

ANALISIS PERBANDINGAN ALGORITMA DECISION TREE DAN RANDOM FOREST TERHADAP KLASIFIKASI KUALITAS BUAH APEL

Dimas Nur Hazikin¹, Hardian Oktavianto², Dewi Lusiana Pater³, Daryanto⁴, Qurotta A'yun⁵

^{1,2,3,4,5}STIKOM Tunas Bangsa
Email: Dimasnh.83@gmail.com

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak

Apel merupakan salah satu komoditas unggulan di Indonesia yang kualitasnya perlu diklasifikasikan secara tepat untuk menjaga nilai jual. Proses klasifikasi manual masih sering digunakan, namun memiliki kelemahan dari sisi waktu, biaya, dan akurasi. Penelitian ini bertujuan membandingkan kinerja algoritma Decision Tree dan Random Forest dalam klasifikasi kualitas buah apel. Data penelitian diperoleh dari dataset Kaggle dengan atribut ukuran, berat, tingkat kemanisan, kerenyahan, kelembapan, kematangan, dan keasaman. Metode penelitian meliputi preprocessing, pembagian data, pelatihan model, serta evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan random forest mendapatkan tingkat akurasi 90% dan tingkat akurasi Decision Tree 81% maka dapat ditarik kesimpulan bahwa algoritma Random Forest merupakan metode yang lebih unggul dibandingkan dengan Decision Tree. Selain itu, dalam penelitian ini juga dilakukan pengujian dengan beberapa rasio pembagian data latih (training) dan data uji (testing), yaitu (90/10, 80/20, 70/30, 60/40, 50/50). Dari hasil pengujian tersebut, diperoleh bahwa rasio 80/20 menghasilkan tingkat akurasi yang paling tinggi dibandingkan rasio lainnya.

Kata kunci: *klasifikasi, apel, decision tree, random forest*

Comparative Analysis of Decision Tree and Random Forest Algorithms for Apple Quality Classification

Abstract

Apples are one of Indonesia's leading agricultural commodities whose quality needs to be accurately classified to maintain their market value. Manual classification processes are still commonly used but have several drawbacks in terms of time, cost, and accuracy. This study aims to compare the performance of the Decision Tree and Random Forest algorithms in classifying apple quality. The research data were obtained from a Kaggle dataset with attributes including size, weight, sweetness level, crispness, moisture, ripeness, and acidity. The research method includes preprocessing, data splitting, model training, and evaluation using accuracy, precision, recall, and F1-score metrics. The results of the study show that the Random Forest algorithm achieved an accuracy of 90%, while the Decision Tree algorithm achieved 81%. Therefore, it can be concluded that the Random Forest algorithm performs better than the Decision Tree algorithm in classifying apple quality. In addition, this study also tested several ratios of training and testing data splits (90/10, 80/20, 70/30, 60/40, and 50/50). Based on the test results, the 80/20 ratio produced the highest accuracy compared to the other ratios.

Keywords: *classification, apple, decision tree, random forest*

1. PENDAHULUAN

Apel (*Malus domestica*) adalah buah yang memiliki tingkat popularitas tinggi di seluruh dunia (NUPUS, 2022). Di banyak negara, apel menjadi pilihan utama konsumen karena rasanya yang manis, asam, dan segar, serta kandungan gizi yang beragam. Buah ini mengandung vitamin C, serat, kalium, dan berbagai senyawa antioksidan yang bermanfaat untuk meningkatkan sistem imun tubuh dan mencegah berbagai penyakit degeneratif seperti penyakit

jantung, diabetes tipe 2, dan kanker. Karena itu, apel sering dianggap sebagai simbol makanan sehat, yang menjadikannya buah yang banyak dibudidayakan, termasuk di Indonesia (Pamungkasih, Ristanti, Ramayanti, & Arini, 2023).

Wilayah terbesar penghasil buah apel di Indonesia adalah Jawa Timur, dengan jumlah produksi mencapai 480.834 ton pada tahun 2019 dan terus meningkat ditahun-tahun selanjutnya (ISLAMIAH, 2024).

Kualitas buah apel sering dikategorikan berdasarkan beberapa parameter fisik yang jelas terlihat oleh produsen atau pedagang. Beberapa faktor yang dipertimbangkan dalam penilaian kualitas apel meliputi ukuran, warna, bentuk, tekstur kulit, kekerasan buah, serta adanya cacat pada kulit buah akibat kerusakan alami atau kerusakan mekanis selama transportasi. Apel berkualitas baik biasanya memiliki ukuran yang seragam, kulit yang cerah dan merata, serta bentuk yang utuh tanpa cacat. Buah apel yang bebas dari kerusakan atau penyakit serta memiliki rasa yang enak umumnya dihargai lebih tinggi (Brahmantyo, Mukson, & Roessali, 2022).

Proses pengklasifikasian kualitas apel biasanya dilakukan secara manual, di mana pekerja memeriksa setiap buah apel secara visual untuk menentukan apakah buah tersebut memenuhi standar kualitas tertentu. Kelemahan utama adalah ketergantungan pada tenaga manusia yang banyak dan memakan waktu. Seiring dengan meningkatnya permintaan pasar dan ekspektasi konsumen yang semakin tinggi, diperlukan sistem otomatis yang dapat mengklasifikasikan kualitas apel dengan lebih cepat, akurat, dan efisien. Kemajuan dalam teknologi kecerdasan buatan (AI) dan pembelajaran mesin (*machine learning*) dapat menjadi solusi untuk mengatasi masalah yang timbul dari sistem klasifikasi manual (Agustyaningrum, Gata, Nurfalah, Radiyah, & Maulidah, 2020).

Dengan menggunakan algoritma machine learning, kualitas apel dapat diprediksi secara otomatis dengan menganalisis berbagai fitur fisik yang dimilikinya. Salah satu algoritma pembelajaran mesin yang populer dalam klasifikasi adalah *Decision Tree*. Algoritma ini bekerja dengan membagi data menjadi subset lebih kecil berdasarkan atribut yang dianggap penting untuk memisahkan data ke dalam kelas yang berbeda. Untuk mengatasi masalah *overfitting* yang sering terjadi pada *Decision Tree*, metode *ensemble learning* seperti *Random Forest* diperkenalkan. *Random Forest* menggabungkan banyak pohon keputusan untuk menghasilkan model yang lebih kuat dan stabil.

Meskipun keduanya memiliki tujuan yang sama dalam mengklasifikasi data, *Decision Tree* dan *Random Forest* memiliki perbedaan yang signifikan. *Decision Tree* lebih sederhana, cepat dilatih, dan mudah dipahami, tetapi rentan terhadap *overfitting*. Sementara itu, *Random Forest* lebih tahan terhadap *overfitting* dan lebih stabil, namun memerlukan lebih banyak waktu dan sumber daya komputasi (Novianti, Mandati, & Andana, 2023).

Oleh karena itu, penting untuk melakukan analisis perbandingan antara kedua algoritma ini dalam konteks klasifikasi kualitas buah apel, guna menentukan algoritma mana yang lebih efektif dan efisien dalam menghasilkan prediksi yang akurat (kanugrahan, putra, & ramadhani, 2024).

Penelitian ini akan mengevaluasi dan membandingkan kinerja algoritma *Decision Tree* dan

Random Forest dalam klasifikasi kualitas buah apel, serta menilai kedua algoritma berdasarkan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Penelitian ini diharapkan dapat memberikan kontribusi signifikan bagi industri pertanian, terutama dalam meningkatkan efisiensi dan akurasi proses klasifikasi kualitas buah apel.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian yaitu klasifikasi dengan algoritma *Random Forest* dan *Decision Tree*. Pendekatan ini dipilih karena penelitian bertujuan untuk membandingkan kinerja dua algoritma klasifikasi, yaitu *Decision Tree* dan *Random Forest*, dalam mengklasifikasikan kualitas buah apel.

Data yang digunakan dalam penelitian ini adalah dataset kualitas buah apel yang diperoleh dari sumber Kaggle

<https://www.kaggle.com/datasets/nelgiriwithana/apple-quality>. Dataset ini berisi atribut-atribut sebagai berikut. Ukuran (Size) Ukuran buah apel, Berat (Weight) Berat buah apel. Tingkat Kemanisan (Sweetness) Tingkat rasa manis pada buah apel, Kekerenyahan (Crunchiness) Tekstur buah yang menunjukkan seberapa renyah apel tersebut, Kelembapan (Juiciness) Tingkat kandungan air atau jus dalam buah apel, Kematangan (Ripeness) Tahap kematangan dari buah apel, Keasaman (Acidity) Tingkat keasaman dari buah apel, Kualitas (Quality) Label kualitas buah apel (baik, dan buruk).

melakukan pra-pemrosesan data sebelum proses analisis. Tahap ini mencakup pembersihan data dengan menghapus nilai hilang, data duplikat, serta outlier yang tidak relevan. Peneliti kemudian melakukan transformasi data dengan menyesuaikan tipe data. Proses pembagian dilakukan menggunakan teknik stratified sampling agar proporsi kelas tetap seimbang. Pada penelitian ini, peneliti menggunakan rasio pembagian umum, yaitu 80/20 peneliti mengevaluasi kinerja model dengan menggunakan data uji. Data uji digunakan untuk menghasilkan prediksi kelas, kemudian hasil prediksi dibandingkan dengan label aktual. Peneliti menghitung metrik evaluasi, meliputi akurasi, presisi, *recall*, dan *F1-score*, guna menilai sejauh mana model mampu melakukan klasifikasi dengan baik.

Confusion Matrix Merupakan salah satu metode evaluasi kinerja model klasifikasi yang penting, terutama untuk memahami secara rinci kesalahan klasifikasi dalam konteks data label biner. Table 1. Confusion matrix

Tabel 1. Confusion Matrix

Aktual/ Prediksi	Baik (Prediksi)	Buruk (Prediksi)
Baik (Aktual)	True Positive (TP)	False Negative (FN)
Buruk (Aktual)	False Positive (FP)	True Negative (TN)

Akurasi merupakan alat mengukur seberapa baik sebuah model klasifikasi dalam memprediksi data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Nilai Presisi merupakan rasiojumlah data kategori positif.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall merupakan nilai untuk menampilkan proporsi data kelas positif yang berhasil di klasifikasi oleh algoritma.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1- Score merupakan gabungan dari presisi dan recall agar lebih seimbang.

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (4)$$

hasil evaluasi dari algoritma Decision Tree dan Random Forest untuk menentukan algoritma yang lebih baik dalam mengklasifikasikan kualitas buah apel. Perbandingan dilakukan dengan menyajikan metrik evaluasi dari kedua model dalam bentuk tabel maupun grafik agar hasilnya lebih mudah dianalisis.

3. HASIL DAN PEMBAHASAN

3.1 Pra-Pemrosesan Data

Tahap ini mencakup pembersihan data dengan menghapus nilai hilang, data duplikat, serta outlier yang tidak relevan.

1. Pada data yang digunakan tidak terdapat data yang duplikat.
2. Terdapat missing value di beberapa kolom.

```
A_id      1
Size      1
Weight    1
Sweetness 1
Crunchiness 1
Juiciness 1
Ripeness  1
Acidity   0
Quality   1
dtype: int64
```

Gambar 1. Cek missing value

Kemudian menghapus missing value

```
A_id Size Weight Sweetness Crunchiness Juiciness Ripeness Acidity Quality
4000 NaN NaN NaN NaN NaN NaN NaN Created_by_Nidula_Elginriyewithana NaN
```

Gambar 2. Penghapusan missing value

3. Mendeteksi outlier pada setiap kolom numerik dari dataframe, pada data terdapat data outlier, menggunakan metode interquartile range IQR.

```
Size      22
Weight    54
Sweetness 32
Crunchiness 47
Juiciness 32
Ripeness  24
Acidity   20
Quality   0
dtype: int64
```

Gambar 3. Jumlah outlier per kolom

outlier diseluruh baris dan juga presentasinya,

mencari presentasen menggunakan rumus

Presentase outlier = (Total outlier : Total data) x 100

Total seluruh outlier sebanyak 231 dan hasil presentasinya adalah 0,72%. Berdasarkan hasil perhitungan, persentase outlier dalam data ini tergolong kecil. Oleh karena itu, peneliti memutuskan untuk tidak menghapus data outlier tersebut. Selain jumlahnya yang minim, outlier juga dapat mengandung informasi penting yang merepresentasikan kondisi ekstrem atau kasus khusus yang tetap relevan dengan analisis. Dengan demikian, penghapusan outlier secara sembarangan tidak dilakukan demi menjaga keutuhan dan validitas data.

3.2 Pembagian Data

dataset telah dibersihkan dan selanjutnya dibagi menjadi data latih (training data) dan data uji (testing data). Tujuan dari pembagian ini untuk melatih model dengan sebagian data dan menguji kinerjanya. Pembagian data dilakukan dengan rasio train test split 80/20, 90/10, 70/30, 60/40, 50/50. Data dengan 80/20

diperoleh nilai performa paling baik (Putri, 2021). Dimana 80% data digunakan untuk melatih model dan 20% untuk menguji performa model.

3.3 Hasil Perbandingan

Dari hasil pengujian yang sudah dilakukan evaluasi algoritma Decision Tree dan Random Forest. Hasil dari algoritma random forest yang memiliki nilai tertinggi dibandingkan dengan algoritma Decision Tree. Berikut tabel Perbandingannya dapat dilihat pada tabel dibawah ini :

Tabel 2. Hasil Perbandingan

Rasio	Model	Aku rasi	Presisi	Recall	F1- Score
80/20	Decision Tree	0,81	0,81	0,81	0,81
	Random forest	0,90	0,90	0,90	0,90
90/10	Decision Tree	0,78	0,79	0,80	0,79
	Random forest	0,90	0,92	0,89	0,90
70/30	Decision Tree	0,79	0,78	0,79	0,79
	Random forest	0,89	0,89	0,89	0,89
60/40	Decision Tree	0,79	0,79	0,77	0,79
	Random forest	0,88	0,88	0,88	0,88
50/50	Decision Tree	0,81	0,80	0,81	0,81
	Random forest	0,88	0,88	0,88	0,88

Dari tabel diatas dapat dilihat perbandingannya, model yang termasuk Classification terbaik yang mana dari ke dua algoritma ini menggunakan pembagian data 80/20 mendapatkan hasil yang baik. Algoritma Random Forest menghasilkan performa yang bagus yaitu 0,90, dibandingkan dengan Decision tree yang memiliki nilai 0,81, namun nilai tersebut juga dapat dikatakan cukup baik. Dari Kinerja ke dua algoritma ini bisa dikatakan algoritma yang tahan dengan Outlier.

4. KESIMPULAN

random forest mendapatkan tingkat akurasi 90% dan tingkat akurasi Decision Tree 81% maka dapat ditarik kesimpulan bahwa algoritma Random Forest merupakan metode yang lebih unggul dibandingkan dengan Decision Tree. dalam penelitian ini juga dilakukan pengujian dengan beberapa rasio pembagian data latih (training) dan data uji (testing), yaitu 90/10, 80/20, 70/30, 60/40, 50/50. Dari hasil pengujian tersebut, diperoleh bahwa rasio 80/20 menghasilkan tingkat akurasi yang paling tinggi dibandingkan rasio lainnya.

5. DAFTAR PUSTAKA

Semua naskah ditelaah secara *double blind-review* oleh mitra bestari (*reviewers*) yang ditunjuk oleh *editor* menurut bidang kepakarannya. Penulis naskah diberi kesempatan untuk melakukan perbaikan (*revisi*) naskah atas dasar rekomendasi/saran dari mitra bestari dan editor. Kepastian pemuatan atau penolakan naskah akan diberitahukan secara tertulis melalui email.

Pemeriksaan dan penyuntingan cetak-coba dikerjakan oleh redaksi dan/atau dengan melibatkan penulis. Naskah yang sudah dalam bentuk cetak-coba dapat dibatalkan pemuatannya oleh redaksi jika diketahui bermasalah.

Segala sesuatu yang menyangkut perijinan pengutipan atau penggunaan software komputer untuk pembuatan naskah atau hal lain yang terkait dengan HAKI yang dilakukan oleh penulis naskah, berikut konsekuensi hukum yang mungkin timbul karenanya, menjadi tanggung jawab penuh penulis naskah tersebut.

6. DAFTAR PUSTAKA

- Agustyaningrum, C. I., Gata, W., Nurfalaha, R., Radiah, U., & Maulidah, M. (2020). KOMPARASI ALGORITMA NAIVE BAYES, RANDOM FOREST DAN SVM UNTUK MEMPREDIKSI NIAT PEMBELANJA ONLINE. *Jurnal Informatika*, 20, 164-173.
- Brahmantyo, I., Mukson, & Roessali, W. (2022). ANALISIS FAKTOR - FAKTOR YANG MEMPENGARUHI KEPUTUSAN KONSUMEN DALAM MEMBELI BUAHAPELIMPOR(STUDI KASUS PASAR MODERN TRANSMART KOTA SEMARANG). *AGROLAND: JURNAL ILMU-ILMUPERTANIAN*, 29, 230-240.

- ISLAMIAH, U. N. (2024). *ANALISIS NILAI TAMBAH DAN STRATEGI PENGEMBANGAN AGROINDUSTRI TENG-TENG APEL (STUDI KASUS HILWA NUSANTARA KOTA BATU)*. UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR SURABAYA .
- kanugrahan, G., putra, v. h., & ramadhani, y. (2024). Analisis Sentimen Aplikasi Gojek Menggunakan SVM, Random Forest dan Decision Tree. *Jurnal infortech*, 6, 171-178.
- Novianti, T., Mandati, S. A., & Andana, E. K. (2023). Peningkatan Evaluasi Risiko Kredit Menggunakan Decision Tree C4.5. *Journal of Maneuvacturing in Industrial Engineering and Technology(MINE-TECH)*, 2, 1-9.
- NUPUS, Y. Y. (2022). *KARAKTERISTIK BUAH DAN NILAI NUTRISI KULTIVAR PISANG BILE (Musa acuminata x Musa balbisiana (ABB) cv. Bile) DI KABUPATEN LOMBOK TIMUR*. UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG .
- Pamungkasih, E., Ristanti, R. F., Ramayanti, K., & Arini, I. Y. (2023). Strategi Pengembangan Komoditas Buah Apel di Kabupaten Malang. *Prosiding Seminar Nasional Pembangunan dan Pendidikan Vokasi Pertanian*, 4, 105-113.
- Putri, n. P. (2021). Analisis Performa Metode K-Nearest Neighbor(KNN) dan Crossvalidationpada Data Penyakit Cardiovascular. *Indonesian Journal of Data and Science*, 2, 21-28.