

Klasifikasi Kualitas Buah Apel Dengan Algoritma *K-Nearest Neighbor* (*K-NN*) Menggunakan Bahasa Pemrograman *Python*

Puji Astuti^{1*}

¹Universitas Nusa Mandiri
Jln. Jatiwaringin Raya No.02 Rt.08 Rw.013 Kelurahan Cipinang Melayu Kecamatan Makasar, Jakarta Timur

e-mail: puji.pat@nusamandiri.ac.id

(*) Corresponding Author

Artikel Info : Diterima : 22-03-2024 | Direvisi : 04-07-2024 | Disetujui : 17-07-2024

Abstrak - Buah termasuk asupan yang penting bagi tubuh manusia, buah apel termasuk kedalam buah yang digemari oleh masyarakat Indonesia. Untuk itu perlu penyediaan buah apel dengan kualitas yang baik, sehingga dapat bermanfaat bagi tubuh. Dengan menggunakan metode *K-NN* yang dirasa mampu melatih data dengan cepat dan efektif untuk data training dan data testing dalam jumlah besar. Penelitian ini dimulai dari pengumpulan dataset yang didapat dari <https://www.kaggle.com/>, kemudian melakukan proses *preprocessing* yang dilanjutkan dengan memisahkan data training dan data testing dengan komposisi 25% data testing dan 75% data training. Kemudian metode *K-NN* diterapkan pada penelitian ini untuk diklasifikasikan berdasarkan beberapa kriteria yang sudah ada, sehingga mendapatkan hasil evaluasi kinerja *K-NN* dengan nilai *accuracy* yang sudah di hitung dengan pemrograman *python*. Dalam melakukan implementasi *datamining* menggunakan Bahasa pemrograman *Python* yaitu dengan memanfaatkan *library* yang sudah disediakan sebagai proses untuk memudahkan implementasi *machine learning*. Dari pengujian *confusion matrix* terdapat 441 data yang diprediksi dengan data benar, dan 440 data yang diprediksi salah. Sedangkan untuk 54 serta 65 data diprediksi kurang tepat dari 1000 data testing. Sehingga didapat nilai *accuracy* dengan metode *K-NN* yaitu sebesar 0.88 atau 88%. Hal ini terlihat bahwa metode *K-NN* dapat bekerja dengan baik, cepat dan efisien dalam melatih jumlah data yang besar.

Kata Kunci : Klasifikasi, *K-NN*, *Python*

Abstracts - Fruit is an important intake for the human body, apples are included in the fruit favored by the people of Indonesia. For this reason, it is necessary to provide apples of good quality, so that they can benefit the body. By using the *k-NN* method that is considered able to train data quickly and effectively for training data and testing data in large quantities. This study began from the collection of datasets obtained from <https://www.kaggle.com/>, then perform a preprocessing process followed by separating the training data and testing data with a composition of 25% testing data and 75% training data. Then the *k-NN* method is applied to this study to be classified based on several existing criteria, so as to obtain the results of performance evaluation *K-NN* with the value of accuracy that has been calculated with python programming. In implementing *datamining* using Python programming language by utilizing the library that has been provided as a process to facilitate the implementation of machine learning. From The Matrix confusion test, there are 441 data predicted with true data, and 440 data predicted incorrectly. As for the 54 and 65 data predicted to be less precise than 1000 testing data. So that the accuracy value obtained by the *k-NN* method is equal to 0.88 or 88%. It is seen that the *k-NN* method can work well, quickly and efficiently in training large amounts of data.

Keywords : Classification, *K-NN*, *Python*

PENDAHULUAN

Buah termasuk asupan yang penting bagi tubuh manusia, buah apel adalah buah yang digemari oleh masyarakat Indonesia. Untuk itu perlu penyediaan buah apel dengan kualitas yang baik, sehingga dapat bermanfaat bagi tubuh (Ridho et al., 2022). Apel termasuk buah yang dengan sejuta macam manfaat, bercitarasa manis, asam, crunchy dan sedikit juicy. Buah ini mengandung serat, kalium, vitamin, magenisium, dan fosfor. Pada buah apel ini dapat menangkal beberapa macam penyakit. Menentukan kematangan buah nanas menggunakan model *K-NN* dengan pendekatan $k=5$ dan $k=9$ mendapatkan hasil *accuracy* sebesar 73% (Reswan et al., 2024). Klasifikasi



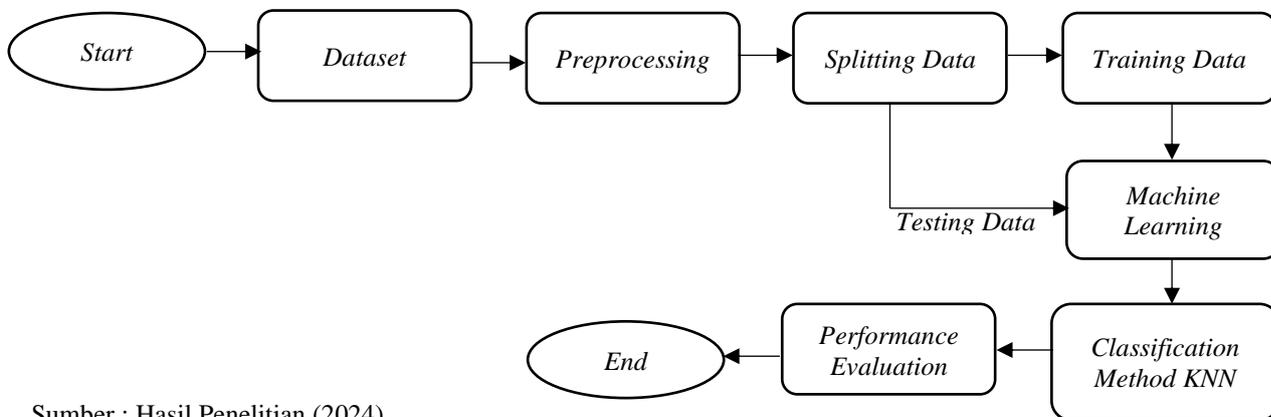
penentuan buah pisang raja berdasarkan fitur warna RGB dengan pendekatan model algoritma *naïve bayes* pada kategori rasa manis mendapatkan nilai 100% sedangkan untuk kategori sepat dan hambar 80% sehingga rata-rata *accuracy* yang didapat mencapai nilai 86.66% (M. Afriansyah et al., 2024). Mengonsumsi buah apel dapat memilih buah yang masih segar dengan kualitas tidak terlalu muda dan tidak busuk. Menggunakan metode K-NN menggunakan fitur meanR, meanG, dengan 117 data memiliki nilai *accuracy* sebesar 88.37% dengan tingkat kemiripan K5 (Suryanti et al., 2024). *K-Nearest Neighbor* (K-NN) merupakan metode yang dapat digunakan untuk pengklasifikasian objek baru berdasarkan jumlah K dengan bilangan terdekat. Algoritma K-NN relative lebih mudah dipahami dan dapat mengklasifikasikan dengan data dalam jumlah besar (Azizah et al., 2023).

Memanfaatkan fitur ekstraksi untuk kematangan buah pinang berdasarkan warna dan tekstur dengan model *Support Vector Machine* (SVM) dengan nilai *accuracy* sebesar 90.90% (Maneno et al., 2023). Klasifikasi buah apel berdasarkan warna RGB untuk menentukan jenis buah apel dengan metode *Naive Bayes*, dengan hasil pengujian dengan tingkat kemanisan 100%, sedikit manis 86.68%, keasaman dengan nilai 80%, dan rata-rata nilai *accuracy* dengan metode *Naive Bayes* memiliki nilai 88.88% (Afriansyah et al., 2023).

Penentuan tingkat kematangan buah kopi dilihat dari tiga tingkatan yaitu matang, setengah matang dan mentah sehingga warna buah kopi dapat menjadi tolak ukur dalam mengenali kualitas buah kopi. Terdapat 300 data untuk menentukan tingkat kematangan buah kopi dengan menggunakan metode K-NN kedekatan tetangga dengan nilai 3 didapat *accuracy* sebesar 95% (Farhan & Setiaji, 2023). Dalam penelitian yang dibuat penulis menerapkan metode K-NN karena dianggap mampu memberikan solusi terhadap kualitas buah apel dalam proses data latih yang besar, dengan Delapan (8) kriteria dan label satu (1). Menggunakan data yang diambil dari <https://www.kaggle.com/> sebanyak 4000 data, dengan pendekatan nilai K pada K-NN dan diambil nilai tetangga terdekat dengan menggunakan Bahasa pemrograman *python*.

METODE PENELITIAN

Penelitian ini dimulai dari pengumpulan *dataset*, kemudian melakukan proses *preprocessing* yang dilanjutkan dengan memisahkan data *training* dan data *testing* dengan komposisi 25% data *testing* dan 75% data *training*. Kemudian metode K-NN diterapkan pada penelitian ini untuk diklasifikasikan berdasarkan beberapa kriteria yang sudah ada, sehingga mendapatkan hasil evaluasi kinerja K-NN dengan nilai *accuracy* yang sudah di hitung dengan pemrograman *python*. Berikut kerangka penelitian :



Sumber : Hasil Penelitian (2024)

Gambar 1. Kerangka Penelitian

1. Dataset

Dataset yaitu merujuk kepada suatu kumpulan data. *Dataset* biasanya lebih dari satu variabel serta menelisis suatu topik tertentu. *Dataset* juga bisa disebut dengan kumpulan data yang berasal dari informasi masa lalu dan siap untuk di olah menjadi informasi baru (Mustakim, 2022).

2. Preprocessing

Tahapan awal dari pengolahan data yaitu *preprocessing*, dimana data yang sudah dikumpulkan dalam bentuk data mentah diolah terlebih dahulu (Nugraha & Astuti, 2023). Proses dari *preprocessing* yaitu menghilangkan data yang tidak bisa digunakan sehingga lebih dipahami oleh sistem komputer. Adapun tahapan dalam proses *preprocessing* ada 4 yaitu :

a. *Data Cleaning*

Data mentah yang sudah dikumpulkan di proses dengan data *cleaning* fungsinya untuk membersihkan data yang tidak lengkap, menyeleksi kembali data.

- b. *Data Integration*
Menggabungkan beberapa data dalam suatu *dataset*.
- c. *Transformasi Data*
Transformasi data dapat dikatakan dengan menyamakan isi dari data, data yang berbeda antara numerik dan *alphabet* harus disamakan formatnya, agar memudahkan dalam proses klasifikasi data.
- d. *Mengurangi Data*
Tahapan akhir yaitu mengurangi data yang redundansi, dengan tidak merubah data asli.

3. *Data Training dan Data Testing*

Data *training* yaitu data asli sesuai dengan fakta dan keadaan yang ada sedangkan data *testing* yaitu data yang digunakan untuk melakukan atau mengukur sampai mana klasifikasi berhasil dengan baik (Baiq Nurul Azmi et al., 2023).

4. *Algoritma K-NN*

K-Nearest Neighbor atau K-NN merupakan algoritma klasifikasi dalam *machine learning* sebagai acuan untuk menentukan kategori atau objek baru. Algoritma K-NN menghitung jarak antara objek baru serta untuk menentukan nilai K yang valid (Fuansah et al., 2023). Rumus yang digunakan dalam menghitung kedekatan jarak antara objek baru dan data yaitu :

$$\text{Similarity}(T, S) = \frac{\sum_{i=1}^n f(T_i, S_i) * W_i}{W_i}$$

Ket :

T = Kasus Baru

S = Kasus yang ada dalam penyimpanan

n = Jumlah atribut dalam setiap kasus

i = Atribut individu antara 1 s.d n

f = Fungsi *similarity* atribut i antara kasus T dan kasus S

w = Bobot yang diberikan pada atribut ke-i

Pendekatan K-NN dilakukan dalam menemukan nilai parameter K. Jarak antara data baru dan semua data dalam set pelatihan dihitung. Kemudian perbandingan dengan data uji dilakukan dan *confusion matrix* dihasilkan (Situmorang et al., 2023). Evaluasi dalam penelitian ini menggunakan *confusion matrix*. *Confusion Matrix* digunakan untuk memperoleh hasil dari nilai *precision*, *recall* dan *nilai accuracy*. Nilai dari *Confusion Matrix* biasanya ditunjukkan dalam persen (%) (Rahayu et al., 2021).

5. *Pengujian*

Proses pengujian hingga mendapatkan nilai *accuracy* menggunakan Bahasa pemrograman *Python*. *Python* adalah Bahasa pemrograman tingkat tinggi yang dibuat oleh Guido Van Rossum dan dirilis pada tahun 1991. *Python* merupakan Bahasa yang sedang populer saat ini. Bahkan *python* juga merupakan Bahasa pemrograman yang multifungsi dibidang *Machine Learning* dan *Deep Learning* (Alfarizi et al., 2023). *Machine Learning* bentuk analisis dari suatu masalah dengan menggunakan model matematika yang digunakan untuk menangkap fitur sebagai kunci saat mesin sedang bekerja, kemudian model matematika menghasilkan solusi dari suatu masalah (Mulia et al., 2023). Sedangkan *Deep Learning* merupakan metode *learning* dimana memanfaatkan *multiple non-linear transformation*, serta *deep learning* dapat digabungkan dengan *machine learning dengan AI (Artificial neural network)* (Raup et al., 2022).

HASIL DAN PEMBAHASAN

1. *Pengumpulan Data*

Dataset diambil dari beberapa kumpulan data yang tersimpan didalam <https://www.kaggle.com/> data kualitas buah apel diambil dari <https://www.kaggle.com/> dengan dataset sebanyak 4000 data yang sudah diklasifikasikan berdasarkan kriteria dan label.

	A_id	Size	Weight	Sweetness	Crunchiness	Juiciness	Ripeness	Acidity	Quality
0	0	-3.970049	-2.512336	5.346330	-1.012009	1.844900	0.329840	-0.491590	good
1	1	-1.195217	-2.839257	3.664059	1.588232	0.853286	0.867530	-0.722809	good
2	2	-0.292024	-1.351282	-1.738429	-0.342616	2.838636	-0.038033	2.621636	bad
3	3	-0.657196	-2.271627	1.324874	-0.097875	3.637970	-3.413761	0.790723	good
4	4	1.364217	-1.296612	-0.384658	-0.553006	3.030874	-1.303849	0.501984	good
...
3995	3995	0.059386	-1.067408	-3.714549	0.473052	1.697996	2.244055	0.137784	bad
3996	3996	-0.293118	1.949253	-0.204020	-0.640196	0.024523	-1.087900	1.854235	good
3997	3997	-2.634515	-2.138247	-2.440461	0.657223	2.199709	4.763859	-1.334611	bad
3998	3998	-4.008004	-1.779337	2.366397	-0.200329	2.161435	0.214488	-2.229720	good
3999	3999	0.278540	-1.715505	0.121217	-1.154075	1.266677	-0.776571	1.599796	good

4000 rows x 9 columns

Sumber : Hasil Penelitian (2024)

Gambar 2. Dataset

2. Preprocessing

Preprocessing dengan menggunakan Bahasa pemrograman python dimana proses preprocessing ini dilakukan untuk membersihkan apakah ada data yang tidak sama, serta membuat data agar isi dari data tersebut sama untuk memudahkan dalam proses pengklasifikasian data.

```
[ ] x = buah.iloc[:, [1, 2, 3, 4, 5, 6, 7]].values
    y = buah.iloc[:, -1].values

print(x)

[[[-3.97004852 -2.51233638  5.34632961 ...  1.84490036  0.3298398
  -0.49159048]
 [-1.19521719 -2.83925653  3.66405876 ...  0.8532858  0.86753008
  -0.72280937]
 [-0.29202386 -1.35128199 -1.73842916 ...  2.83863551 -0.03803333
  2.62163647]
 ...
 [-2.6345153  -2.13824672 -2.44046129 ...  2.19970859  4.76385918
  -1.33461139]
 [-4.00800374 -1.77933711  2.36639697 ...  2.16143512  0.21448838
  -2.22971981]
 [ 0.27853965 -1.71550503  0.12121725 ...  1.2666774  -0.77657147
  1.59979646]]
```

Sumber : Hasil Penelitian (2024)

Gambar 3. Proses Preprocessing

Terdapat 9 kriteria dari data buah apel yang diambil dari <https://www.kaggle.com>, dimana untuk menentukan kualitas buah apel digunakan semua kriteria yang akan diolah berdasarkan masing-masing kriteria. Dilihat dari bentuk, ukuran, tingkat kemanisan, tingkat kadar air, tingkat keasaman dan sampai ke hasil berupa kualitas dari buah apel tersebut.

3. Data Training dan Data Testing

Dari 4000 data yang didapat dari <https://www.kaggle.com>, dibagi menjadi beberapa data antara data training dan data testing yang akan diklasifikasikan menjadi sebuah data yang terstruktur. Dari 4000 data diambil 25% untuk data testing dan selebihnya untuk diuji pada data training.

```
[ ] from sklearn.model_selection import train_test_split
    x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=0)
```

Sumber : Hasil Penelitian (2024)

Gambar 4. Data Training dan Data Testing

```
[ ] len(y_train)
3000

[ ] len(y)
4000

[ ] len(y_test)
1000
```

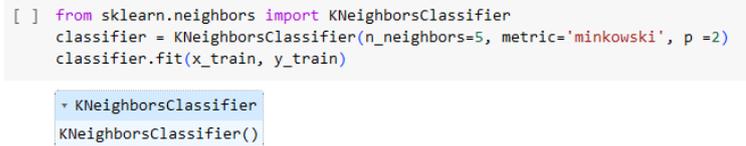
Sumber : Hasil Penelitian (2024)

Gambar 5. Hasil Pembagian Data Training dan Data Testing

4. Algoritma K-NN

Setelah melakukan *preprocessing* dan kemudian melakukan pembagian data antara data *training* dan data *testing* dari 4000 data didapat pada gambar 5 bahwa data *testing* menjadi 1000 data dan data *training* menjadi 3000 data. Pada *machine learning* Bahasa pemrograman *python* dilakukan pemanggilan fungsi K-NN, hal ini agar terbaca pada saat *accuracy* data metode yang digunakan dengan menggunakan *minkowski* sebagai nilai ketetanggaan terdekat yaitu 2.

```
[ ] from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=5, metric='minkowski', p =2)
classifier.fit(x_train, y_train)
```



Sumber : Hasil Penelitian (2024)

Gambar 6. Proses Algoritma K-NN

5. Hasil Pengujian

Setelah melakukan prediksi atau klasifikasi dengan algoritma K-NN, maka selanjutnya dilakukan evaluasi dan validasi untuk melihat seberapa akurat data kualitas buah apel dengan menggunakan algoritma K-NN. Langkah awal membuat confusion matrix dari data testing didapat hasil.

```
[ ] from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

```
[[441  65]
 [ 54 440]]
```

Sumber : Hasil Penelitian (2024)

Gambar 7. Hasil Confusion Matrix

Terdapat 441 data yang diprediksi dengan data benar, dan 440 data yang diprediksi salah. Sedangkan untuk 54 serta 65 data diprediksi kurang tepat dari 1000 data testing.

Setelah itu dilanjutkan dengan melihat klasifikasi *report* untuk melihat *accuracy* secara detail.

```
from sklearn.metrics import classification_report
akurasi = classification_report(y_test, y_pred)
print(akurasi)
```

Sumber : Hasil Penelitian (2024)

Gambar 8. Mencari Nilai Accuracy

Tabel 1. Hasil Accuracy dengan Metode K-NN

Sumber : Hasil Penelitian (2024)

Dari hasil *accuracy* diatas terlihat *precision* perbandingan data *precision*, *recall*, *f1-score* dengan hasil *accuracy* sebesar 0.88% atau 88%.

KESIMPULAN

Hasil *accuracy* yang dilakukan dengan algoritma K-NN dengan jumlah data 1000 data testing pada kualitas buah apel bahwa didapat nilai sebesar 0.88 atau 88%. Hal ini menunjukkan bahwa klasifikasi dengan menggunakan data sebanyak 1000 didapatkan nilai *accuracy* yang cukup baik. Pada penelitian selanjutnya dapat dilakukn

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Bad</i>	0.89	0.87	0.88	506
<i>Good</i>	0.87	0.89	0.88	494
<i>Accuracy</i>			0.88	1000
<i>Macro avg</i>	0.88	0.88	0.88	1000
<i>Weighted avg</i>	0.88	0.88	0.88	1000

dengan data yang sama dapat di terapkan metode SVM atau *Naive Bayes* dengan optimasi seperti *Genetic Algorithm*

(GA) atau dengan *feature selection* PSO.

REFERENSI

- Afriansyah, M., Saputra, J., Sa'adati, Y., & Valian Yoga Pudya Ardhana. (2023). Optimasi Algoritma Naive Bayes Untuk Klasifikasi Buah Apel Berdasarkan Fitur Warna RGB. *Bulletin of Computer Science Research*, 3(3), 242–249. <https://doi.org/10.47065/bulletincsr.v3i3.251>
- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, 2(1), 1–6.
- Azizah, N., Riyad Firdaus, M., Suyaningsih, R., & Indrayatna, F. (2023). Penerapan Algoritma Klasifikasi K-Nearest Neighbor pada Penyakit Diabetes. *Prosiding Seminar Nasional Statistika Aktuaria*, 2(1), 119–126. [https://prosidingnsa.statistics.unpad.ac.id/?journal=prosidingnsa&page=article&op=view&path\[\]=344](https://prosidingnsa.statistics.unpad.ac.id/?journal=prosidingnsa&page=article&op=view&path[]=344)
- Baiq Nurul Azmi, Arief Hermawan, & Donny Avianto. (2023). Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 281–290. <https://doi.org/10.35746/jtim.v4i4.298>
- Farhan, N. M., & Setiaji, B. (2023). Indonesian Journal of Computer Science. *Indonesian Journal of Computer Science*, 12(2), 284–301. <http://ijcs.stmkindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- Fuansah, Y. S., Meileni, H., Novianti, L., Informatika, J. M., Sriwijaya, P. N., Neighbor, K., & Matrix, C. (2023). Implementasi Metode K-Nearest Neighbor untuk Menentukan Klasifikasi Status Ekonomi Penerima Bantuan. *JUPITER*, 15(2), 929–938.
- M. Afriansyah, Joni Saputra, Ardhana, V. Y. P., & Yuan Sa'adati. (2024). Algoritma Naive Bayes Yang Efisien Untuk Klasifikasi Buah Pisang Raja Berdasarkan Fitur Warna. *Journal of Information Systems Management and Digital Business*, 1(2), 236–248. <https://doi.org/10.59407/jismdb.v1i2.438>
- Maneno, R., Baso, B., Manek, P. G., & Fallo, K. (2023). Deteksi Tingkat Kematangan Buah Pinang Menggunakan Metode Support Vector Machine Berdasarkan Warna Dan Tekstur. *Journal of Information and Technology*, 3(2), 60–66. <https://doi.org/10.32938/jitu.v3i2.5323>
- Mulia, S. B., Wisma Nugraha, N., Robbani, M. H., Otomasi, T., Mekatronika, M. &, & Manufaktur Bandung, P. (2023). Implementasi Machine Learning Untuk Identifikasi Orang Batuk/Bersin. *Journal of Energy and Electrical Engineering (Jeee)*, 81(2), 81–86.
- Mustakim. (2022). *Empat Sumber Dataset untuk Belajar dan Penelitian Bidang Data Mining*. <https://Mustakim.Irpi.or.Id/2022/05/18/Empat-Sumber-Dataset-Untuk-Belajar-Dan-Penelitian-Bidang-Data-Mining/>.
- Nugraha, D., & Astuti, P. (2023). Analisis Sentimen Cyberbullying Pada Sosial Media Instagram Menggunakan Metode Support Vector Machine. *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS: Journal of Information System*, 8(2), 153. <https://doi.org/10.51211/isbi.v8i2.2535>
- Rahayu, W. I., Prianto, C., & Novia, E. A. (2021). Perbandingan Algoritma K-Means dan Naive Bayes untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan pada PT. Pertamina (Persero). *Jurnal Teknik Informatika*, 13(2), 1–8.
- Raup, A., Ridwan, W., Khoeriyah, Y., Supiana, S., & Zaqiah, Q. Y. (2022). Deep Learning dan Penerapannya dalam Pembelajaran. *JiIP - Jurnal Ilmiah Ilmu Pendidikan*, 5(9), 3258–3267. <https://doi.org/10.54371/jiip.v5i9.805>
- Reswan, Y., Toyib, R., Witriyono, H., & Anggraini, A. (2024). Klasifikasi Tingkat Kematangan Buah Nanas Berdasarkan Fitur Warna Menggunakan Metode K-Nearest Neighbor (KNN). *Jurnal Media Infotama*, 20(1), 280–287. <https://jurnal.unived.ac.id/index.php/jmi/article/view/5689>
- Ridho, A., Setyadjit, K., & Hariadi, B. (2022). Menentukan kualitas buah apel malang berdasarkan kulitnya memanfaatkan pengolahan citra digital. *Jurnal Universitas 1945 Surabaya*, 25(2), 1–12.
- Situmorang, R., Rahayu, W. I., Nuraini, R., Fathonah, S., Studi, P., Terapan, S., Informatika, T., Logistik, U., Internasional, B., Bandung, K., & Barat, J. (2023). *Model Algoritma K-Nearest Neighbor (K-Nn) Dan Naive Bayes*. 7(1), 250–254.
- Suryanti, C., Rohman, Mg., Kunci -Buah apel, K., & Knn, K. (2024). Klasifikasi Kualitas Buah Apel Berdasarkan Warna dan Bentuk Menggunakan Metode KNN. *Generation Journal*, 8(1), 2580–4952.