



Penggunaan Bahasa Pemrograman Python Dalam Menentukan RGB Pada Buah Apel Dengan Algoritma K-Nearest Neighbor (K-NN)

Chairul Imam¹, Feliani Waruwu², Okhta Alex Ndruru³, Indra Putra Jaya Laia⁴

^{1,2,3,4} Teknologi Informasi/ Fakultas Teknik & Ilmu Komputer, Universitas Pembinaan Masyarakat Indonesia, Indonesia

Correspondensi : chairulimam313@gmail.com

ABSTRAK

Buah sangat penting bagi kesehatan manusia, dan apel adalah salah satu buah yang sangat disukai oleh masyarakat Indonesia. Oleh karena itu, penyediaan apel berkualitas baik sangat diperlukan agar memberikan manfaat bagi kesehatan. Dalam penelitian ini, digunakan metode K-NN yang dianggap efisien untuk melatih data dengan cepat pada data pelatihan dan pengujian yang berjumlah besar. Penelitian dimulai dengan pengumpulan dataset dari <https://www.kaggle.com/>, dilanjutkan dengan proses pra-pemrosesan, kemudian memisahkan data menjadi data pelatihan dan pengujian dengan proporsi 25% untuk pengujian dan 75% untuk pelatihan. Setelah itu, metode K-NN diterapkan untuk mengklasifikasikan data berdasarkan kriteria yang sudah ditentukan, menghasilkan evaluasi kinerja K-NN yang nilainya dihitung menggunakan pemrograman Python. Selama implementasi data mining dalam Bahasa pemrograman Python, digunakan library yang tersedia untuk mempermudah proses penerapan machine learning. Dari analisis confusion matrix, terdapat 441 data yang diprediksi benar, dan 440 data diprediksi salah. Selain itu, terdapat 54 dan 65 data yang diperkirakan tidak tepat dari 1000 data pengujian. Sehingga, nilai akurasi dengan metode K-NN diperoleh sebesar 0.88 atau 88%. Ini menunjukkan bahwa metode K-NN mampu bekerja dengan baik, cepat, dan efisien dalam melatih jumlah data yang besar.

Kata kunci: Klasifikasi, K-NN, Python.

PENDAHULUAN

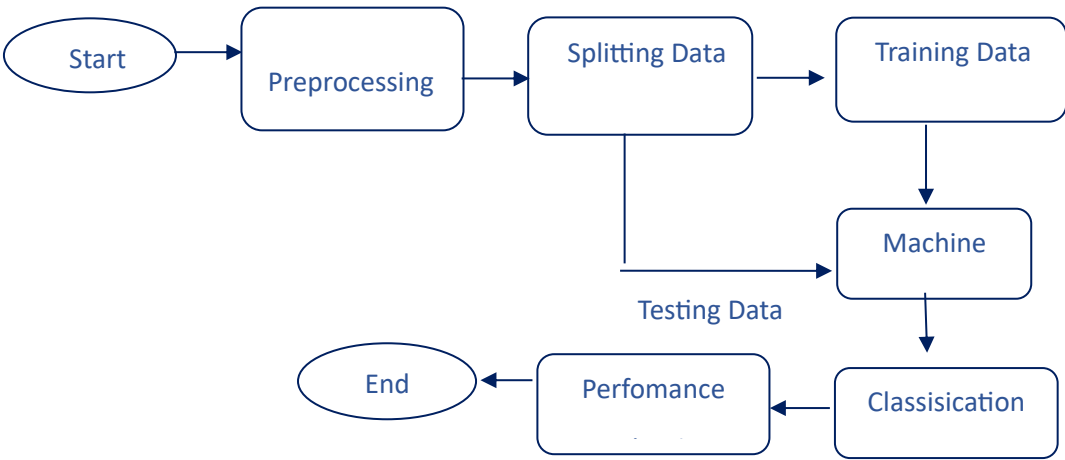
Buah-buahan merupakan komponen penting dalam diet manusia, dan apel adalah salah satu buah yang banyak disukai oleh masyarakat Indonesia. Oleh karena itu, sangat penting untuk menyediakan apel berkualitas tinggi agar dapat memberikan manfaat, dengan rasa manis, asam, tekstur renyah, dan sedikit berair. Buah ini akan serat, kalium, vitamin, magnesium, dan fosfor. Terdapat beberapa jenis penyakit yang dapat dicegah dengan mengonsumsi apel ini. Menentukan tingkat kematangan buah nanas dengan menggunakan model K-NN melalui pendekatan K=5 dan K=9 menghasilkan tingkat akurasi sebesar 73% (Reswan et al., 2024). Klasifikasi kategori rasa manis memperoleh skor 100% sementara kategori rasa sepat dan hambar mencapai 80% sehingga rata-rata akurasi yang diperoleh adalah 86.66%. Saat mengonsumsi buah apel, penting untuk memilih buah yang masih segar, dengan kualitas tidak terlalu muda dan tidak busuk. Penggunaan metode K-NN dengan fitur meanR dan meanG , berdasarkan 117 data, menghasilkan tingkat akurasi 88.37% dengan tingkat kesamaan. K-Nearest Neighbor (K-NN) adalah metode yang dapat digunakan untuk mengklasifikasikan objek baru berdasarkan nilai dengan pihak terdekat. Algoritma K-NN relatif mudah dipahami dan mampu mengklasifikasikan data dalam jumlah besar.

Memanfaatkan fitur ekstraksi untuk menilai kematangan buah pinang berdasarkan aspek warna dan tekstur menggunakan model Support Vector Machine berhasil mencapai tingkat akurasi sebesar 90.90%. Klasifikasi buah apel didasarkan pada analisis warna RGB untuk mengidentifikasi jenis apel dengan menerapkan metode Naive Bayes, di mana hasil pengujian menunjukkan tingkat kemanisan dengan nilai 80%. Rata-rata tingkat akurasi yang diperoleh dengan metode Naive Bayes adalah 88.88%. Penentuan kematangan buah kopi dapat dibedakan menjadi tiga kategori, yaitu matang, setengah matang, dan mentah. Oleh karena itu, warna buah kopi berfungsi

sebagai indikator untuk menganalisis kualitasnya. Penelitian ini menggunakan 300 data untuk menentukan tingkat kematangan buah kopi melalui metode K-NN yang merupakan teknik kedekatan tetangga dengan nilai K sebesar 3, menghasilkan tingkat akurasi sebesar 95% dalam studi ini, penulis memilih metode K-NN karena dianggap efektif untuk menilai kualitas buah apel pada data pelatihan yang besar dengan menggunakan delapan (8) kriteria dan satu (1) label. Data yang digunakan diambil sebanyak 4000 data dengan pendekatan nilai K dalam K-NN, serta nilai tetangga terdekat diolah menggunakan bahasa pemrograman python.

METODE PENELITIAN

Studi ini dimulai dengan pengumpulan data, lalu dilanjutkan dengan proses preprocessing yang melibatkan pemisahan antara data pelatihan dan data pengujian dengan rasio 25% untuk data pelatihan dan data pengujian dan 75% untuk data pelatihan. Selanjutnya metode K-NN diimplementasikan dalam penelitian ini untuk melakukan klasifikasi berdasarkan beberapa kriteria yang telah ditentukan, sehingga diperoleh hasil evaluasi kinerja K-NN dengan nilai akurasi yang dihitung menggunakan pemrograman python. Berikut adalah kerangka penelitian:



Gambar 1. Kerangka Penelitian

1. Kumpulan Data
Kumpulan data merujuk pada sekumpulan informasi. kumpulan data umumnya melibatkan lebih dari satu variabel dan mengeksplorasi suatu tema tertentu. kumpulan data juga dapat diartikan sebagai koleksi informasi data yang berasal dari masa lalu dan siap untuk diproses menjadi pengetahuan yang baru (Mustakim, 2022).
2. Preprocessing
Langkah pertama dalam pengolahan data adalah preprocessing, di mana data yang sudah dikumpulkan dalam keadaan mentah diproses sebelumnya. Tahapan preprocessing melibatkan penghilangan data yang tidak relevan agar lebih dapat dimengerti oleh sistem komputer. Terdapat empat langkah dalam proses processing yaitu:
 - a. Data
data mentah yang telah dikumpulkan melalui proses data cleaning bertujuan untuk menghapus data yang kurang lengkap dan melakukan seleksi ulang terhadap data.
 - b. Integrasi Data
Menggabungkan berbagai jenis data ke dalam satu kumpulan data.
 - c. Transformasi Data
Transformasi data dapat diartikan sebagai penyesuaian isi data, di mana format data yang berbeda, baik numerik maupun alfabet, harus diseragamkan untuk mempermudah klasifikasi data.
 - d. Pengurangan Data
Langkah terakhir adalah mengurangi data redundan tanpa mengubah data asli.
3. Pelatihan Data dan pengujian data
Data pelatihan merujuk pada data yang merupakan representasi akurat dari kondisi dan fakta yang ada sedangkan data pengujian adalah data yang dimanfaatkan untuk mengevaluasi seberapa efektif klasifikasi yang telah dilakukan (Baiq Nurul Azmi et al., 2023).
4. Algoritma K-NN
K-Nearst Neighbor, singkatan dari K-NN, adalah algoritma klasifikasi dalam pembelajaran mesin yang digunakan untuk menetapkan kategori pada objek atau entitas baru. Algoritma K-NN menghitung jarak antara objek baru dan

data yang ada untuk menentukan nilai K yang tepat (Fuansah et al., 2023). Formula yang digunakan untuk menghitung sejauh mana objek baru mendekati data yang ada adalah:

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

Keterangan:

- T = kasus baru
- S = kasus yang ada dalam penyimpanan
- n = jumlah atribut pada masing-masing kasus
- i = atribut individu antara 1 sampai n
- f = Fungsi kesamaan atribut I antara kasus T dan kasus S
- w = Bobot yang diberikan pada atribut ke-i

Pendekatan K-NN digunakan dalam menentukan nilai parameter K. jarak antara data baru dan seluruh data dalam set pelatihan dihitung. Selanjutnya, perbandingan dilakukan dengan data uji, dan confusion matrix dihasilkan (situramorang al.,2023). Penilaian dalam studi ini menggunakan confusion matrix. confusion Matrix berfungsi hntuk mendapatkan hasil dari nilai presisi, recal, dan akurasi. nilai dari Confusion Matrix biasanya di sajikan dala persentase (%) (rehayu et al.,2021).

5. Pengujian

Proses evaluasi untuk memperoleh nilai akurasi dilakukan dengan menggunakan bahasa pemograman python. Python adalah Bahasa pemograman yang memiliki tingkat tinggi yang dikembangkan oleh Guido Van Rossum dan diperkenalkan pada tahun 1991. Saat ini, python menjadi bahasa yang sangat populer. Selain itu, python juga merupakan Bahasa pemograman yang memiliki banyak kegunaan dalam bidang machine learning dan dep learning (Alfarizi et al,2023). Machine learning adalah suatu metode analisis terhadap masalah yang menggunakan model matematika untuk menangkap fitur-fitur yang menjadi kunci saat mesin beroperasi, dan kemudian model matematika ini memberikan solusi untuk suatu permasalahan (Mulia et al.,2023). Sementara itu Deep learning adalah suatu pendekatan pembelajaran yang memanfaatkan trasformasi nonlinier yang beragam, serta Deep dapat diintegrasikan dengan machine Learning menggunakan AI (Articial neural network).

HASIL

1. Pengumpulan Informasi

Kumpulan data diambil dari berbagai sumber yang tersimpan di <https://www.kaggle.com/>. Data tentang kualitas apel diperoleh dari <https://www.kaggle.com/> dengan total 4000 data yang telah dikategorikan sesuai dengan 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	1.697986	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 × 9 columns

Gambar 2. Dataset

2. Persiapan Data

Persiapan data menggunakan bahasa pemrograman Python, di mana langkah-langkah ini dilakukan untuk membersihkan data agar tidak ada yang tidak konsisten, serta menyamakan isi data agar lebih mudah dalam proses pengelompokan informasi.

```
[ ] 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]]
```

Gambar 3. Proses Persiapan Data

Ada 9 kriteria dari data apel yang diambil dari <https://www.kaggle.com>, yang digunakan untuk menilai kualitas apel. Semua kriteria ini akan diproses sesuai dengan masing-masing aspek. Aspek yang dilihat meliputi bentuk, ukuran, tingkat kemanisan, kadar air, keasaman, hingga hasil akhir yang menunjukkan kualitas apel itu sendiri.

3. Pelatihan Data dan Pengujian Data

Dari 4000 data yang diperoleh dari <https://www.kaggle.com>, dipisahkan menjadi dua bagian, yaitu data pelatihan dan data pengujian yang akan diorganisir menjadi data yang terstruktur. Dari 4000 data, 25% diambil untuk data pengujian dan sisanya digunakan untuk melakukan pengujian pada data pelatihan.

```
[ ] 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)
```

Gambar 4. Pelatihan Data dan Pengujian Data

```
[ ] len(y_train)
3000

[ ] len(y)
4000

[ ] len(y_test)
1000
```

Gambar 5. Hasil Pembagian Pelatihan Data dan Pengujian Data.

4. Algoritma K-NN

Setelah menyelesaikan proses pembersihan data dan membagi data menjadi data pelatihan dan data uji dari 4000 data, terlihat pada gambar 5 bahwa data uji terdiri dari 1000 data, sedangkan data pelatihan memiliki 3000 data. Dalam machine learning menggunakan bahasa pemrograman Python, kita memanggil fungsi K-NN agar dapat menghitung akurasi metode yang digunakan, dengan mendasarkan pada nilai kedekatan minkowski yaitu 2.

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

KNeighborsClassifier
KNeighborsClassifier()
```

Gambar 6. Proses Algoritma K-NN

5. Hasil Uji

Setelah melakukan prediksi atau klasifikasi dengan metode K-NN, tahap berikutnya adalah evaluasi dan validasi untuk mengukur seberapa tepat kualitas data buah apel menggunakan algoritma K-NN. Langkah pertama adalah membuat matriks kebingungan dari data pengujian yang telah diperoleh.

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

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

Gambar 7. Hasil Confusion Matriks

Ada 441 data yang diprediksi dengan benar, sedangkan 440 data diprediksi salah. Untuk 54 dan 65 data, prediksinya tidak terlalu tepat dari 1000 data yang diuji. Selanjutnya, kita akan mengecek laporan klasifikasi untuk melihat akurasi secara lebih rinci.

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

Gambar 8. Mencari Nilai Accuracy

Dari hasil akurasi yang disebutkan, terlihat perbandingan antara data precision, recall, dan f1-score, dengan akurasi yang mencapai 0.88% atau 88%.

KESIMPULAN

Hasil akurasi yang dilakukan menggunakan algoritma K-NN dengan 1000 data uji pada kualitas buah apel menunjukkan nilai sebesar 0,88 atau 88%. Ini menunjukkan bahwa klasifikasi dengan menggunakan 1000 data menghasilkan akurasi yang cukup memuaskan. Pada penelitian berikutnya, data yang sama bisa diterapkan dengan metode SVM atau Naïve Bayes yang dioptimalkan menggunakan algoritma genetik. Buah sangat penting bagi kesehatan manusia, dan apel adalah salah satu buah yang sangat disukai oleh masyarakat Indonesia. Oleh karena itu, penyediaan apel berkualitas baik sangat diperlukan agar memberikan manfaat bagi kesehatan. Dalam penelitian ini, digunakan metode K-NN yang dianggap efisien untuk melatih data dengan cepat pada data pelatihan dan pengujian yang berjumlah besar. Penelitian dimulai dengan pengumpulan dataset dari <https://www.kaggle.com/>, dilanjutkan dengan proses pra-pemrosesan, kemudian memisahkan data menjadi data pelatihan dan pengujian dengan proporsi 25% untuk pengujian dan 75% untuk pelatihan. Setelah itu, metode K-NN diterapkan untuk mengklasifikasikan data berdasarkan kriteria yang sudah ditentukan, menghasilkan evaluasi kinerja K-NN yang nilainya dihitung menggunakan pemrograman Python. Selama implementasi data mining dalam Bahasa pemrograman Python, digunakan library yang tersedia untuk mempermudah proses penerapan machine learning. Dari analisis confusion matrix, terdapat 441 data yang diprediksi benar, dan 440 data diprediksi salah. Selain itu, terdapat 54 dan 65 data yang diperkirakan tidak tepat dari 1000 data pengujian. Sehingga, nilai akurasi dengan metode K-NN diperoleh sebesar 0.88 atau 88%. Ini menunjukkan bahwa metode K-NN mampu bekerja dengan baik, cepat, dan efisien dalam melatih jumlah data yang besar.

DAFTAR PUSTAKA

Afriansyah, M., Saputra, J., Sa’adati, Y., & Valian Yoga Pudya Ardhana. (2023). Improving the Naive Bayes Algorithm for Classifying Apples by RGB Color Features. *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). Utilizing Python as a Programming Language for Machine Learning and Deep Learning. *Scientific Work of Students Bertauhid (KARIMAH TAUHID)*, 2(1), 1–6.

Azizah, N., Riyad Firdaus, M., Suyaningsih, R., & Indrayatna, F. (2023). Implementing the K-Nearest Neighbor Classification Algorithm for Diabetes Disease. *Proceedings of National Seminar on Actuarial Statistics*, 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). Analyzing the Impact of Training and Testing Data Composition on PCA and Decision Tree Algorithms for Liver Disease Classification. *JTIM: Journal of Information Technology and 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.stmikindonesia.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). Implementing the K-Nearest Neighbor Method to Classify Economic Status of Assistance Recipients. JUPITER, 15(2), 929–938.
- M. Afriansyah, Joni Saputra, Ardhana, V. Y. P., & Yuan Sa'adati. (2024). An Efficient Naive Bayes Algorithm for Classifying King Banana Based on Color Features. 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). Detecting the Ripeness Level of Areca Nut Using Support Vector Machine Based on Color and Texture. 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). Implementing Machine Learning for Identifying Individuals Coughing/Sneezing. Journal of Energy and Electrical Engineering (Jeee), 81(2), 81–86.
- Mustakim. (2022). Four Sources of Datasets for Learning and Research in Data Mining. <https://Mustakim.Irpi.or.Id/2022/05/18/Empat-Sumber-Dataset-Untuk-Belajar-Dan-Penelitian-BidangData-Mining/>.
- Nugraha, D., & Astuti, P. (2023). Sentiment Analysis of Cyberbullying on Instagram Using Support Vector Machine Method. 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). Comparison of K-Means and Naive Bayes Algorithms for Predicting Hospital Bill Payment Priorities Based on Importance Levels at PT. Pertamina (Persero). Journal of Informatics Engineering, 13(2), 1–8.
- Raup, A., Ridwan, W., Khoeriyah, Y., Supiana, S., & Zaqiah, Q. Y. (2022). Deep Learning and Its Applications in JJIP - Journal of Scientific Educational Studies, 5(9), 3258–3267. <https://doi.org/10.54371/jiip.v5i9.805>
- Reswan, Y., Toyib, R., Witriyono, H., & Anggraini, A. (2024). Classifying the Ripeness Level of Pineapple Based on Color Features Using the K-Nearest Neighbor (KNN) Method. 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). Evaluating the Quality of Malang Apples Based on Their Skin Using Digital Image Processing. Journal of 1945 University 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). K-Nearest Neighbor (K-NN) and Naive Bayes Algorithm Model. 7(1), 250–254.
- Suryanti, C., Rohman, Mg., Kunci -Apple Fruit, K., & Knn, K. (2024). Classifying the Quality of Apples by Color and Shape Using the KNN Method. Generation Journal, 8(1), 2580–4952.