

PENGGUNAAN PRINCIPAL COMPONENT ANALYSIS (PCA) DAN K-NEAREST NEIGHBORS (KNN) DALAM KLASIFIKASI SAYURAN MENTIMUN, PARE, DAN TERONG

Ahmad Syarif¹, Agung Ramadhanu²

Universitas Putra Indonesia YPTK, Padang

e-mail: ¹ahmadsyarif.skom@gmail.com, ²agung_ramadhanu@upiptk.ac.id

Abstract: *The classification of vegetable types is an important aspect of the agricultural industry to improve the efficiency of product management, packaging, and distribution. This study aims to implement the Principal Component Analysis (PCA) and K-Nearest Neighbors (KNN) methods in the vegetable classification process, especially for cucumbers, squash, and eggplants. PCA is used to reduce the data dimensions and extract the key significant features that distinguish vegetable categories. Meanwhile, KNN is applied as a classification algorithm based on the proximity of these key features. The research dataset consists of digital images of vegetables extracted into color, texture, and shape attributes. The results show that the combination of PCA and KNN is able to significantly improve the classification accuracy by minimizing computational complexity. Experiments are carried out with various numbers of main components in PCA and variations in k parameter values in KNN to determine the optimal configuration. In the best configuration, this method achieves a classification accuracy of 90%, with PCA effectively reducing the data dimension by 95% without losing important information. In conclusion, this approach has great potential to be implemented in vegetable classification automation systems to support efficiency in the agricultural supply chain.*

Keywords: *Vegetables classification, Principal Component Analysis (PCA), K-Nearest Neighbors (KNN), cucumber, bitter melon, eggplant.*

Abstrak: Klasifikasi jenis sayuran merupakan salah satu aspek penting dalam industri pertanian untuk meningkatkan efisiensi pengelolaan produk, pengemasan, serta distribusi. Penelitian ini bertujuan untuk mengimplementasikan metode Principal Component Analysis (PCA) dan K-Nearest Neighbors (KNN) dalam proses klasifikasi sayuran, khususnya mentimun, pare, dan terong. PCA digunakan untuk mengurangi dimensi data dan mengekstraksi fitur utama yang signifikan dalam membedakan kategori sayuran. Sementara itu, KNN diterapkan sebagai algoritma klasifikasi berdasarkan kedekatan fitur-fitur utama tersebut. Dataset penelitian terdiri atas citra digital sayuran yang diekstraksi menjadi atribut warna, tekstur, dan bentuk. Hasil penelitian menunjukkan bahwa kombinasi PCA dan KNN mampu meningkatkan akurasi klasifikasi secara signifikan dengan meminimalkan kompleksitas komputasi. Eksperimen dilakukan dengan berbagai jumlah komponen utama pada PCA dan variasi nilai parameter k pada KNN untuk menentukan konfigurasi optimal. Pada konfigurasi terbaik, metode ini mencapai akurasi klasifikasi sebesar 90%, dengan PCA secara efektif mereduksi dimensi data hingga 95% tanpa kehilangan informasi penting. Kesimpulannya, pendekatan ini memiliki potensi yang besar untuk diimplementasikan dalam sistem otomatisasi klasifikasi sayuran guna mendukung efisiensi dalam rantai pasok agrikultur.

Kata kunci: Klasifikasi sayuran, Principal Component Analysis (PCA), K-Nearest Neighbors (KNN), mentimun, pare, terong.

PENDAHULUAN

Klasifikasi objek berbasis citra telah menjadi salah satu topik penelitian yang berkembang pesat dalam bidang pengolahan citra digital dan kecerdasan buatan. Dalam konteks agrikultur, klasifikasi jenis sayuran memiliki peran penting dalam mendukung proses otomatisasi, mulai dari identifikasi produk, pengelompokan berdasarkan kategori, hingga pengemasan dan distribusi. Pendekatan ini tidak hanya meningkatkan efisiensi operasional, tetapi juga dapat mengurangi potensi kesalahan manusia dalam proses klasifikasi.

Sayuran seperti mentimun, pare, dan terong memiliki karakteristik morfologi dan tekstur yang unik, sehingga memerlukan metode yang akurat dan efisien untuk mengidentifikasi serta mengklasifikasikannya. Tantangan utama dalam klasifikasi ini adalah variasi alami dalam bentuk, ukuran, warna, dan kondisi permukaan akibat faktor lingkungan atau perlakuan pascapanen. Oleh karena itu, diperlukan pendekatan yang mampu menangani variasi tersebut sekaligus menghasilkan klasifikasi yang andal.

Metode Principal Component Analysis (PCA) dan K-Nearest Neighbors (KNN) merupakan kombinasi teknik yang menjanjikan dalam menangani permasalahan tersebut. PCA berfungsi untuk mereduksi dimensi data dengan mengekstraksi fitur-fitur utama yang paling relevan, sehingga dapat meningkatkan efisiensi komputasi tanpa kehilangan informasi signifikan. Sementara itu, KNN, sebagai algoritma berbasis instance learning, memungkinkan klasifikasi yang fleksibel dengan memanfaatkan kedekatan data berdasarkan metrik tertentu, seperti jarak Euclidean.

Penelitian ini bertujuan untuk mengevaluasi kinerja kombinasi metode PCA dan KNN dalam klasifikasi tiga jenis sayuran, yaitu mentimun, pare, dan terong. Studi ini tidak hanya mengukur tingkat akurasi klasifikasi, tetapi juga menganalisis pengaruh jumlah komponen

utama PCA dan parameter kkk pada KNN terhadap hasil akhir. Dengan pendekatan ini, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam pengembangan teknologi klasifikasi berbasis citra untuk aplikasi agrikultur.

METODE

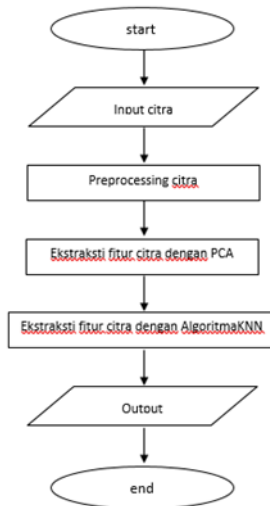
Tahapan Penelitian

1. Studi Pustaka Pada studi literatur merupakan suatu kegiatan terkait metode yang dijalankan dalam penelitian. Literatur bias didapat dari sumber seperti buku-buku, jurnal, artikel, situs ilmiah. Pada penelitian ini, literatur mengenai pengolahan citra pada jamur serta algoritma PCA dan KNN dijadikan sebagai bahan acuannya.
2. Pengumpulan Data Untuk pengumpulan datanya akan berupa data citra dari Terdiri dari 3 jenis sayuran : Mentimun, Pare dan Terong, berformat JPG berupa citra RGB. Citra di ambil sebanyak 16 citra untuk data latih ,untuk data testing atau data uji menggunakan 10 citra yang berarti total semuanya adalah 26 citra , data nantinya akan di ekstraksi menggunakan matlab R2023b.
3. Analisa Kebutuhan Sistem Analisa Pada tahapan ini mempunyai tujuan untuk mengetahui kebutuhan dalam pembuatan sistem tentang implementasi Pricipal Component Analysis dan KNearest Neighbor. Kebutuhan yang diperlukan dalam penelitian ini yaitu kebutuhan hardware (perangkat keras) dan kebutuhan software (perangkat lunak).
4. Tahapan Perancangan Sistem Tahapan sistem meliputi penginputan citra berupa format RGB, lalu proses segmentasi dan morfologi. Selanjutnya ekstraksi ciri, yaitu Tahap ekstraksi informasi yang terdapat pada objek dalam citra digital. Fitur yang digunakan yaitu

RGB, Hue, Saturation, Value dan Area. Mengekstraksi ciri warna dari R'G'B' (Red, Green, Blue) menjadi H'S'V' (Hue, Saturation, Value)

Flowchart

Tahapan pada proses klasifikasi dibuat seperti gambar dibawah ini :



Gambar flowchart

Principal Component Analysis (PCA)

Principal Components Analysis (PCA) adalah metode mengekstraksi atribut penting dari kumpulan data dengan memisahkan atau menguraikan data untuk menghasilkan koefisien yang tidak relevan. PCA sendiri disebut juga dengan transformasi Kahunen-Loeve, transformasi Hotelling, atau teknik Eigenface (Dunteman. 1989: 7)

Analisis komponen utama (PCA) adalah teknik untuk menyelaraskan bentuk set awal variabel dengan satu set kecil variabel yang tidak terkait yang dapat mewakili informasi dari set variabel. Tujuan dari analisis komponen utama (PCA) adalah untuk menjelaskan variabilitas dari sekumpulan variabel. Variabel asli yang telah dimodifikasi atau disederhanakan menjadi beberapa variabel.

Ekstraksi ciri menggunakan PCA, Menghitung rata-rata: menjumlahkan M training image (T) lalu dibagi banyaknya M

Menghitung selisih rata-rata, yaitu dengan mencari selisih antara training image(T) dengan nilai mean. Dimana nilai hasil yang kurang dari 1 diubah menjadi nol

$$\phi = T_n - \mu$$

Menentukan matriks kovarian

$$C = \frac{1}{M} \sum_{n=1}^M \phi_n \phi_n^T = AA^T$$

Hitung kovarian:

$$C = A^T A, A = [\phi_1, \phi_2, \dots, \phi_n]$$

Selanjutnya mencari eigenvalue dan eigen vector. Nilai akan diurutkan dari besar ke terkecil untuk mendapatkan nilai PCA. $\det(C - \lambda I) = 0$ Mencari EigenValue $v = [\lambda_1 \lambda_2 \lambda_3]$ Mencari Eigen Vektor.

Setelah proses ekstraksi ciri citra, akan dibandingkan jarak kedekatannya training image dan testing image. Pada tahap pengujian data training image akan di uji dengan menggunakan PCA yang sama untuk mendapatkan hasil pengenalan dan klasifikasi dengan KNN.

KNN

K-Nearest Neighbour atau disingkat (KNN) adalah suatu cara klasifikasi terhadap objek berdasarkan data pelatihan atau tetangga yang paling dekat dengan objek tersebut. K-Nearest Neighbor, sebuah algoritma supervised learning yang proses pembelajarannya didasarkan pada data latih, dimana diharapkan output yang keluar sudah diketahui sebelumnya. Identifikasi Algoritma K-Nearest Neighbor dilakukan di data uji, yang mana nilai jarak dari data latih yang paling dekat dengan objek tersebut adalah hasil dari output klasifikasi Algoritma K-Nearest Neighbor. Klasifikasi dilakukan berdasarkan memori di data latih. Algoritma K-Nearest Neighbor menggunakan pengklasifikasi tetangga terdekat untuk memprediksi data baru, nilai prediksi dari nilai instance yang baru tersebut nantinya akan sebagai parameter untuk menentukan kelas. (Rahmadianto et al., 2019)

Pada metode KNN bertujuan

mengklasifikasikan suatu objek baru menggunakan atribut dan contoh dari data latih. Algoritma KNN dihitung dari jarak minimum data baru ke data latih. Hasil prediksi dari data tersebut diperoleh dari nilai yang paling sering muncul. Mencari jarak nilai k tetangga menggunakan Euclidian. Persamaan yang digunakan adalah yaitu Euclidian distance terdapat dalam persamaan berikut :

$$d_{p,q} = \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2}$$

$$d_{p,q} = \sqrt{\sum_{i=1}^n (pi - qi)^2}$$

$d_{p,q}$ = Hasil Euclidian distance (nilai k tetangga terdekat)

$p1q1...$ = nilai atribut pertama

$p2q2...$ = nilai atribut kedua n = urutan atribut

i = total perhitungan atribut

Ekstraksi Fitur

Ekstraksi fitur merupakan nilai fitur yang terkandung di dalam sebuah citra yang mewakili ciri khusus dari sebuah citra. Nilai yang didapatkan dari hasil pengestraksian fitur dari sebuah citra kemudian akan di proses untuk diidentifikasi. Ekstaksi Fitur yang digunakan di dalam penelitian ini adalah RGB (Red, Green, Blue) ke HSV (Hue, Saturation, Value).

RGB (Red, Green, Blue) adalah suatu model warna yang terdiri dari 3 warna primer yaitu merah, hijau, dan biru. Model warna RGB berdasar pada sistem koordinat Cartesius. RGB mempunyai nilai primer yaitu warna sekunder kuning, cyan, dan magenta. hitam di daerah asal sedangkan yang berada di sudut terjauh dari daerah asalnya adalah putih. Model RGB, skala abu-abu menghubungkan titik hitam ke titik putih yang berada disepanjang garis diagonal yang.

$$R=R/255$$

$$G=G/255$$

$$B=B/255$$

$$C \text{ max} = \max (R', G', B')$$

$$C \text{ min} = \min (R', G', B')$$

$$\Delta = C \text{ max} - C \text{ min}$$

HASIL DAN PEMBAHASAN

Input Citra

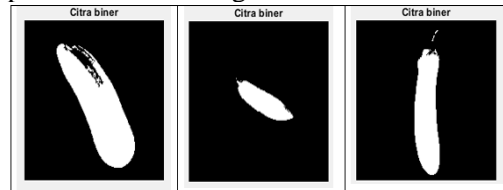
Inputan yang digunakan berupa citra berformat RGB sebanyak 26 data. Yaitu 16 citra latih dan 10 citra uji dari 3 jenis sayur yaitu mentimun, pare dan terong. Citra diambil dalam berbagai posisi yang berbeda. Citra latih merupakan citra yang digunakan untuk proses pelatihan sebelum tahap pengenalan.



Gambar Sampel Input Citra

Konversi Citra

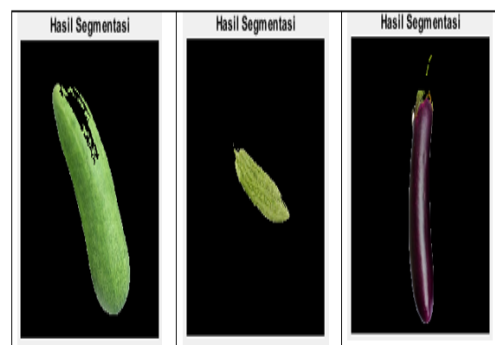
Citra RGB dikonversi menjadi citra biner, yaitu memisahkan background dengan objek. Proses ini bisa disebut proses Thresholding.



Gambar Konversi Citra Biner

Segmentasi Citra

Pada proses segmentasi citra dari sayur akan dicropping dan akan menghasilkan gabungan proses thresholding dan citra biner, sehingga bentuk citra terlihat lebih jelas.



Gambar Hasil Segmentasi

Ekstraksi Ciri

Ekstraksi ciri merupakan proses pengambilan/mengekstrak ciri dari sayuran mentimun, pare dan terong. Ekstraksi ciri yang digunakan RGB, HSV dan Area. Nilai RGB akan dikonversi menjadi HSV. Juga mencari nilai dari area yang digunakan untuk membedakan ukuran sayur satu dengan lainnya.

Hasil Ekstraksi Ciri		Hasil Ekstraksi Ciri		Hasil Ekstraksi Ciri	
Ciri	Nilai	Ciri	Nilai	Ciri	Nilai
1 Red	120.0937	1 Red	144.6941	1 Red	75.4252
2 Green	164.2598	2 Green	161.9126	2 Green	51.7731
3 Blue	85.2640	3 Blue	75.1772	3 Blue	62.0838
4 Hue	0.2613	4 Hue	0.1996	4 Hue	0.7847
5 Saturation	0.4949	5 Saturation	0.5475	5 Saturation	0.4805
6 Value	0.6442	6 Value	0.6350	6 Value	0.2987
7 Area	93874	7 Area	22942	7 Area	55211

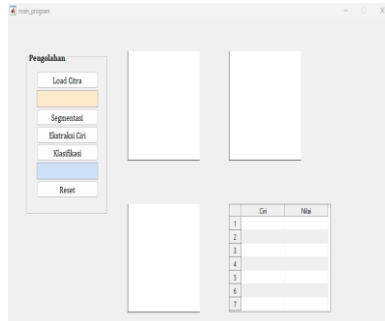
Gambar Hasil Ekstraksi Ciri

Klasifikasi

Klasifikasi citra sayuran mentimun, pare dan terong menggunakan KNN dilakukan untuk mendapatkan hasil klasifikasi setelah melalui pelatihan dan pengujian menggunakan PCA dan KNN.

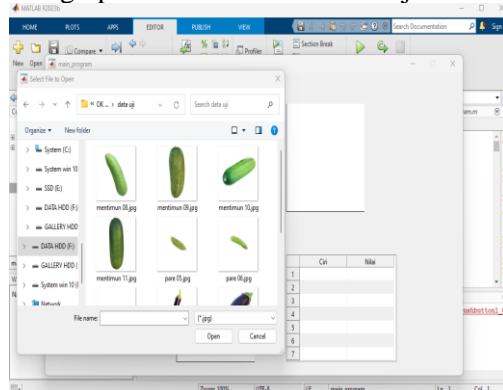
Tampilan GUI MATLAB

Tampilan awal GUI matlab untuk klasifikasi jenis sayuran.



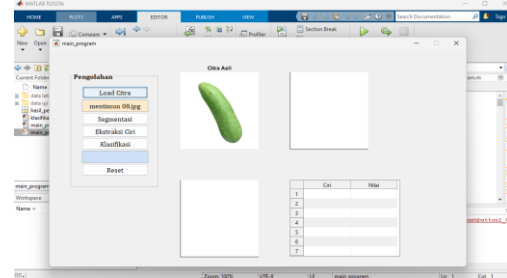
Gambar Tampilan Awal GUI

Tampilan GUI Matlab saat proses menginput citra dari folder data uji.



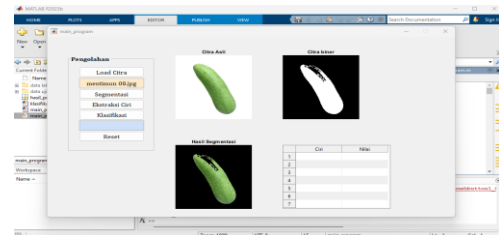
Gambar Menginput Citra

Hasil dari proses input citra.



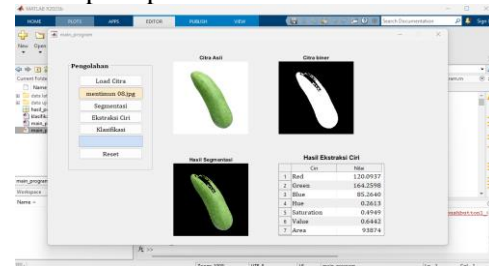
Gambar Hasil Input Citra

Hasil dari proses segmentasi, dimana citra RGB diubah menjadi citra biner.



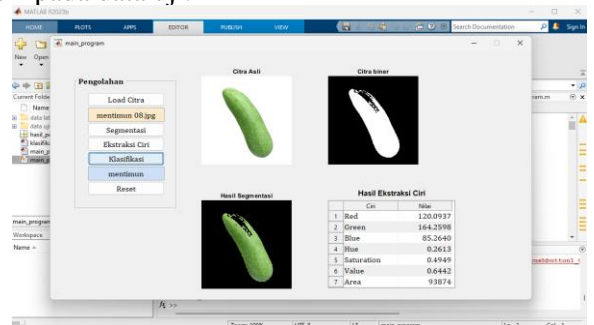
Gambar Proses Segmentasi

Hasil proses ekstraksi ciri dari jamur menunjukkan nilai RGB, HSV dan Area seperti pada tabel dibawah ini.



Gambar Ekstraksi Fitur PCA

Proses klasifikasi pada jenis mentimun, pare dan terong. Berdasarkan ekstraksi ciri yang didapat, algoritma akan menentukan jenis mentimun, pare dan terong yang mempunyai kemiripan di data latih pada data uji.



Gambar Hasil Klasifikasi

Pengujian Akurasi

Menghitung tingkat akurasi terhadap klasifikasi jenis sayuran mentimun, pare dan terong. Terdapat 10 data uji citra sayuran. Yaitu masing-masing berjumlah 4 jenis mentimun, 3 jenis pare dan 3 jenis terong.

$$\text{akurasi} = \frac{\text{banyak data yang sesuai}}{\text{banyak data masukan}} \times 100\%$$

$$\text{akurasi} = \frac{9}{10} \times 100\% = 90\%$$

Berdasarkan perhitungan diatas, diperoleh akurasi sebesar 92%. Terdapat 23 citra yang benar.

No.	Kelas Asli	Kelas Keluaran
1	Mentimun 08	Mentimun
2	Mentimun 09	Mentimun
3	Mentimun 10	Mentimun
4	Mentimun 11	Mentimun
5	Pare 05	Pare
6	Pare 06	Pare
7	Pare 07	Mentimun
8	Terong 06	Terong
9	Terong 07	Terong
10	Terong 08	Terong

SIMPULAN

Pada penelitian ini dapat disimpulkan bahwa :

1. Penerapan Algoritma Principal Component Analysis (PCA) dan K-nearest neighbors (KNN) mampu membedakan 4 jenis mentimun, 2 jenis pare dan 3 jenis terong.
2. Akurasi yang dihasilkan dari Algoritma Principal Component Analysis (PCA) dan Knearest neighbors (KNN) untuk klasifikasi adalah 90% dari 10 data uji

DAFTAR PUSTAKA

- Permana Putra, I. (2020). Scleroderma spp. in Indonesia : Poisoning Case and Potential Utilization. *Justek : Jurnal Sains Dan Teknologi*, 3(2), 37. <https://doi.org/10.31764/justek.v3i2.3517>
- Wardhani, I. P., & Widayati, S. (2019). Segmentasi Warna Citra HSV dan Deteksi Objek Kupu-Kupu Dengan. 3
- Novianto, D., & Sugihartono, T. (2020). Sistem Deteksi Kualitas Buah Jambu Air Berdasarkan Warna Kulit Menggunakan Algoritma Principal Component Analysis (Pca) dan K-Nearest Neighbor (K-NN). 11(2), 42–47.
- Widians, J. A., Pakpahan, H. S., Budiman, E., Haviluddin, H., & Soleha, M. (2019). Klasifikasi Jenis Bawang Menggunakan Metode K-Nearest Neighbor Berdasarkan Ekstraksi Fitur Bentuk dan Tekstur. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 3(2), 139. <https://doi.org/10.30872/jurti.v3i2.3213>
- Hasan, M. A., & Liliana, D. Y. (2020). Pengenalan Motif Songket Palembang Menggunakan Deteksi Tepi Canny, PCA dan KNN. *Multinetics*, 6(1), 1–7. <https://doi.org/10.32722/multinetics.v6i1.27000>
- F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 273–281, 2024, doi: 10.57152/malcom.v4i1.1085
- D. Nurnaningsih, D. Alamsyah, A. Herdiansah, A. Aristo, and J. Sinlae, "Identifikasi Citra Tanaman Obat Jenis Rimpang dengan Euclidean Distance Berdasarkan Ciri Bentuk

- dan Tekstur,” vol. 3, no. 3, pp. 171–178, 2021, doi: 10.47065/bits.v3i3.1019.
- Primandari, P. N., & Hardiansyah, B. (2018). Ekstraksi Fitur Menggunakan Principal Component Analysis (PCA). Seminar Nasional Hasil Penelitian Dan Pengabdian Masyarakat, 1(1), 66–74
- Q. A. A'yuniyah and M. Reza, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru,” Indones. J. Inform. Res. Softw. Eng., vol. 3, no. 1, pp. 39–45, 2023, doi: 10.57152/ijirse.v3i1.484.
- “Klasifikasi Jenis Sayuran Menggunakan Algoritma PCA dan KNN | Pemrograman Matlab.” <https://pemrogramanmatlab.com/2019/01/01/klasifikasi-jenis-sayuran-menggunakan-algoritma-pca-dan-knn/> accessed Januari, 11, 2024
- Andrean Nugraha, R., Wahyu Hidayat, E., & Nur Shofa, R. (2023). Klasifikasi Jenis Buah Jambu Biji Menggunakan Algoritma Principal Component Analysis dan K-Nearest Neighbor. Generation Journal, 7(1), 1–7. <https://doi.org/10.29407/gj.v7i1.17900>
- Hasym, I. E., & Susilawati, I. (2021). Konvergensi Teknologi dan Sistem Informasi Klasifikasi Jenis Ikan Cupang Menggunakan Algoritma Principal Component Analysis (PCA) Dan K-Nearest Neighbors (KNN). KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi, 168–179.
- Nurfitri, K., Pradana, A. D., & Widaningrum, I. (2021). Jurnal Rekayasa Teknologi dan Komputasi PENERAPAN ALGORITMA PRINCIPAL COMPONENT ANALYSIS (PCA) DAN K-NEAREST NEIGHBORS (KNN) PADA KLASIFIKASI. Jurnal Rekayasa Teknologi Dan Komputasi, 1(1), 1–19