

Penerapan Metode Support Vector Machine Dalam Klasifikasi Bunga Iris

Anita Desiani*, Irmeilyana, Herlina Hanum, Yuli Andriani, Sri Indra Maiyanti, Clarita Margo Uteh, Ira Rayani
Jurusan Matematika FMIPA Universitas Sriwijaya

*Email: anita_desiani@unsri.ac.id

Info Artikel

Kata Kunci :

Support Vector Machine,
Klasifikasi, Bunga Iris

Keywords :

Support Vector Machine,
Classification, Iris Flowers

Tanggal Artikel

Dikirim : 26 Mei 2022

Direvisi : 27 September 2022

Diterima : 30 November 2022

Abstrak

Data mining adalah proses melatih komputer untuk mengenali suatu pola menggunakan teknik statistika maupun matematika. Salah satu teknik data mining yang sering digunakan adalah klasifikasi, yakni mengelompokkan data ke dalam suatu label menggunakan atribut. Pada klasifikasi, Support Vector Machine (SVM) merupakan salah satu metode yang paling banyak digunakan. Penelitian ini akan memanfaatkan metode SVM dalam melakukan klasifikasi bunga Iris. Data yang diteliti menggunakan sebanyak 150 data dengan menggunakan dua metode data latih, yakni percentage split dan k-fold cross validation. Data diolah melalui tahap *pre-processing*, lalu diklasifikasi menggunakan metode SVM melalui 2 metode data latih, percentage split sebesar 80% dan k-fold corss validation dengan $k=10$, perhitungan hasil prediksi menggunakan *confusion matrix*. Pada metode *percentage split* diperoleh nilai akurasi sebesar 96,7%, presisi 97,6%, *recall* sebesar 95,3%, dan F1-score sebesar 96,3%. Pada metode k-fold cross validation diperoleh nilai akurasi sebesar 92,6%, presisi 92,6%, *recall* sebesar 92,6%, dan F1-score sebesar 92,3%. Dengan demikian metode SVM menggunakan kernel polynomial dengan metode data latih percentage split dapat diimplementasikan ke dalam sistem klasifikasi bunga Iris.

Abstract

Data mining is the process of training a computer to recognize a pattern using statistical and mathematical techniques. One of the data mining techniques that are often used is classification, which is to group data into the label using attributes. In classification, the Support Vector Machine (SVM) is one of the most widely used methods. This research will utilize the SVM method in classifying Iris flowers. The data studied used 150 data using two training data methods, percentage split and k-fold cross validation. The data is processed through the pre-processing stage, then classified using the SVM method through 2 training data methods, percentage split of 80% and k-fold cross validation with $k = 10$, and calculation of prediction results using a confusion matrix. In the percentage split method, the accuracy is 96.7%, precision is 97.6%, recall is 95.3%, and F1-score is 96.3%. In the k-fold cross validation method, the accuracy is 92.6%, precision is 92.6%, recall is 92.6%, and F1-score is 92.3%. So that the SVM method using a polynomial kernel with the percentage split training data method can be implemented into the iris classification system.

1. PENDAHULUAN

Data mining merupakan proses atau teknik pembelajaran komputer dalam menganalisa pengetahuan secara otomatis dengan melatih komputer menggunakan teknik pengenalan pola seperti teknik Statistik dan Matematika sehingga ditemukan pola baru agar mudah digunakan oleh komputer [1,2]. Salah satu teknik pengolahan data mining yang sering digunakan adalah klasifikasi.

Klasifikasi merupakan teknik dalam data mining yang digunakan untuk mengurutkan dan mengelompokkan data ke dalam suatu label atau kelas dengan membedakannya menggunakan atribut yang tersedia [3,4]. Salah satu contoh metode pada klasifikasi adalah *Support Vector Machine* (SVM).

Metode SVM adalah sistem pembelajaran komputer untuk memberikan label ke suatu objek tertentu dengan memaksimalkan fungsi matematika berupa fungsi linier dalam sebuah ruang fitur yang memiliki dimensi tinggi [5,6]. Salah satu kelebihan dari metode SVM, yakni mampu menghasilkan klasifikasi dengan akurasi yang tinggi [5]. Hal ini didukung dengan penelitian yang telah dilakukan sebelumnya, antara lain pengkategorian penelitian [5] dengan nilai akurasi 90%, klasifikasi tweet [7] dengan akurasi sebesar 97,54%, klasifikasi kualitas pengelasan SMAW [8] dengan akurasi sebesar 96,2%.

Metode SVM awalnya hanya dapat mengklasifikasi pada data linier, namun setelah dikembangkan, SVM juga dapat melakukan klasifikasi pada data non linier menggunakan pendekatan kernel, salah satunya ialah kernel *polynomial* [9]. Pengklasifikasian metode SVM menggunakan kernel *polynomial* telah dilakukan oleh beberapa peneliti, antara lain mendeteksi intrusi [10] dengan akurasi sebesar 89,81%, klasifikasi akreditasi Sekolah Dasar [9] dengan akurasi sebesar 92,683%, dan klasifikasi sentimen tweet public figure [11] dengan akurasi sebesar 72%. Kelebihan dari metode SVM adalah metode ini memiliki landasan teori yang berbasis statistic, implementasi yang cukup mudah, dan bersifat generalisasi.

Bunga Iris merupakan bunga hias yang berasal dari Jepang yang memiliki bermacam-macam warna dan bentuk, dan dapat tumbuh di darat maupun air seperti daerah gurun ataupun rawa [12,13,14,15]. Data bunga Iris pertama kali diteliti oleh Ronald Fisher pada tahun 1936 dengan memperoleh 150 data bunga Iris, 3 jenis bunga, yakni Iris Setosa, Iris Versicolor, dan Iris Virginica, serta 4 atribut yang terdiri dari panjang sepal, lebar sepal, panjang petal, dan lebar petal [16]. Data bunga Iris telah banyak digunakan oleh para peneliti dalam melakukan klasifikasi menggunakan berbagai metode, metode *K-Nearest Neighbour* (KNN) [4] dengan akurasi sebesar 92%, algoritma *decision tree* [17] dengan akurasi sebesar 95,33%, model *Deep Neural Network* (DNN) [12] dengan akurasi 96%, dan metode kombinasi MLP-APSO [18] dengan akurasi 96%. Kekurangan dari beberapa penelitian yang dilakukan sebelumnya, yakni hanya menampilkan nilai akurasi saja.

Berdasarkan penelitian-penelitian sebelumnya dengan hasil yang akurasi yang baik, maka pada penelitian ini akan dilakukan pembuatan sebuah sistem klasifikasi bunga Iris menggunakan metode SVM dengan kernel *polynomial* serta membandingkan dua metode latihan data, yakni *percentage split* dan *k-fold cross validation* dengan mengukur hasil tingkat akurasi, tingkat presisi, *recall*, dan *F1-score* dengan harapan dapat memberikan hasil terbaik dibandingkan dengan metode klasifikasi lainnya dan dapat membandingkan metode data latihan terbaik diantara *percentage split* dan *k-fold cross validation*.

2. METODE PENELITIAN

2.1. Deskripsi Data

Data yang digunakan adalah dataset bunga iris yang bersumber dari *kaggle* dapat diakses melalui <https://www.kaggle.com/datasets/arshid/iris-flower-dataset>, total data bunga iris yang digunakan berjumlah 150 dengan lima atribut yaitu *Sepallength*, *Sepalwidth*, *Petalength*, *Petalwidth*, dan *Species* dengan masing-masing 50 data untuk setiap kelas *Iris Setosa*, *Iris Versicolor*, dan *Iris Virginica*. Ditampilkan masing-masing dua data teratas dari setiap kelas dataset bunga Iris pada Tabel 1.

Tabel 1. Entry Dataset

<i>Sepal Length</i>	<i>Sepal Width</i>	<i>Petal Length</i>	<i>Petal Width</i>	<i>Species</i>
5.1	3.5	1.4	0.2	Iris Setosa
4.9	3.0	1.4	0.2	Iris Setosa
7.0	3,2	4,7	1,4	Iris Versicolor
6.4	3.2	4.5	1.5	Iris Versicolor
6.3	3.3	6.0	2.5	Iris Virginica
5.8	2.7	5.1	1.9	Iris Virginica

2.2. Pre-processing Data

Pada *preprocessing data* dilakukan proses pengecekan *missing value* pada data bunga Iris. Selanjutnya, data akan dibagi menjadi 2 bagian, yaitu *data training* dan *data testing*. Data tersebut akan dilatih menggunakan 2 metode pelatihan data, *percentage split* dan *k-fold cross validation*. Sebesar 80% data *training* dan 20% data *testing*

digunakan pada *percentage split* dan nilai k sebesar 10 digunakan pada *k-fold cross validation* untuk pengujian dan pelatihan data.

2.3. Metode SVM

SVM merupakan metode *learning machine* yang bekerja menggunakan prinsip *Structural Risk Minimization* bertujuan untuk dapat menemukan *hyperplane* yang terbaik dalam memisahkan dua *class* pada *input space*. SVM menggunakan hipotesis fungsi linier dalam suatu ruang fitur yang berdimensi tinggi, kemudian mengimplementasikan *learning* bias yang berasal dari teori statistik [19].

SVM membagi dataset menjadi dua kelas yang dipisah oleh *hyperplane* kelas pertama bernilai (1) dan kelas kedua bernilai (-1)[19].

$$X_i \cdot W + b \geq 1 \text{ untuk } Y_i = 1 \quad (1)$$

$$X_i \cdot W + b \leq -1 \text{ untuk } Y_i = -1 \quad (2)$$

Rumus yang digunakan untuk menghitung hasil prediksi dengan *hyperplane* tunggal dan sesuai dengan b dan w yang akan diperoleh adalah sebagai berikut.

$$f(x) = \text{sign}(w \cdot \phi(x) + b) \quad (3)$$

$$f(x) = \text{sign} \left(\sum_{i=1}^n a_i y_i \phi(x_i)^T \cdot \phi(x) + b \right) \quad (4)$$

Keterangan :

X_i = data ke i

W = nilai bobot *support vector* yang tegak lurus dengan *hyperplane*

b = nilai bias

Y_i = kelas data ke i

Menentukan nilai optimal dari *hyperplane* dua kelas menggunakan persamaan sebagai berikut.

$$\text{Minimize } \int 1[w] = \frac{1}{2} \|w\|^2 \quad (5)$$

2.4. Evaluasi Hasil

Proses evaluasi hasil menggunakan *confusion matrix* dengan nilai akurasi, presisi, recall, dan *F1-score*. *Confusion matrix* merupakan tabel klasifikasi jumlah data uji yang bernilai benar dan salah [20]. Contoh *confusion matrix* klasifikasi biner pada Tabel 2.

Tabel 2. *Confusion Matrix*

	Prediction	
	Positive	Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Akurasi merupakan angka prediksi yang benar dibuat oleh model melalui kumpulan data, rumus menghitung akurasi sebagai berikut.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif, rumus menghitung presisi sebagai berikut.

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (7)$$

Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif, rumus menghitung *recall* sebagai berikut.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

F1-score digunakan untuk mengukur kombinasi nilai hasil dari presisi dan *recall*, rumus menghitung *F1-score* sebagai berikut.

$$F1 - Score = \frac{2(Precision \times Recall)}{(Precision + Recall)} \quad (9)$$

3. HASIL DAN PEMBAHASAN

Dataset yang digunakan pada penelitian ini sebanyak 150 data bunga Iris dengan empat atribut yang menjadi label klasifikasi yaitu *Iris Setosa*, *Iris Versicolor*, dan *Iris Virginica* dengan *percentage split* dan *k-fold cross validation* sebagai metode pelatihan yang digunakan.

3.1. Percentage Split

Percentage Split menggunakan sebesar 80% data *training* dan 20% data *testing* dengan *confusion matrix* yang dihasilkan seperti terlihat pada Tabel 3.

Tabel 3. *Confusion Matrix* menggunakan *Percentage Split*

		Prediksi		
		<i>Iris Setosa</i>	<i>Iris Versicolor</i>	<i>Iris Virginica</i>
Aktual	<i>Iris Setosa</i>	14	0	0
	<i>Iris Versicolor</i>	1	6	0
	<i>Iris Virginica</i>	0	0	9

Berdasarkan Tabel 3, SVM memprediksi dengan benar sebanyak 14 data *Iris Setosa*, 6 data *Iris Versicolor*, 9 data *Iris Virginica*, dan 1 data *Iris Versicolor* diprediksi *Iris Setosa*. Hasil presisi yang dihasilkan untuk *Iris Setosa* sebesar 93%, *Iris Versicolor* sebesar 100%, dan *Iris Virginica* sebesar 100%. Hasil *recall* yang dihasilkan untuk *Iris Setosa* sebesar 100%, *Iris Versicolor* sebesar 86%, dan *Iris Virginica* sebesar 100%. Hasil *F1-score* yang dihasilkan untuk *Iris Setosa* sebesar 97%, *Iris Versicolor* sebesar 92%, dan *Iris Virginica* sebesar 100%. Nilai akurasi yang diperoleh dengan menggunakan *percentage split* sebesar 96,7%, berarti *Support Vector Machine* sudah sangat baik dalam memprediksi hasil klasifikasi dari bunga iris.

3.2. K-fold Cross Validation

K-fold cross validation menggunakan nilai k sebesar 10 dengan *confusion matrix* yang dihasilkan seperti terlihat pada Tabel 4.

Tabel 4. *Confusion Matrix* menggunakan *K-fold Cross Validation*

		Prediksi		
		<i>Iris Setosa</i>	<i>Iris Versicolor</i>	<i>Iris Virginica</i>
Aktual	<i>Iris Setosa</i>	50	0	0
	<i>Iris Versicolor</i>	6	42	2
	<i>Iris Virginica</i>	0	3	47

Berdasarkan Tabel 4, SVM memprediksi dengan benar sebanyak 50 data *Iris Setosa*, 42 data *Iris Versicolor*, 47 data *Iris Virginica*, 6 data *Iris Versicolor* diprediksi *Iris Setosa*, 2 data *Iris Versicolor* diprediksi *Iris Virginica*, dan 3 data *Iris Virginica* diprediksi *Iris Versicolor*. Hasil presisi yang dihasilkan untuk *Iris Setosa* sebesar 89%, *Iris Versicolor* sebesar 93%, dan *Iris Virginica* sebesar 96%. Hasil *recall* yang dihasilkan untuk *Iris Setosa* sebesar 100%, *Iris Versicolor* sebesar 84%, dan *Iris Virginica* sebesar 94%. Hasil *F1-score* yang dihasilkan untuk *Iris Setosa* sebesar 94%, *Iris Versicolor* sebesar 88%, dan *Iris Virginica* sebesar 95%. Nilai akurasi yang diperoleh dengan menggunakan *k-fold cross validation* sebesar 92,6%, berarti SVM sudah sangat baik dalam memprediksi hasil klasifikasi dari bunga iris.

3.3. Perbandingan Hasil

Berdasarkan hasil prediksi menggunakan metode *percentage split* dan *k-fold cross validation*, diperoleh bahwa kedua metode tersebut memberikan hasil akurasi yang sangat baik. Perbandingan kedua metode tersebut dapat dilihat pada Tabel 5 dan Tabel 6.

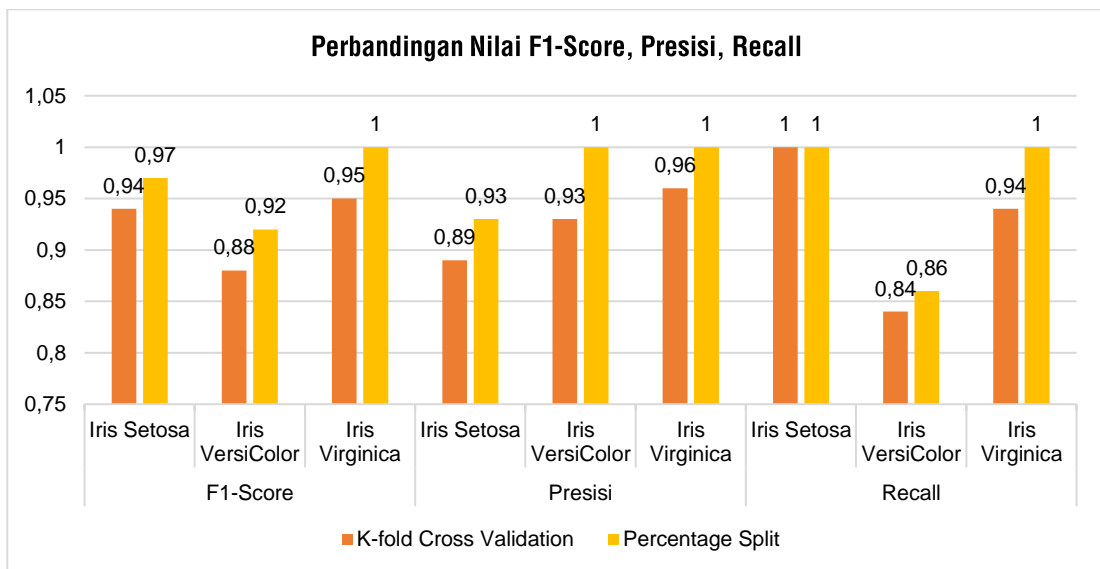
Tabel 5. Perbandingan Nilai Presisi dan Akurasi

	<i>Presisi</i>			<i>Akurasi</i>
	<i>Iris Setosa</i>	<i>Iris Versicolor</i>	<i>Iris Virginica</i>	
<i>Percentage Split</i>	93%	100%	100%	96,7%
<i>K-Fold Cross Validation</i>	89%	93%	96%	92,6%

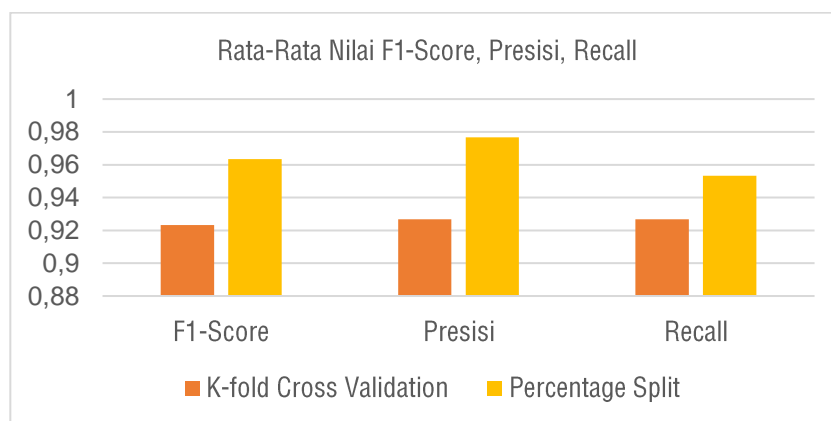
Tabel 6. Perbandingan Nilai Recall dan F1-score

	<i>Recall</i>			<i>F1-score</i>		
	<i>Iris Setosa</i>	<i>Iris Versicolor</i>	<i>Iris Virginica</i>	<i>Iris Setosa</i>	<i>Iris Versicolor</i>	<i>Iris Virginica</i>
<i>Percentage Split</i>	100%	86%	100%	97%	92%	100%
<i>K-Fold Cross Validation</i>	100%	84%	94%	94%	88%	95%

Berdasarkan Tabel 5 dan Tabel 6, nilai akurasi, presisi, recall, dan F1-score jauh lebih tinggi pada *percentage split* dibanding *k-fold cross validation*. Nilai evaluasi minimum yang diperoleh menggunakan *percentage split* adalah recall pada kelas *versicolor* dengan nilai sebesar 86% dengan kategori baik. Nilai evaluasi minimum tersebut tidak jauh berbeda dengan nilai minimum yang diperoleh menggunakan *k-fold cross validation*. Nilai presisi, recall, dan F1-score disajikan dalam bentuk diagram pada Gambar 1.



Gambar 1. Perbandingan Nilai F1-score, Presisi dan Recall Hasil Prediksi



Gambar 2. Rata-rata Nilai F1-score, Presisi dan Recall Hasil Prediksi

Pada Gambar 2 terlihat bahwa metode *percentage split* memiliki nilai presisi, *recall*, dan *F1-score* yang lebih unggul dibandingkan metode *k-fold cross validation* dengan nilai lebih dari 95%, namun kedua metode tersebut merupakan metode terbaik dalam melakukan klasifikasi menggunakan SVM pada bunga Iris karena memiliki nilai akurasi, presisi, *recall*, dan *F1-score* lebih dari 90%.

Tabel 7. Perbandingan dengan Penelitian Lain

Metode	Evaluasi			
	Akurasi	Presisi	Recall	F1-score
KNN [4]	92%	-	-	-
DNN [12]	96%	-	-	-
Decision Tree [17]	95,33%	-	-	-
MLP-APSO [18]	96%	-	-	-
Decision Tree [21]	86,67%	90,48%	86,67%	-
<i>Purpose Method</i>				
Percentage Split	96,7%	97,6%	95,3%	96,3%
K-Fold Cross Validation	92,6%	92,6%	92,6%	92,3%

Berdasarkan Tabel 7, dapat terlihat bahwa klasifikasi bunga iris memberikan nilai evaluasi di atas 85% dengan kategori baik. Klasifikasi bunga iris menggunakan SVM dengan metode *percentage split* memiliki nilai akurasi serupa menggunakan metode DNN dan MLP-APSO dengan nilai akurasi sebesar 96%. Metode SVM menggunakan *k-fold cross validation* juga memiliki nilai akurasi serupa dengan metode KNN dengan nilai akurasi 92% meskipun nilai tersebut lebih rendah dibandingkan dengan metode SVM menggunakan *percentage split*. Pada nilai presisi dan recall, metode SVM memiliki nilai evaluasi yang lebih tinggi dengan kategori sangat baik dibandingkan dengan metode *Decision Tree*. Berdasarkan hasil tersebut, maka metode SVM menggunakan *percentage split* merupakan metode terbaik dalam melakukan klasifikasi pada bunga iris.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, klasifikasi bunga Iris menggunakan metode SVM dengan kernel *polynomial* dengan metode data latih *percentage split* memiliki nilai akurasi, presisi, *recall*, dan *F1-score* di atas 95% dengan kategori sangat baik dan metode data latih *k-fold cross validation* memiliki nilai akurasi, presisi, recall, dan *F1-score* di atas 90% dengan kategori sangat baik. Nilai evaluasi yang diperoleh kedua metode tersebut menunjukkan bahwa metode SVM mampu melakukan klasifikasi bunga iris dan dapat menggunakan metode latih *percentage split* untuk menghasilkan nilai evaluasi terbaik.

DAFTAR PUSTAKA

- [1] J. Eska, "Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5". *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, vol. 2(2), pp. 9–13, 2016.
- [2] D.H. Kamagi & S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa". *Jurnal ULTIMATICS*, vol. 6(1), pp. 15–20, 2014.
- [3] S. Mutfrofin, A. Izzah, A. Kurniawardhani & M. Masrur. "Optimasi Teknik Klasifikasi Modified K Nearest Neighbor Menggunakan Algoritma Genetika". *Jurnal Gamma*, vol. 10(1), pp.130–134, 2014.
- [4] P. Putra, Akim M H Pardede & S. Syahputra. "Analisis Metode K-Nearest Neighbour (KNN) dalam Klasifikasi". *Jurnal Teknik Informatika Kaputama (JTIK)*, vol.6(1), pp. 297–305, 2022.
- [5] F.S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian". *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 1(1), pp. 19–25, 2017.
- [6] W.S. Noble, (2006). "What is a support vector machine?". *Nature Biotechnology*, vol. 24(12), pp. 1565–1567, 2006.
- [7] I.A. Muis & M. Affandes, "Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet". *Sains, Teknologi Dan Industri*, vol. 12(2), pp. 189–197, 2015.
- [8] A. S. Ritonga & E.S. Purwaningsih, "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan SMAW (Shield Metal Arc Welding)". *Ilmiah Edutic*, vol. 5(1), pp. 17–25, 2018.

- [9] P. A. Octaviani, Y. Wilandari & D. Ispriyanti, (2014). "Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang". *Jurnal Gaussian*, vol. 3(8), 811–820, 2014.
- [10] A. Jacobus & E. Winarko, "Penerapan Metode Support Vector Machine pada Sistem Deteksi Intrusi secara Real-time". *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 8(1), p. 13, 2014.
- [11] A. Mustika & M. Affandes, "Penerapan Metode Support Vector Machine Dalam Klasifikasi Sentimen Tweet Public Figure". *Sentra*, pp. 978–979, 2015.
- [12] A. Eldem, H. Eldem & D. Üstün, "A Model of Deep Neural Network for Iris Classification With Different Activation Functions". *2018 International Conference on Artificial Intelligence and Data Processing*, 2018.
- [13] I. Novelisari, R. Fitriana & H. Susanti, "Analisis Makna Motif Bunga pada Kimono". *Jurnal Studi Jepang*, vol. 2(1), pp. 1–13, 2020.
- [14] H. Pramaditya & T.I. Sugiharto, "Aplikasi Pembelajaran Augmented Reality Bunga Iris Berbasis Unity". *Seminar Nasional Sistem Informasi*. pp. 2517–2525, 2020.
- [15] E. Retnoningsih & R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python". *Bina Insani ICT Journal*, vol. 7(2), p. 156, 2020.
- [16] M. Lachowicz (2019). "Decision Tree Approach for IRIS Database Classification". *CEUR Workshop Proceedings*, pp. 10–14, 2019.
- [17] P.K. Handayani, (2020). "Penerapan Principal Component Analysis untuk Peningkatan Kinerja Algoritma Decision Tree pada Iris Dataset". *Indonesian Journal of Technology, Informatics and Science (IJTIS)*, vol. 1(2), pp. 55–58, 2020.
- [18] M. Mentari, E.K.R. Sari & S. Mutrofin (2014). "Klasifikasi Menggunakan Kombinasi Multilayer Perceptron dan Aligment Particle Swarm Optimization". *Senastik*, vol. 2014(September), pp. 10–11, 2014.
- [19] I.M. Parapat, M.T. Furqon & Sutrisno, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak". *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 2(10), pp. 3163–3169, 2018.
- [20] D. Normawati & S.A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter". *Jurnal Sains Komputer & Informatika (J-Sakti)*, vol. 5(2), pp. 697–711, 2021.
- [21] Y. Verawati & M.S. Hasibun, (2021). "Perbandingan Data Set Iris dengan Aplikasi Rapid Miner dan Orange Menggunakan Algoritma Klasifikasi". *Jurnal Darmajaya*, vol. 11(1), pp. 158-163, 2021.