) dan *Fisher Discriminant Analysis* adalah dua metode klasifikasi yang berbeda dalam pendekatan dan implementasi. Meskipun keduanya bertujuan untuk memprediksi kelas dari sebuah sampel, terdapat perbedaan dalam cara kedua metode tersebut memproses informasi dan mengklasifikasikan sampel. Proses pelatihan LVQ berfokus pada penyesuaian neuron pada output layer sehingga setiap neuron merepresentasikan kelas yang tepat dan meminimalkan *error* dari klasifikasi. Pada saat pengujian, sampel baru diuji dan diklasifikasikan berdasarkan jarak *euclidean* dengan neuron terdekat pada output layer. Sementara itu, pada metode *Fisher*, fokus utama adalah pada pemilihan dimensi/fitur yang paling relevan dan signifikan dalam membedakan kelas pada dataset. Proses klasifikasi pada *Fisher* dilakukan dengan menghitung *diskriminan* pada fitur yang paling *signifikan* dan memisahkan kelas pada dataset. Pada saat pengujian, sampel baru diuji dan diklasifikasikan berdasarkan *diskriminan* yang telah dihitung. Berikut ini paparan lebih lanjut mengenai hasil klasifikasi pada masing-masing algoritma yang disajikan dalam Tabel 1, Tabel 2, Tabel 3, dan Tabel 4.

Tabel 1. Hasil klasifikasi algoritma *Learning Vector Quantization*

Ekspe- rimen	Pro- porsi	Iterasi	Aku- rasi	Running Time (second)
1	60 : 40	4	94.12%	4.04
2	70 : 30	4	94.73%	4.21
3	80 : 20	4	96.29%	4.56
4	60 : 40	8	94.12%	7.83
5	70 : 30	8	94.73%	8.32
6	80 : 20	8	96.29%	8.44

Tabel 2. Kinerja algoritma *Learning Vector Quantization* yang dievaluasi menggunakan *confusion matrix*

Ekspe- rimen		Set- osa	Versi- color	Vir- ginica
1	Setosa	19	1	0
	Versicolor	0	11	1
	Virginica	0	1	18
2	Setosa	10	1	0
	Versicolor	0	10	0
	Virginica	0	1	16

3	<i>Setosa</i>	8	0	0
	<i>Versicolor</i>	0	6	0
	<i>Virginica</i>	0	1	12
4	<i>Setosa</i>	19	1	0
	<i>Versicolor</i>	0	11	1
	<i>Virginica</i>	0	1	18
5	<i>Setosa</i>	8	0	0
	<i>Versicolor</i>	0	6	0
	<i>Virginica</i>	0	1	12
6	<i>Setosa</i>	8	0	0
	<i>Versicolor</i>	0	6	0
	<i>Virginica</i>	0	1	12

Tabel 3. Hasil Klasifikasi algoritma *Diskriminan Fisher*

Ekспери- men	Pro- por-si	Dim- ensi	Aku- ra-si	Running Time (second)
1	60:40	4	96.67%	0.113
2	70:30	4	97.78%	0.091
3	80:20	4	96.67%	0.111
4	60:40	8	100 %	0.100
5	70:30	8	97.78%	0.107
6	80:20	8	96.67%	0.101

Tabel 4. Kinerja algoritma *Diskriminan Fisher* yang dievaluasi menggunakan *confusion matrix*

Ekspere- rimen		<i>Set- osa</i>	<i>Versi- color</i>	<i>Vir- ginica</i>
1	<i>Setosa</i>	20	0	0
	<i>Versicolor</i>	0	18	0
	<i>Virginica</i>	0	2	20
2	<i>Setosa</i>	15	0	0
	<i>Versicolor</i>	0	15	1
	<i>Virginica</i>	0	1	14
3	<i>Setosa</i>	10	0	0
	<i>Versicolor</i>	0	10	1
	<i>Virginica</i>	0	0	9

4	<i>Setosa</i>	20	0	0
	<i>Versicolor</i>	0	20	0
	<i>Virginica</i>	0	0	20
5	<i>Setosa</i>	15	0	0
	<i>Versicolor</i>	0	14	0
	<i>Virginica</i>	0	1	15
6	<i>Setosa</i>	10	0	0
	<i>Versicolor</i>	0	9	0
	<i>Virginica</i>	0	1	10

Berdasarkan beberapa eksperimen yang telah dilakukan pada tabel 1, terlihat bahwa jumlah iterasi pada klasifikasi LVQ tidak berpengaruh terhadap akurasi. Hal ini disebabkan oleh fakta bahwa algoritma LVQ cukup stabil dalam mencapai titik konvergensi pada iterasi awal, sehingga tidak diperlukan banyak iterasi untuk mencapai akurasi yang optimal. Namun demikian, jumlah iterasi pada LVQ cukup mempengaruhi waktu proses algoritma. Semakin banyak iterasi yang dilakukan, semakin lama waktu yang dibutuhkan untuk mengklasifikasikan data. Oleh karena itu, dalam praktiknya jumlah iterasi pada LVQ perlu diatur dengan baik agar tidak memakan waktu yang terlalu lama, namun tetap menghasilkan akurasi klasifikasi yang memadai. Berdasarkan eksperimen klasifikasi dengan metode LVQ diperoleh akurasi terbaik pada proporsi data sebesar 80:20 dengan akurasi sebesar 96.29% dengan waktu proses sebesar 4.56 second.

Pada klasifikasi dengan algoritma *Diskriminan Fisher*, diperoleh hasil seperti yang disajikan pada tabel 3. Berdasarkan tabel tersebut proporsi data 60:40 menghasilkan akurasi terbaik dari semua eksperimen yang dilakukan, yaitu dengan akurasi sebesar 100%. Dapat disimpulkan bahwa jumlah proporsi data serta inisiasi dimensi awal berpengaruh signifikan terhadap akurasi yang dihasilkan. Dari beberapa eksperimen ini algoritma *Diskriminan Fisher* bekerja sangat baik dalam melakukan klasifikasi pada dataset bunga iris. Selanjutnya untuk melihat dengan lebih jelas perbandingan hasil klasifikasi, maka kami sajikan tabel perbandingan kedua metode klasifikasi dengan atribut proporsi data, akurasi dan *running time*. Hasil perbandingan pengujian lebih lengkap dapat dilihat pada Tabel 5.

Tabel 5. Hasil perbandingan kinerja kedua metode

Algo- ritma	Eksp- erimen	Pro- por-si	Akurasi	Running Time (second)
----------------	-----------------	----------------	---------	-----------------------------

LVQ	1	60 : 40	94.12%	4.04
	2	70 : 30	94.73%	4.21
	3	80 : 20	96.29%	4.56
	4	60 : 40	94.12%	7.83
	5	70 : 30	94.73%	8.32
	6	80 : 20	96.29%	8.44
Fisher	1	60:40	96.67%	0.113
	2	70:30	97.78%	0.091
	3	80:20	96.67%	0.111
	4	60:40	100 %	0.100
	5	70:30	97.78%	0.107
	6	80:20	96.67%	0.101

Berdasarkan hasil perbandingan kinerja kedua metode pada tabel 5, dapat disimpulkan bahwa performa metode klasifikasi yang telah dilakukan mampu melakukan klasifikasi pada dataet iris dengan akurasi yang tinggi. Namun, terdapat perbedaan dalam kinerja kedua metode tersebut. Metode LVQ menghasilkan akurasi tertinggi sebesar 96.29% dengan waktu proses 4.56 *second*, sedangkan metode *Diskriminan Fisher* menghasilkan akurasi sebesar 100% dengan waktu proses hanya 0.1 *second*. Hal ini menunjukkan bahwa metode *Diskriminan Fisher* lebih unggul dalam melakukan klasifikasi pada data bunga iris. Selain itu, hasil pengukuran waktu eksekusi menunjukkan bahwa metode *Diskriminan Fisher* lebih cepat dalam melakukan klasifikasi dibandingkan dengan metode LVQ. Namun, perlu diingat bahwa hasil tersebut tidak dapat langsung digeneralisasi ke data lainnya, karena kinerja metode klasifikasi sangat bergantung pada karakteristik dari data yang digunakan.

4. KESIMPULAN

Berdasarkan beberapa eksperimen yang dilakukan dengan menggunakan kedua metode dapat ditarik suatu kesimpulan bahwa proporsi data *training* dan *testing* menentukan hasil akurasi klasifikasi. Proporsi data sebesar 80:20 pada metode LVQ memberikan akurasi sebesar 96.29% dengan waktu proses sebesar 4.56 *second*, sedangkan pada pengklasifikasian dengan metode *Diskriminan Fisher* proporsi data 60:40 menghasilkan akurasi terbaik dari semua eksperimen, yaitu dengan akurasi 100%. Dapat disimpulkan bahwa perubahan proporsi data serta

inisiasi dimensi awal berpengaruh *signifikan* terhadap akurasi yang diperoleh. Metode klasifikasi *Diskriminan Fisher* lebih efektif dan efisien dalam melakukan klasifikasi pada data bunga iris dibandingkan dengan metode LVQ. Namun, kinerja metode klasifikasi sangat bergantung pada karakteristik dari data yang digunakan. Oleh karena itu, perlu dilakukan penelitian lebih lanjut untuk membandingkan kedua metode klasifikasi pada data lainnya dengan karakteristik yang berbeda.

5. REFERENCES

- [1] A. Jain, D. Somwanshi, K. Joshi, and S. S. Bhatt, "A Review: Data Mining Classification Techniques," in *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, Apr. 2022, pp. 636–642. doi: 10.1109/ICIEM54221.2022.9853036.
- [2] Qian Li, Hao Peng, M. Dontcheva, and K. Karahalios, "A Survey on Text Classification: From Traditional to Deep Learning," *Proc ACM Hum Comput Interact*, vol. 5, no. CSCW1, Apr. 2021, doi: 10.1145/1122445.1122456.
- [3] B. K. Varghese, D. Amali, and K. S. Devi, "Prediction of parkinson's disease using machine learning techniques on speech dataset," *Res J Pharm Technol*, vol. 12, no. 2, pp. 644–648, 2019.
- [4] P. M. Usha and N. V Balaji, "A comparative study on machine learning algorithms for employee attrition prediction," *IOP Conf Ser Mater Sci Eng*, vol. 1085, no. 1, p. 012029, Feb. 2021, doi: 10.1088/1757-899X/1085/1/012029.
- [5] O. Arifin, K. Saputra, and H. Fathoni, "Implementation of Data Mining Using Naïve Bayes Classifier in Food Crop Prediction," *Scientific Journal of Informatics*, vol. 8, no. 1, 2021, doi: 10.15294/sji.v8i1.xxxxx.
- [6] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Rehan, "A comparative review on deep learning models for text classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, p. 325, Jul. 2020, doi: 10.11591/ijeecs.v19.i1.pp325-335.
- [7] Y.-Y. Shen, Y.-M. Zhang, X.-Y. Zhang, and C.-L. Liu, "Online semi-supervised learning with learning vector quantization," *Neurocomputing*, vol. 399, pp. 467–478, Jul. 2020, doi: 10.1016/j.neucom.2020.03.025.
- [8] E. Akarslan, "Learning Vector Quantization based predictor model selection for hourly load

- demand forecasting,” *Appl Soft Comput*, vol. 117, p. 108421, Mar. 2022, doi: 10.1016/j.asoc.2022.108421.
- [9] Y. Aprizal, R. I. Zainal, and A. Afriyudi, “Perbandingan Metode Backpropagation dan Learning Vector Quantization (LVQ) Dalam Menggali Potensi Mahasiswa Baru di STMIK PalComTech,” *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 18, no. 2, pp. 294–301, May 2019, doi: 10.30812/matrik.v18i2.387.
- [10] Arnita, M. S. Sinaga, and Elmanani, “Classification and diagnosis of diabetic with neural network algorithm learning vector quantization (LVQ),” in *Journal of Physics: Conference Series*, Apr. 2019, vol. 1188, no. 1. doi: 10.1088/1742-6596/1188/1/012091.
- [11] V. E. de Almeida *et al.*, “Scores selection via Fisher’s discriminant power in PCA-LDA to improve the classification of food data,” *Food Chem*, vol. 363, p. 130296, Nov. 2021, doi: 10.1016/j.foodchem.2021.130296.
- [12] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, “Robust discriminant analysis with feature selective projection and between-classes structural incoherence,” *Digit Signal Process*, vol. 134, p. 103896, Apr. 2023, doi: 10.1016/j.dsp.2022.103896.
- [13] D. AbuZeina and F. S. Al-Anzi, “Employing fisher discriminant analysis for Arabic text classification,” *Computers & Electrical Engineering*, vol. 66, pp. 474–486, Feb. 2018, doi: 10.1016/j.compeleceng.2017.11.002.
- [14] M. A. Afifi, T. M. Ghazal, M. A. M. Afifi, and D. Kalra, “Data Mining and Exploration: A Comparison Study among Data Mining Techniques on Iris Data Set Educational Revamp View project E-Governance View project Data Mining and Exploration: A Comparison Study among Data Mining Techniques on Iris Data Set,” *Talent Development & Excellence*, vol. 12, no. 1, pp. 3854–3861, 2020, [Online]. Available: <http://www.iratde.com>
- [15] P. Putra, A. M. H Pardede, and S. Syahputra, “Analisis Metode K-Nearest Neighbour (Knn) Dalam Klasifikasi Data Iris Bunga,” *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 6, no. 1, 2022.