

"Klasifikasi jenis Ikan Menggunakan Teknik Pembelajaran Mesin Dengan Algoritma SVM"

Vixy Ardian Listianto (A11.2018.11515)

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian
Nuswantoro

Abstract

Problem identification: In this step we find problems according to the information technology department, in this case the classification of fish types in the sea. The goal of this step is to collect relevant sources so that this research can contribute to new research. Support Vector Machine (SVM) is one of the best machine learning-based classification methods that can be used to solve classification and regression problems using master data as input. The SVM method is widely used because it is simpler and can provide good and accurate results in problem classification C: 0.1 The kernel test coefficient for the CI variable (0.386244, 0.415423) shows a significant difference. Models with linear kernels have better results. The variable kernel test coefficient C:1 CI (-0.013760, -0.010127) shows that the RBF kernel model is better than the linear kernel model. This confidence interval shows that the mean difference is quite significant and the RBF model provides better results. Kernel test coefficient C:10 for CI variables (-0.056773, -0.054336) The results show that the RBF kernel model is also better than the linear kernel model. The confidence interval shows significant mean differences and the RBF model provides better results. Hypothesis Null hypothesis (H0): Linear variables have a significant effect on RBF simultaneously. Hypothesis The hypothesis used in the partial test is as follows: Variable x_1 is s_1 Null hypothesis (H0): Linear variables have no significant effect on RBF. Alternative (H1): Linear variables have a significant effect on RBF

Abstrak

Identifikasi masalah: Pada langkah ini kita menemukan permasalahan menurut departemen teknologi informasi, dalam hal ini klasifikasi jenis ikan di laut. Tujuan dari langkah ini adalah mengumpulkan sumber-sumber yang relevan agar penelitian ini dapat memberikan kontribusi pada penelitian baru. Support Vector Machine (SVM) merupakan salah satu metode klasifikasi berbasis machine learning terbaik yang dapat digunakan untuk menyelesaikan masalah klasifikasi dan regresi dengan menggunakan data master sebagai masukannya. Metode SVM banyak digunakan karena lebih sederhana dan dapat memberikan hasil yang baik dan akurat dalam klasifikasi masalah C: 0.1 Koefisien uji kernel untuk variabel CI (0.386244, 0.415423) menunjukkan perbedaan yang signifikan. Model dengan kernel linier mempunyai hasil yang lebih baik. Koefisien uji kernel variabel C:1 CI (-0.013760, -0.010127) menunjukkan bahwa model kernel RBF lebih baik dibandingkan model kernel linier. Interval kepercayaan ini menunjukkan bahwa perbedaan mean cukup signifikan dan model RBF memberikan hasil yang lebih baik. Koefisien uji kernel C:10 untuk variabel CI (-0.056773, -0.054336) Hasil menunjukkan bahwa model kernel RBF juga lebih baik dibandingkan model kernel linier. Interval kepercayaan menunjukkan perbedaan mean yang signifikan dan model RBF memberikan hasil yang lebih baik. Hipotesis Hipotesis null (H0): Variabel linier berpengaruh signifikan terhadap RBF secara simultan. Hipotesis Hipotesis yang digunakan dalam uji parsial adalah sebagai berikut: Variabel x_1 adalah s_1 Hipotesis nol (H0): Variabel linier tidak berpengaruh signifikan terhadap RBF. Alternatif (H1): Variabel linier berpengaruh signifikan terhadap RBF

Kata kunci: jenis ikan, SVM, ikan, jenis, HOG, SVM kernel, kernel linear, kernel RBF

1. PENDAHULUAN

Pengelompokan dan estimasi ikan adalah hal penting dalam bidang perikanan dan akuakultur. Pengelompokan ikan dapat memberikan informasi yang berharga bagi para petani ikan, peneliti, dan pihak terkait lainnya untuk mengelola stok ikan dengan lebih efektif.

Salah satu pendekatan yang umum digunakan untuk melakukan estimasi berat ikan adalah menggunakan algoritma SVM. Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang biasa digunakan untuk tugas klasifikasi dan regresi.

SVM sangat efektif dalam menangani data berdimensi tinggi dan cocok untuk situasi dimana jumlah fitur lebih banyak daripada jumlah sampel.

Algoritme ini termasuk dalam kelompok teknik pembelajaran mesin yang dikenal sebagai ``pembelajaran yang diawasi.

""(1)

Pendekatan SVM untuk klasifikasi berat ikan didasarkan pada pengumpulan data ikan yang lengkap. Data ini mencakup Gambar pada ikan yang sebenarnya. Dengan menggunakan algoritma SVM, kita dapat mengembangkan model matematis yang memprediksi klasifikasi ikan berdasarkan data gambar tersebut.

Dengan menggunakan algoritma SVM, Diharapkan dapat menghasilkan model yang dapat digunakan untuk klasifikasi ikan berdasarkan gambar. Namun, penting untuk diingat bahwa model ini didasarkan pada asumsi-asumsi tertentu, seperti asumsi bahwa hubungan antara variabel independen dan dependen adalah dengan metode HOG. Juga, pastikan untuk memperhatikan interpretasi hasil model dan menjelaskan batasan-batasan yang mungkin ada.

Dalam praktiknya, ada banyak variasi dan peningkatan yang dapat dilakukan dalam analisis klasifikasi dari gambar ikan menggunakan algoritma SVM. Misalnya, UJicoba dapat menggabungkan variabel tambahan

Dengan demikian, analisis klasifikasi dari gambar ikan menggunakan algoritma SVM dapat menjadi alat yang berguna dalam memprediksi berat ikan berdasarkan pengukuran yang diberikan.(2)

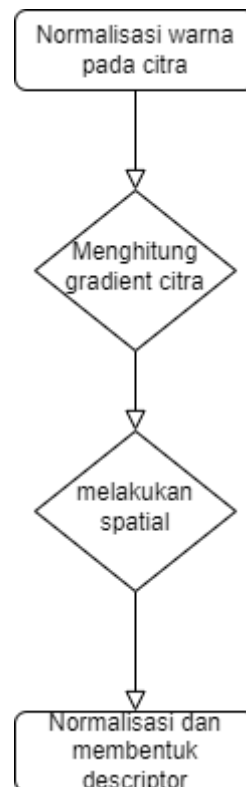
2. METODOLOGI PENELITIAN

Metode penelitian untuk klasifikasi ikan dari data yang di berikan menggunakan algoritma SVM melibatkan pengumpulan data gambar ikan serta jenis ikan yang terkait. Berikut adalah langkah-langkah umum yang dapat diikuti dalam metode ini:

1. Identifikasi Masalah: Pada tahap ini yaitu menemukan permasalahan sesuai bidang Teknik informatika,dalam hal ini adalah klasifikasi jeis ikan laut. Tahap ini ditujukan untuk

mengumpulkan referensi yang relevan sehingga penelitian ini dapat berkontribusi untuk penelitian baru.

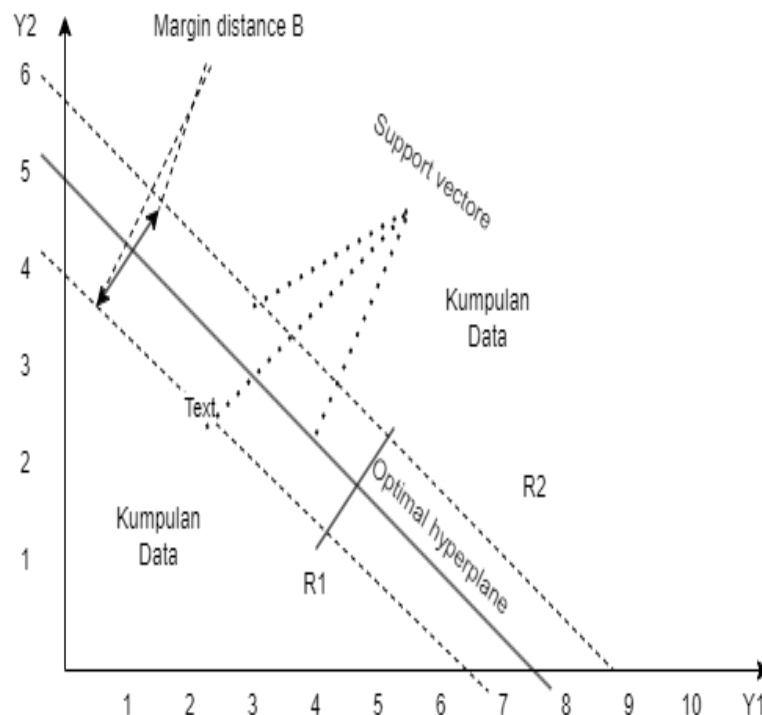
2. Histogram of Oriented Gradients: Metode Histogram of Oriented Gradients (HOG) digunakan untuk mengekstraksi bentuk citra dalam bentuk histogram. HOG secara sempurna merepresentasikan fitur bentuk karena dapat menangkap gradien lokal pada gambar (3). HOG pertama-tama menghitung nilai gradien. Normalisasi warna pada citra dilakukan, setelah itu blok-blok citra digabungkan dan diurutkan setiap bagiannya hingga menjadi representasi fitur (4). Berkat nilai mask, histogram dijumlahkan dan normalisasi blok dilakukan untuk menghindari tumpang tindih. Tahapan implementasi HOG ditunjukkan pada Gambar 1.



Gambar 1 Tahapann Implementasi HOG(5)

3. Support Vector Machine: Support Vector Machine (SVM) merupakan salah satu metode klasifikasi berbasis machine learning terbaik yang dapat digunakan untuk menyelesaikan masalah klasifikasi dan regresi dengan menggunakan data master sebagai masukannya. Metode SVM banyak digunakan karena lebih sederhana dan dapat memberikan hasil yang baik dan akurat ketika terjadi masalah klasifikasi (6). Tujuan utama dari metode klasifikasi SVM adalah untuk mendapatkan batasan berupa

hyperband terbaik yang dapat memisahkan kelas data berdasarkan jumlah fitur. Gambar 2 menunjukkan evolusi hyperplane dari distribusi terbaik vektor pendukung data SVM. Hyperplane SVM terbaik diperoleh dengan memaksimalkan margin atau jarak dari garis pemisah hyperplane ke distribusi data latih terdekat. Selain itu, SVM memiliki fungsi dasar seperti linier dan polinomial yang dapat mempelajari data untuk menemukan hyperplane terbaik sebagai model klasifikasi.



Gambar 2 Pembentukan hyperlane pada SVM(7)

4. Pengumpulan Data: Data gambar ikan laut diambil dari segmentasi dan klasifikasi Fish Large Dataset publik (8). Mengumpulkan 9000 gambar ikan laut. Gambar tersebut memperlihatkan sembilan jenis ikan laut, yaitu black sea spart, gilt head bream, hourse mackerel, red mullet, red sea bream, sea bass, shrimp, striped red mullet dan ikan trout. Terdapat 9000 gambar berukuran 590 x 445 piksel untuk setiap jenis ikan. Jenis dan jumlah ikan laut yang digunakan ditunjukkan pada Tabel 1.

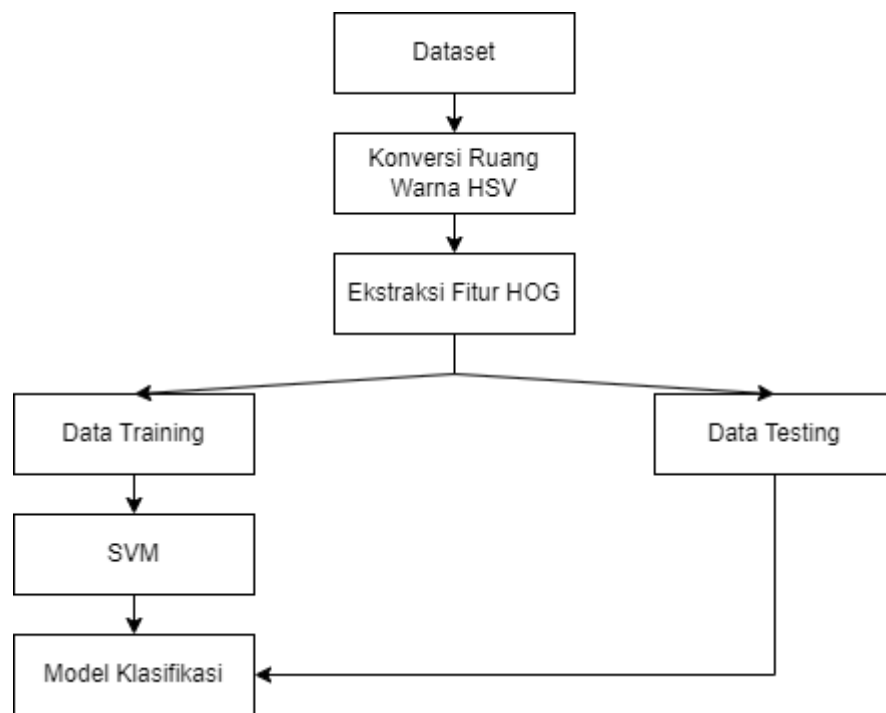
no	Jenis Ikan Laut	Citra Ikan Laut	Jumlah Citra
----	-----------------	-----------------	--------------

1	Black Sea Sprat		1000
2	Gilt Head Bream		1000
3	Hourse Mackerel		1000
4	Red Mullet		1000
5	Red Sea Bream		1000
6	Sea Bass		1000
7	Shrimp		1000
8	Striped Red Mullet		1000
9	Trout		1000

Tabel 1 Jenis dan Jumlah Dataset Ikan Laut Yang Digunakan

- Perancangan system: Pada langkah ini, HOG melakukan konversi warna citra RGB ke format HSV dan ekstraksi fitur, dan proses klasifikasi dilanjutkan dengan SVM. Proses penelitian ini mempunyai dua tahap yaitu pelatihan dan pengujian. Saat mengerjakan ini dalam pelatihan. Kemudian citra setiap jenis ikan yang terdapat di laut diubah menjadi citra HSV, kemudian fitur HOG diekstraksi dan diperoleh hasilnya. Ekstraksi fitur HOG digunakan sebagai fitur pada proses pelatihan menggunakan SVM. Di akhir pelatihan diperoleh model klasifikasi yang digunakan

dalam pengujian. Pada proses pengujian, citra ikan laut diubah menjadi citra HSV dan kemudian fungsi ini dihilangkan. HOG dan fitur turunannya digunakan untuk menguji model klasifikasi. Diagram struktur sistem klasifikasi ikan laut ditunjukkan pada Gambar 3.



Gambar 3 Perancangan Sistem Klasifikasi Ikan

6. Implementasi: Setelah proses perancangan sistem, dilakukan tahap implementasi atau tahap kode program. Konversi data menjadi sampel HSV kemudian dilanjutkan pada langkah ekstraksi fitur HOG.

3. HASIL DAN PEMBAHASAN

3.1. Model Support Vector Machine

Berdasarkan hasil Analisa data citra menggunakann SVM , diperoleh hasil untuk dua jenis kernel, Linear dan RBF.

3.1.1 Model Support Vector Machine dengan Kernel Linear

Dengan menggunakan kernel linear, model SVM menghasilkan hasil sebagai berikut: Model SVM: Kernel Linear

Parameter Model :

- C (Regulasi Parameter):0.1
- Kernel:Linear

Hasil Analisis :

- Mean Test Score: 0.8756
- Standard Deviation Test Score: 0.0003
- Mean Training Score: 1.0
- Standard Deviation Training Score: 0.0

Model SVM dengan kernel linear memiliki akurasi rata – rata sekitar 87.56% pada data uji.Nilai Deviasi standar menunjukan konsistensi performa model.Pada data latih,model mencapai akurasi sempurna (1.0) tanpa deviasi.

3.1.2 Model Support Vector Machine dengan kernel RBF

Dengan menggunakan kernel RBF,model SVM memberikan hasil sebagai berikut: Model SVM: Kernel RBF

Parameter Model :

- C (Regulasi Parameter):0.1
- Kernel: RBF

Hasil Analisis :

- Mean Test Score: 0.4747
- Standard Deviation Test Score: 0.0069
- Mean Training Score: 0.5261
- Standard Deviation Training Score: 0.0067

Model SVM dengan kernel RBF memiliki akurasi rata-rata sekitar 47.47% pada data uji.Nilai deviasi standar yang cukup tinggi menunjukkan variasi performa model pada data uji.Pada data latih, model memiliki akurasi sekitar 52.61% dengan deviasi standar 0.0067.

3.2 Hasil Analisis

3.2.1 Hipotesis

Hipotesis Nol (H_0)

Tidak ada perbedaan besar dalam kemampuan versi SVM untuk mengkategorikan jenis ikan dari gambar menggunakan kernel linier atau RBF.Hipotesis Alternatif (H_1):

Penggunaan kernel linier dan kernel RBF memungkinkan model SVM mengklasifikasikan spesies ikan berdasarkan gambar.

3.2.2 Daerah Kritis

Pada tingkat signifikansi α yang telah ditentukan ($\alpha = 0.05$). Jika perbedaan antara dua model melebihi 1.96 kali nilai std_test_score,Maka uji hipotesis menolak hipotesis nol.

3.2.3 Analisis Uji Kernel C:0.1

Tabel Gambar 1 berikut merupakan hasil uji kernel linear dan RBF dengan C = 0.1

Tabel Gambar 1. Hasil Uji Kernel C:0.1

	params	mean_test_score	std_test_score	mean_train_score	std_train_score
0	{'C': 0.1, 'kernel': 'linear'}	0.875556	0.000278	1.000000	0.000000
1	{'C': 0.1, 'kernel': 'rbf'}	0.474722	0.006944	0.526111	0.006667

- Model dengan Kernel Linear
($\bar{x}_1 = 0.875556, s_1 = 0.000278$)
- Model dengan Kernel RBF
($\bar{x}_2 = 0.474722, s_2 = 0.006944$)

$$\text{Rumus } (CI = (\bar{x}_1 - \bar{x}_2) \pm Z \times \sqrt{s_1^2 + s_2^2})$$

$$CI = (0.875556 - 0.474722) \pm 1.96 \times \sqrt{(0.000278)^2 + (0.006944)^2}$$

$$CI = 0.400834 \pm 1.96 \times 0.006946$$

$$CI = (0.386244, 0.415423)$$

Oleh karena itu, perbedaan reliabilitas rata-rata antara kedua model SVM adalah sekitar (0,386, 0,415). Jika nol tidak berada dalam interval kepercayaan ini, hasil uji menolak hipotesis nol.

3.2.4 Analisis Uji Kernel C:1

Tabel Gambar 2 berikut merupakan hasil uji kernel linear dan RBF dengan C = 1

Tabel Gambar 2. Hasil Uji Kernel C:1

2	{'C': 1, 'kernel': 'linear'}	0.875556	0.000278	1.000000	0.000000
3	{'C': 1, 'kernel': 'rbf'}	0.887500	0.000833	0.986389	0.001111

- Model dengan Kernel Linear
($\bar{x}_1 = 0.875556, s_1 = 0.000278$)
- Model dengan Kernel RBF
($\bar{x}_2 = 0.887500, s_2 = 0.000833$)

$$\text{Rumus } (CI = (\bar{x}_1 - \bar{x}_2) \pm Z \times \sqrt{s_1^2 + s_2^2})$$

- $CI = (0.875556 - 0.887500) \pm 1.96 \times \sqrt{(0.000278)^2 + (0.000833)^2}$
- $CI = -0.011944 \pm 1.96 \times 0.000874$
- $CI = (-0.013760, -0.010127)$

Oleh karena itu, interval kepercayaan dari perbedaan rata-rata antara kedua model SVM adalah kira-kira (-0,014, -0,010). Karena nilai nol tidak termasuk dalam interval kepercayaan ini, hasil uji menolak hipotesis nol.

3.2.5 Analisis Uji Kernel C:10

Tabel Gambar 3 berikut merupakan hasil uji kernel linear dan RBF dengan C = 10

Tabel Gambar 3. Hasil Uji Kernel C:10

4	{'C': 10, 'kernel': 'linear'}	0.875556	0.000278	1.000000	0.000000
5	{'C': 10, 'kernel': 'rbf'}	0.931111	0.000556	1.000000	0.000000

- Model dengan Kernel Linear
($\bar{x}_1 = 0.875556, s_1 = 0.000278$)
- Model dengan Kernel RBF
($\bar{x}_2 = 0.931111, s_2 = 0.000556$)

$$\text{Rumus } (CI = (\bar{x}_1 - \bar{x}_2) \pm Z \times \sqrt{s_1^2 + s_2^2})$$

- $CI = (0.875556 - 0.931111) \pm 1.96 \times \sqrt{(0.000278)^2 + (0.000556)^2}$
- $CI = -0.055555 \pm 1.96 \times 0.000624$
- $CI = (-0.056773, -0.054336)$

Oleh karena itu, interval kepercayaan dari perbedaan rata-rata antara kedua model SVM adalah kira-kira (-0,057, -0,054). Karena nilai nol tidak termasuk dalam interval kepercayaan ini, hasil uji menolak hipotesis nol.

3.2.6 Hasil Akhir

Tabel Gambar 4 berikut adalah hasil akhir dari uji akhir SVM

Tabel Gambar 4. Hasil Uji Akhir SVM

	params	mean_test_score	std_test_score	mean_train_score	std_train_score
0	{'C': 0.1, 'kernel': 'linear'}	0.875556	0.000278	1.000000	0.000000
1	{'C': 0.1, 'kernel': 'rbf'}	0.474722	0.006944	0.526111	0.006667
2	{'C': 1, 'kernel': 'linear'}	0.875556	0.000278	1.000000	0.000000
3	{'C': 1, 'kernel': 'rbf'}	0.887500	0.000833	0.986389	0.001111
4	{'C': 10, 'kernel': 'linear'}	0.875556	0.000278	1.000000	0.000000
5	{'C': 10, 'kernel': 'rbf'}	0.931111	0.000556	1.000000	0.000000

3.3. Koefisien Determinasi (R-squared) Pada Model SVM

Nilai R-squared yang ditunjukkan pada Tabel 1 yaitu sebesar 0.885 mengindikasikan bahwa sekitar 93,5% dari variasi dalam persentase jenis ikan yang dapat dijelaskan oleh model SVM. Artinya, sebagian besar variasi dalam persentase berat ikan dapat diatribusikan kepada variabel-variabel yang digunakan dalam model regresi tersebut. Namun, sekitar 10,% dari variasi masih tidak dapat dijelaskan oleh model ini dan mungkin dipengaruhi oleh faktor-faktor lain yang tidak dimasukkan ke dalam model.

4. SIMPULAN

Metode penelitian untuk klasifikasi jenis ikan dari pengambilan gambar menggunakan algoritma SVM dengan fitur ekstraksi HOG melibatkan pengumpulan data pengukuran ikan berupa gambar beberapa jenis ikan yang terkait.

Estimasi jenis: Setelah model terbentuk, gunakan variabel independen berupa(foto bagian kanan, foto bagian kiri, foto bagian atas) dari ikan yang sudah di klompokkan untuk di jadikan data training.

C: 0.1 Koefisien uji kernel untuk variabel CI (0.386244,0.415423) menunjukkan perbedaan yang signifikan. Model dengan kernel linear memiliki hasil yang lebih baik.

C: 1 Koefisien uji karnel untuk variable CI (-0.013760,-0.010127) menunjukkan bahwa model kernel RBF lebih baik dibandingkan model kernel linier. Interval kepercayaan ini menunjukkan bahwa perbedaan mean cukup signifikan dan model RBF memberikan hasil yang lebih baik.

C: 10 koefisien uji kernel untuk variable CI (-0.056773,-0.054336) hasilnya menunjukkan bahwa model kernel RBF juga lebih baik dibandingkan model kernel linier. Interval kepercayaan menunjukkan perbedaan mean yang signifikan dan model RBF memberikan hasil yang lebih baik.

Hipotesis Hipotesis Nol (H_0): Adanya pengaruh simultan yang signifikan dari variabel Linear terhadap RBF.

Hipotesis Hipotesis yang digunakan dalam uji parsial adalah sebagai berikut: Variabel x_1 terhadap s_1 Hipotesis Nol (H_0) : Tidak ada pengaruh yang signifikan variabel Linear terhadap RBF. Alternatif (H_1) : Terdapat pengaruh yang signifikan variabel Linear terhadap RBF..

Nilai R-squared yang ditunjukkan pada Tabel 1 yaitu sebesar 0.885 mengindikasikan bahwa sekitar 93,5% dari variasi dalam persentase jenis ikan.

DAFTAR PUSTAKA

1. Kristiani SA, Rainarli E. DOCUMENT CLASSIFICATION WITH SUPPORT VECTOR MACHINE AND MUTUAL INFORMATION.
2. Handayanto A, Latifa K, Saputro ND, Waliansyah RR. Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi. Juita. 2019 Nov 26;7(2):71.
3. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) [Internet]. San Diego, CA, USA: IEEE; 2005 [cited 2023 Dec 26]. p. 886–93. Available from: <http://ieeexplore.ieee.org/document/1467360/>
4. Yohannes Y, Pribadi MR, Chandra L. Klasifikasi Jenis Buah dan Sayuran Menggunakan SVM Dengan Fitur Saliency-HOG dan Color Moments. ELKHA. 2020 Oct 11;12(2):125.
5. Yohannes Y, Al Rivani ME. Penggunaan Global Contrast Saliency dan Histogram of Oriented Gradient Sebagai Fitur untuk Klasifikasi Jenis Hewan Mamalia. petir. 2020 Mar 21;13(1):80–5.
6. Klyueva I. Improving Quality of the Multiclass SVM Classification Based on the Feature Engineering. In: 2019 1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency

- (SUMMA) [Internet]. Lipetsk, Russia: IEEE; 2019 [cited 2023 Dec 26]. p. 491–4. Available from: <https://ieeexplore.ieee.org/document/8947599/>
7. Paliouras G, Karkaletsis V, Spyropoulos CD, editors. Machine learning and its applications: advanced lectures. Berlin ; New York: Springer; 2001. 324 p. (Lecture notes in computer science ; Lecture notes in artificial intelligence).
 8. Ulucan O, Karakaya D, Turkan M. A Large-Scale Dataset for Fish Segmentation and Classification. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU) [Internet]. Istanbul, Turkey: IEEE; 2020 [cited 2023 Dec 26]. p. 1–5. Available from: <https://ieeexplore.ieee.org/document/9259867/>
 9. Leonardo L, Yohannes Y, Hartati E. Klasifikasi Sampah Daur Ulang Menggunakan Support Vector Machine Dengan Fitur Local Binary Pattern. algoritme. 2020 Oct 10;1(1):78–90.
 10. Astrianda N. Klasifikasi Kematangan Buah Tomat Dengan Variasi Model Warna Menggunakan Support Vector Machine. VCT. 2020 Apr 13;1(2):45–52.
 11. Prasthio R, Yohannes Y, Devella S. Penggunaan Fitur HOG Dan HSV Untuk Klasifikasi Citra Sel Darah Putih. algoritme. 2022 Apr 11;2(2):120–32.