KLASIFIKASI KUALITAS BUAH APEL MENGGUNAKAN METODE RANDOM FOREST

Andicho Putra Argadianata, Doni Abdul Fatah, Hanifudin Sukri
Sistem Informasi, Universitas Trunojoyo Madura
Jl. Raya Telang, Kecamatan Kamal, Bangkalan, Jawa Timur, Indonesia
Andikoputra766@gmail.com

ABSTRAK

Di Indonesia perkebunan merupakan salah satu sektor perekonomian yang penting terutama untuk kebun apel, selain rasanya yang cukup enak serta penyebarannya cukup baik apel juga cukup bermanfaat untuk tubuh, namun tidak semua apel layak di konsumsi, walaupun terlihat baik banyak apel yang kuang baik dalam penggolongan nya, Penelitian ini bertujuan mengklasifikasikan kualitas apel berdasarkan fitur fisik seperti Size, Weight, Sweetness, Crunchiness, Juiciness, Ripeness, dan Acidity, dengan Quality sebagai target klasifikasi. Data yang digunakan berasal dari Kaggle, terdiri dari 4000 data. Model klasifikasi yang digunakan adalah Random Forest karena kemampuannya yang baik dalam menangani data dengan banyak fitur. Hasil evaluasi menunjukkan bahwa model mencapai akurasi 88,5%, precision 88,1%, recall 89,0%, dan F1-Score 88,5%. Berdasarkan confusion matrix, terdapat 48 data Bad yang salah diprediksi sebagai Good (False Positive) dan 44 data Good yang salah diprediksi sebagai Bad (False Negative). Penelitian ini membuktikan bahwa Random Forest adalah metode yang andal untuk klasifikasi kualitas apel berdasarkan fitur fisiknya.

Kata kunci : *Apel, Machine learning, Klasifikasi, Random forest, confusion Matrix*

1. PENDAHULUAN

Di Indonesia perkebunan merupakan salah satu sektor perekonomian yang penting. Perkebunan memberikan arti yang penting dalam pembangunan serta pertumbuhan ekonomi Masyarakat [1].

Perkebunan apel merupakan salah satu sektor pertanian yang memiliki nilai ekonomi tinggi dan berkontribusi besar terhadap sektor agribisnis di banyak negara. Apel (Malus domestica) dikenal sebagai buah yang kaya nutrisi dan memiliki beragam manfaat kesehatan, sehingga menjadi salah satu komoditas hortikultura yang banyak diminati di pasar lokal maupun internasional. Apel dapat tumbuh dan berbuah baik di daerah dataran tinggi yang memiliki temperatur rendah. Tanaman apel menghendaki lingkungan dengan karakteristik yaitu temperatur rendah, kelembaban Udara rendah dan curah hujan tidak terlalu tinggi (Soelarso, 1996).

Kualitas apel sangat ditentukan oleh berbagai faktor, mulai dari pemilihan varietas, teknik budidaya, hingga penanganan pascapanen. Varietas apel seperti Fuji, Granny Smith, dan Rome Beauty memiliki karakteristik unik yang memengaruhi rasa, tekstur, dan daya tahan buah. Penilaian kualitas apel biasanya mencakup aspek-aspek seperti ukuran buah, warna kulit, rasa, aroma, dan tekstur daging buah. Apel berkualitas tinggi cenderung memiliki warna kulit yang menarik, tekstur renyah, serta rasa yang seimbang antara manis dan asam. Kualitas ini tidak hanya memengaruhi preferensi konsumen, tetapi juga menentukan nilai jual apel di pasar.

Klasifikasi adalah salah satu teknik dalam machine learning yang digunakan untuk mengategorikan data ke dalam kelas-kelas tertentu berdasarkan atribut yang dimiliki. Dalam konteks pertanian, klasifikasi dapat digunakan untuk menentukan kualitas buah berdasarkan karakteristik yang terlihat, seperti ukuran, warna, tekstur, dan tingkat kematangan. Salah satu metode yang sering digunakan untuk tugas klasifikasi adalah *Random Forest*.

Random Forest adalah algoritma pembelajaran mesin yang berbasis pada pohon keputusan (decision tree) dan termasuk dalam kategori metode ensemble. Algoritma ini bekerja dengan membangun sejumlah besar pohon keputusan yang terlatih pada subset data yang berbeda, kemudian menggabungkan hasil klasifikasi dari pohon-pohon tersebut untuk menghasilkan keputusan yang lebih akurat dan robust. Keuntungan utama dari *Random Forest* adalah kemampuannya dalam menangani data yang kompleks dan tidak memerlukan pemrosesan data yang rumit.

Dalam aplikasi klasifikasi kualitas buah apel, *Random Forest* dapat digunakan untuk mengklasifikasikan buah apel ke dalam kategori kualitas tertentu, seperti "baik", "sedang", atau "buruk", berdasarkan data karakteristik seperti warna kulit, ukuran, tekstur, dan parameter lainnya. Melalui pendekatan ini, kualitas buah apel dapat dievaluasi secara otomatis, memberikan manfaat bagi produsen, pengepul, dan konsumen dalam memilih buah yang memiliki kualitas terbaik.

Pendekatan ini sangat relevan untuk industri pertanian, di mana otomatisasi dan efisiensi sangat diutamakan dalam proses pengelolaan hasil pertanian. Dengan menggunakan *Random Forest*, proses klasifikasi kualitas buah apel dapat dilakukan dengan lebih cepat dan akurat, sekaligus mengurangi subyektivitas dalam penilaian kualitas.

Tujuan dari penelitian ini untuk mengetahui kualitas dari buah apel dengan menggunakan metode random forest.

2. TINJAUAN PUSTAKA

2.1. Perkebunan (Second-Level Heading)

Perkebunan merupakan salah satu sektor yang penting Di Indonesia. Perkebunan ialah kegiatan yang berhubungan dengan tanaman tertentu pada tanah atau media tumbuh lainnya di dalam area yang sesuai, mengelola dan memasarkan barang dan jasa hasil tanaman tersebut dengan bantuan ilmu teknologi serta pengetahuan, pemodalan serta pengelolaan untuk mewujudkan hasil yang baik bagi pelaku usaha perkebunan serta masyarakat [1]. Salah satu produk perkebunan buah yang popular di Indonesia adalah buah apel.

Buah apel disukai karena menggandung banyak serat yang baik bagi kesehatan pencernaan dan vitamin yang diperlukan untuk menjaga kesehatan tubuh. Kebutuhan masyarakat terhadap buah apel dapat dilihat dari tingkat konsumsi nasional sebesar 1,04 kg per kapita per tahun pada tahun 2016 [2]. Buah apel disukai karena dapat digunakan menjadi berbagai olahan seperti jus, kripik, selai yang dapat diperjualbelikan untuk meningkatkan ekonomi.

2.2. *Machine Learning*

Machine learning merupakan salah satu teknologi *artificial intelligence* atau kecerdasan buatan yang deprogram untuk mengikuti kecerdasan manusia. *Machine learning* digunakan untuk mengembangkan sistem yang dapat berjalan dan membuat keputusan sendiri tanpa harus di program ulang. *Machine learning* dapat berjalan dengan memasukan input data untuk dianalisa yang dapat menghasilkan pola tertentu dalam data besar atau *big data*. Di dalam *machine learning*, terdapat konsep *data training* dan *data testing*, dimana *data tra*ining digunakan untuk melatih algoritma *machine learning* dan *data testing* digunakan untuk menguji performa algoritma yang telah dilatih ketika menemukan data baru yang belum pernah diberikan dalam data training [3].

Pada *machine learning* dapat mendapatkan hasil yang akurat menggunakan teknik cerdas. Pada *machine learning* terdapat beberapa teknik nya seperti *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement learning*. Salah satu teknik *machine learning* adalah *unsupervised learning*, *unsupervised learning* merupakan teknik yang mengambil Kesimpulan dari dataset tanpa memiliki label respons terkait[4].

2.3. *Data Mining*

Data mining merupakan proses yang digunakan untuk mengekstraksi data mengumpulkan informasi dari sekumpulan data berukuran besar. Pada data mining terdapat beberapa proses seperti pemahaman data (*Data Understanding*), persiapan data (*Data Preparation*), pemodelan data (*Data Modeling*), Evaluasi (*Evaluation*), dan penerapan (*Deployment*)[5].

Pada *data mining* untuk pengambilan informasi dari basis data yang besar dengan cara menggabungkan pendekatan dari berbagai disiplin ilmu, termasuk pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi.

Data mining menggunakan teknik-teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin mendapatkan informasi dari berbagai basis data besar. Proses data mining melibatkan beberapa teknik untuk menemukan pola, hubungan, dan tren yang tersembunyi dalam data. Berdasarkan jenis tugasnya, data mining dapat dikategorikan menjadi lima jenis utama, yaitu Exploratory Data Analysis (EDA) atau Memahami karakteristik data secara mendalam., Descriptive juga disebut Mendeskripsikan data dengan menggunakan model statistik, Modelling atau Membangun model untuk memprediksi nilai atau kelas suatu variabel, Predictive Modelling, Penemuan Pola serta Aturan yang digunakan, dan juga Pemanggilan Konten [6]. Data mining juga digunakan untuk mengidentifikasi keterkaitan dan hubungan yang tidak terlihat antar variabel.

2.4. Klasifikasi

Klasifikasi merupakan suatu metode analisis yang bertujuan untuk mengelompokkan data ke dalam kategori-kategori tertentu berdasarkan sifat atau karakteristiknya[7].

Klasifikasi sering digunakan untuk membangun model prediktif untuk membantu pengambilan keputusan. Klasifikasi adalah proses pengelompokan data ke dalam kelas-kelas tertentu berdasarkan aturan yang telah ditentukan. Klasifikasi berbeda dengan clustering, clustering tidak memerlukan label pada data keluarannya, namun klasifikasi memerlukan data berlabel untuk proses pembelajaran nya. Pada klasifikasi terdapat beberapa algoritma yang dapat digunakan seperti *decision tree*, naïve bayes, *k-nearest neighbor* (KNN), dan support *vector machine* (SVM). Algoritma-algoritma ini belajar dari data untuk menemukan pola yang menghubungkan antara ciri-ciri data dengan kategori yang sesuai[8].

Pada klasifikasi terdapat beberapa tahapan seperti pembangunan model, penerapan model, dan evaluasi. Pembuatan model dilakukan dengan membangun model menggunakan data latih yang telah memiliki atribut dan kelas. Data baru kemudian diterapkan ke model untuk menentukan kelas nya. Pada tahap akhir, model dievaluasi untuk mengukur tingkat akurasi saat membangun dan menerapkan model pada data baru. Terdapat dua tahap pada proses klasifikasi yaitu tahap pelatihan dan tahap pengujian. Tahap pelatihan merupakan tahap di mana data digunakan untuk mengembangkan model, sedangkan tahap pengujian digunakan untuk menguji model yang telah dibuat dengan data yang berbeda untuk mengetahui sejauh mana tingkat akurasinya[9].

2.5. *Random Forest*

Random forest merupakan salah satu algoritma *machine learning* yang digunakan untuk klasifikasi

dengan mengembangkan metode decision tree yang melakukan pemilihan atribut secara acak pada setiap node. Random forest menggabungkan konsep bagging (Bootstrap Aggregation) dan pohon keputusan (decision tree) untuk menciptakan model yang kuat, stabil, dan akurat. Pada setiap *tree*, metode *bagging* digunakan dengan memberikan sampel data pelatihan [10].

Pada penggunaan Random Forest untuk melakukan pembagian data pada pohon keputusan menggunakan gini index. Gini index merupakan metrik yang digunakan untuk mengukur impurity atau ketidakmurnian suatu node dalam pohon Keputusan [11].

Bootstrap Sampling: Membuat beberapa subset dari data pelatihan dengan pengambilan sampel acak dengan pengembalian. Feature Randomness: Untuk setiap pohon, hanya sebagian fitur yang dipilih secara acak sebagai kandidat split di setiap node. Voting atau Averaging: Menggabungkan prediksi semua pohon dengan voting mayoritas (klasifikasi) atau rata-rata (regresi). Secara umum ada tiga tahap dalam random forest

a. Bootstrap Sampling: Membuat beberapa subset dari data pelatihan dengan pengambilan sampel acak dengan pengembalian.
b. Feature Randomness: Untuk setiap pohon, hanya sebagian fitur yang dipilih secara acak sebagai kandidat split di setiap node.
c. Voting atau Averaging: Menggabungkan prediksi semua pohon dengan voting mayoritas (klasifikasi) atau rata-rata (regresi).

Rumus dasar:

$$y = mode\{h_1(x), h_2(x), \dots, h_n(x)\} \quad (1)$$

di mana $hi(x)$ adalah prediksi dari pohon ke-i.

Regresi: Hasil akhir (y) adalah rata-rata prediksi:

$$y = \frac{1}{2}\sum_{i=1}^{n} h_1(x) \quad (2)$$

2.6. *Overfitting*

Overfitting adalah kondisi ketika model pembelajaran mesin terlalu "menghafal" data pelatihan sehingga berkinerja sangat baik pada data tersebut, tetapi buruk saat diterapkan pada data baru atau data uji. Model yang overfitting gagal menangkap pola umum dalam data karena terlalu fokus pada detail atau noise yang spesifik pada data pelatihan.

2.7. Evaluasi Model

Pada tahap evaluasi model, digunakan Confusion Matrix untuk menganalisis kinerja model klasifikasi. Confusion Matrix adalah tabel yang membandingkan prediksi model dengan data aktual, memberikan informasi tentang jumlah prediksi yang benar (True Positive dan True Negative) serta kesalahan prediksi (False Positive dan False Negative).

2.8. Jupyter Notebook

Jupyter notebook merupakan aplikasi yang digunakan untuk membuat dan berbagi dokumen yang berbasis web dan dapat digunakan secara gratis. Jupyter Notebook menjadi salah satu aplikasi yang paling populer dalam bidang ilmu data. Alat ini memungkinkan pengguna untuk mengembangkan kode Python secara interaktif, sehingga mereka dapat menampilkan output, visualisasi, dan teks dalam satu dokumen yang terintegrasi. dibuat di tahun 2014 oleh seorang bernama Fernando Perez, ini sangat dibutuhkan dalam pengembangan data karena fungsinya untuk menangani data besar, hubungan dengan database, dan berbagai sistem analisis data [12].

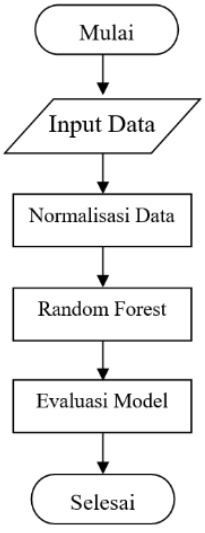
2.9. Penelitian Terdahulu

Havidz Yasaf Al-Afghoni meneliti tentang klasifikasi jenis kacang menggunakan Decision tree C4.5, SMOTE, Confusion Matrix dan dari penelitian itu menghasilkan tingkat akurasi sebesar 94% presisi sebesar 94%, recallnya 94%, dan f1- scorenya adalah 94%[13], pada tahun 2024 rif'at chusnul maafi telah melakukan penelitian tentang processing gambar menggunakan metode cnn, dengan menggunakan evaluasi model Confusion matrix, dan penelitian tersebut menghasilkan nilai tingkat akurasi sebesar 76.92%.[14], tahun 2023 kurniawan melakukan penelitian dengan Tujuan menemukan akurasi tertinggi di antara algoritma klasifikasi prediksi yang diusulkan penerima bantuan raskin memakai tools python machine learning dan di implementasikan melalui suatu website, menunjukkan algoritma klasifikasi RF memiliki hasil precision, recall, f-measure yaitu nilai 97%, nilai accuracy sebesar 97,26% dan nilai ROC 0,970[15].

3. METODE PENELITIAN

Pada penelitian ini menggunakan pemrosesan data dengan metode random forest, tapi sebelum itu data di normalisasi dan dipisahkan menjadi data training dan evaluasi model menggunakan confusion matrix.

3.1. Input Data



Gambar 1. flowchart Penelitian

Pada tahap ini data yang digunakan di inputkan ke dalam sistem untuk di proses perhitungan, data yang digunakan adalah data tentang apel dan gambarannya sebagai berikut:

Tabel 1. Data apel

Nama kolom	penjelasan
Size	Mengacu pada ukuran fisik apel, seperti diameter atau volume.
Weight	Berat apel dalam gram atau kilogram.
Sweetness	Mengukur kadar gula dalam apel.
Crunchiness	Mengukur tekstur apel saat digigit, yaitu seberapa renyah apel tersebut.
Juiciness	Mengukur seberapa banyak jus atau cairan yang dihasilkan apel.
Ripeness	Menunjukkan tingkat kematangan apel.
Acidity	Mengukur kadar asam dalam apel, seperti asam malat.
Quality	Label atau hasil akhir evaluasi kualitas apel, yang bisa berupa kategori seperti "good" (baik), "bad" (buruk).

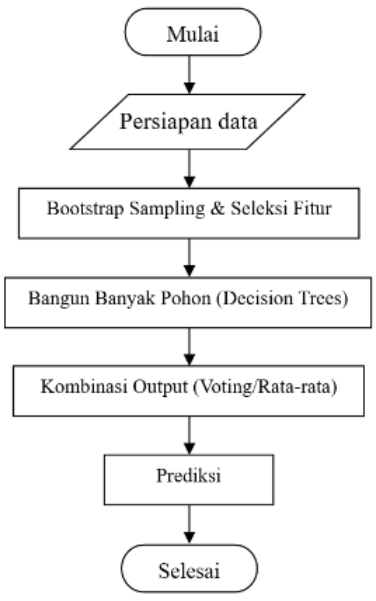
3.2. Normalisasi Data

Normalisasi data adalah proses mengubah nilai-nilai data menjadi skala tertentu, biasanya dalam rentang tertentu, seperti 0 hingga 1 atau -1 hingga 1. Normalisasi digunakan untuk memastikan bahwa semua fitur atau variabel memiliki skala yang sebanding, sehingga tidak ada fitur yang mendominasi perhitungan dalam model machine learning karena memiliki nilai yang jauh lebih besar daripada yang lain. Normalisasi dilakukan menggunakan MinMaxScalling, Rumusnya sebagai berikut :

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Dimana X adalah data yang di pilih, Xmin adalah data terkecil, dan Xmax data terbesar.

3.3. Random Forest



Gambar 2. Flowchart Random Forest

Random Forest adalah algoritma machine learning berbasis ensemble yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bekerja dengan membangun banyak decision tree secara acak selama pelatihan, lalu menggabungkan hasilnya (melalui voting untuk klasifikasi atau rata-rata untuk regresi) untuk meningkatkan akurasi dan mengurangi overfitting.

a. Persiapan Data
   Dari data yang telah di proses dan di persiapkan, data dilakukan pemisahan data menjadi data testing dan data training, dan dilakukan perhitungan klasifikasi.
b. Bootstrap sampling dan seleksi fitur
   Bootstrap sampling adalah teknik pengambilan sampel secara acak dengan pengembalian dari dataset asli untuk membentuk subset data latih. Pada setiap pembentukan node dalam pohon keputusan, subset fitur dipilih secara acak untuk menentukan split terbaik. Hal ini bertujuan untuk mengurangi korelasi antar pohon dan meningkatkan variasi model.
   Dengan menggunakan rumus

$$f = \frac{\sum_{t=1}^{T} \sum_{n \in N_t} I(f,n) \Delta impurity(n)}{\sum_{t=1}^{T} \sum_{n \in N_t} \Delta impurity(n)} \quad (4)$$

   Dimana
   f=feature important
   T=Jumlah total pohon dalam Random Forest.
   N_t=Node dalam pohon t
   $I(f,n)$=Indikator apakah fitur f digunakan untuk split di node n. Jika digunakan maka nilainya 1, jika tidak, maka nilainya 0
   ΔImpurity(n): Pengurangan impurity di node n.
c. Bangun Banyak Pohon (Decision Trees)
   Setiap subset data bootstrap digunakan untuk membangun pohon keputusan. Pohon ini berkembang hingga kriteria tertentu tercapai, seperti kedalaman maksimum atau semua node menjadi homogen. Setiap pohon dilatih secara independen, menciptakan hutan pohon (forest).
d. Kombinasi Output (Voting/Rata-rata)
   Hasil akhir dari Random Forest diperoleh dengan menggabungkan output dari semua pohon, dalam kasus ini peneliti menggunakan Klasifikasi, Menggunakan mayoritas suara (majority voting) dari semua pohon untuk menentukan label akhir.
e. Hasil Akhir: Evaluasi atau Prediksi
   Setelah model selesai dilatih, evaluasi dilakukan menggunakan metrik seperti akurasi, precision, recall, atau confusion matrix. Untuk data baru, model digunakan untuk menghasilkan prediksi berdasarkan pola yang telah dipelajari.
f. Overfitting
   Overfitting adalah sesuatu yang harus dihindari dalam proses pembangunan model, termasuk dalam Random Forest. Overfitting terjadi ketika model terlalu "menghafal" data latih sehingga kinerjanya buruk pada data uji atau data baru.

3.4. Evaluasi Model

Dalam penelitian ini menggunakan confusion matrix, Confusion Matrix adalah metode evaluasi model klasifikasi yang berbentuk tabel matriks untuk membandingkan hasil prediksi model dengan data aktual. Matriks ini memberikan gambaran detail mengenai prediksi yang benar dan salah untuk setiap kelas, sehingga memudahkan analisis performa model.

Dari Confusion Matrix, dapat menghitung metrik evaluasi seperti:

- True Positive (TP): Data aktual "Good" dan diprediksi sebagai "Good".
- True Negative (TN): Data aktual "Bad" dan diprediksi sebagai "Bad".
- False Positive (FP): Data aktual "Bad" tetapi diprediksi sebagai "Good".
- False Negative (FN): Data aktual "Good" tetapi diprediksi sebagai "Bad".

Akurasi: Persentase prediksi yang benar.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Presisi: Tingkat ketepatan prediksi positif.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall: Kemampuan model mendeteksi data positif.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

F1-Score: Harmonik rata-rata antara presisi dan recall.

$$F1 - Score = 2\,x\frac{precision\,x\,recall}{recall+precision}\,x\,100\% \quad (8)$$

Confusion Matrix memberikan hasil menyeluruh terhadap performa model, baik pada dataset seimbang maupun tidak seimbang, sehingga menjadi metode evaluasi yang umum digunakan.

4. HASIL DAN PEMBAHASAN

4.1. Input Data

Data yang digunakan berasal dari Kaggle, terdiri dari 4000 data, atribut yang digunakan ada 8 yaitu, Size, Weight, Sweetness, Crunchiness, Juiciness, Ripeness, dan Acidity, dengan Quality sebagai target klasifikasi.

4.2. Normalisasi

Normalisasi menggunakan metode Min-Max Scaling, semua atribut memiliki rentang nilai yang seragam. Data dibagi menjadi 2 data training, dan data testing Berikut adalah sampel hasil transformasi data sesudah normalisasi:

Tabel 1. Contoh Hasil Normalisasi

Atribut	Data baris 1	Data baris 2
Size	0.998750	0.105776
Weight	0.636832	0.527487
Sweetness	0.371471	0.581343
Crunchiness	0.519578	0.436464
Juiciness	0.554514	0.442786
Ripeness	0.455305	0.447378
Acidity	0.643043	0.447596
Quality	0.401053	0.335658

4.3. Random Forest

Pada tahap ini, model Random Forest dilatih menggunakan data training yang telah diproses sebelumnya. Data training dibagi dari dataset utama dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Model dibangun dengan parameter utama sebagai berikut:

a. Nilai Fitur yang digunakan pada pohon pertama
Fitur disini adalah atribut yang digunakan, apabila nilai nya tidak memenuhi maka atribut tidak digunakan pada pohon ini.

Tabel 3. fitur yang digunakan pohon pertama

Fitur ke-	Nilai
1	0.04309046
2	0.1750597
3	0.11114144
4	0.18061476
5	0.11774975
6	0.13750195
7	0.13335862
8	0.10148332

Dihitung menggunakan rumus

$$f = \frac{\sum_{t=1}^{T}\sum_{n\in N_t} I(f,n)\Delta impurity(n)}{\sum_{t=1}^{T}\sum_{n\in N_t}\Delta impurity(n)} \quad (9)$$

Lalu hasilnya digunakan untuk memilih fitur yang digunakan.

b. Jumlah pohon yang dibangun: 100
c. Hasil kombinasi output (prediksi):

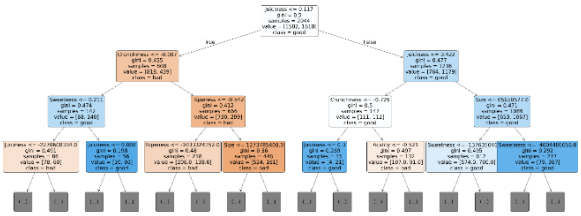
Tabel 4. hasil kombinasi prediksi

Data ke-	Actual	Prediksi
1	good	good
2	good	good
3	bad	good
4	good	bad
5	good	bad
6	bad	good
7	good	bad
8	good	bad
9	bad	bad
10	bad	bad

Pemilihan prediksi berdasarkan nilai dari data yang menjadi paling umum, contoh jika nilai weight = a dan nilai sweetness = b dan hasilnya good, hal ini berulang ulang maka hasil prediksi juga menjadi good tergantung banyak data yang menyerupai.

d. Hasil perhitungan Random Forest:
Accuracy: 0.885
Precision: 0.8850413773274747
Recall: 0.885
F1-Score: 0.8849985624910155
Model yang digunakan pada data latih menggunakan algoritma Random Forest. Algoritma ini menggabungkan banyak pohon keputusan untuk meningkatkan akurasi prediksi dan mencegah overfitting. Hasil evaluasi

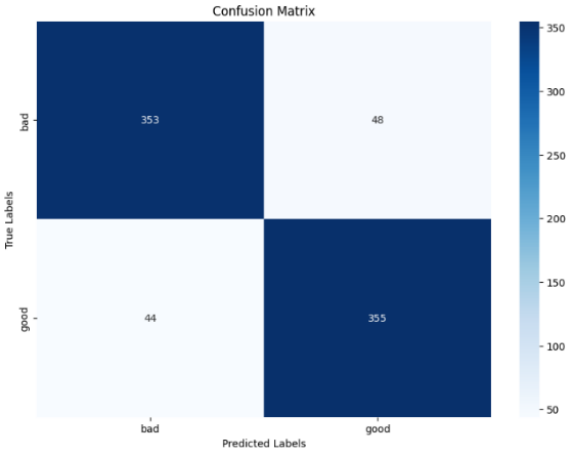
menunjukkan bahwa model telah berhasil mempelajari pola yang ada dalam data pelatihan.



Gambar 3, hasil pohon keputusan random forest

4.4. Evaluasi Confusion Matrix

Dari hasil perhitungan menggunakan Random Forest dapat di simpulkan dalam gambar berikut:



Gambar 4, hasil evaluasi model

Dari gambar 4 hasil dari perhitungan confusion matrix, berikut analisis dan kesimpulan yang dapat diambil:

- True Positive (TP): 355
- True Negative (TN): 353
- False Positive (FP): 48
- False Negative (FN): 44

Dari hasil ini dapat dihitung:

1. Accuracy (Akurasi):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

Apabila data dari evaluasi dimasukan maka:

$$Accuracy = \frac{355+353}{355+353+48+44} = \frac{708}{800} = 0.885 \quad (11)$$

Nilai akurasinya adalah 0.885 atau kalau di persentase adalah 88.5%

2. Presisi (Precision)

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

Jika di masukan data dari evaluasi adalah:

$$Precision = \frac{355}{355+48} = \frac{355}{408} = 0.881 \quad (13)$$

Nilai presisinya adalah 0.881 atau kalau di persentase adalah 88.1%

3. Recall

e. Hasil visual dari pembentukan pohon menggunakan Random Forest

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

Apabila data dari evaluasi dimasukan maka:

$$Recall = \frac{355}{355+44} = \frac{355}{399} = 0.890 \quad (15)$$

Nilai recallnya adalah 0.890 atau kalau di persentase adalah 89.0%

4. F1-Score

$$F1 - Score = 2\,x\frac{precision\ x\ recall}{recall+precision}\ x\ 100\% \quad (16)$$

Jika di masukan data dari evaluasi adalah:

$$F1 - Score = 2\,x\frac{0.881\ x\ 0.890}{0.881+0.890} = 0.885 \quad (17)$$

Nilai f1 scorenya adalah 0.885 atau kalau di persentase adalah 88.5%

Jadi hasil evaluasi menggunakan Decision tree dari data model Random Forest penelitian ini. menunjukkan performa yang baik dengan akurasi 88,5%, precision 88,1%, recall 89,0%, dan F1-Score 88,5%.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengklasifikasikan kualitas apel berdasarkan fitur fisik (Size, Weight, Sweetness, Crunchiness, Juiciness, Ripeness, dan Acidity) menggunakan model Random Forest. Model menunjukkan performa yang baik dengan akurasi 88,5%, precision 88,1%, recall 89,0%, dan F1-Score 88,5%. Berdasarkan confusion matrix, model mampu memprediksi kualitas apel dengan cukup andal, meskipun masih terdapat kesalahan, yaitu 48 data Bad yang salah diprediksi sebagai Good dan 44 data Good yang salah diprediksi sebagai Bad. Hasil ini menunjukkan bahwa Random Forest merupakan metode yang efektif untuk klasifikasi kualitas apel. Namun, terdapat peluang untuk meningkatkan akurasi model, seperti dengan menggunakan teknik optimasi hyperparameter, menambah data pelatihan, atau mencoba algoritma lain, seperti Gradient Boosting atau Neural Networks. Penelitian berikutnya juga disarankan untuk mengeksplorasi fitur tambahan atau menggunakan teknik pra-pemrosesan data yang lebih canggih untuk meningkatkan performa model.

DAFTAR PUSTAKA

[1] A. B. Widodo and M. Mahagiyani, “Analisis kebangkrutan dan mitigasi risiko pada perusahaan perkebunan,” *J. Pengelolaan Perkeb.*, vol. 3, no. 1, pp. 25–35, 2022, doi: 10.54387/jpp.v3i1.13.

[2] N. Herlina and F. Amrullah, “Hubungan Curah Hujan dengan Produktivitas Apel (Malus sylvestris Mill .) di Kabupaten Pasuruan , Jawa Timur The Correlation of Rainfall With Apple (Malus sylvestris Mill .) Productivity in Pasuruan,” vol. 44, no. 1, pp. 11–18, 2020.

[3] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan

Python," *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.

[4] B. Aktavera and H. O. L. Wijaya, "Klasifikasi Produk Menggunakan Algoritma Decision Tree," *J. Teknol. Inf. Mura*, vol. 15, no. 1, pp. 24–29, 2024, doi: 10.32767/jti.v15i1.2264.

[5] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda)," *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.

[6] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.

[7] Marlina Haiza, Elmayati, Zulius Antoni, and Wijaya Harma Oktafia Lingga, "Penerapan Algoritma Random Forest Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo," *Penerapan Kecerdasan Buatan*, vol. 4, no. 2, pp. 138–143, 2023.

[8] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.

[9] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.

[10] M. R. Akbar Ariyadi, S. Lestanti, and S. Kirom, "Klasifikasi Balita Stunting Menggunakan Random Forest Classifier Di Kabupaten Blitar," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3846–3851, 2024, doi: 10.36040/jati.v7i6.7822.

[11] K. Abdi, A. Warjaya, I. Muthmainnah, and P. H. Pahutar, "Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 2, pp. 81–88, 2024, doi: 10.54082/jiki.81.

[12] N. Istiqomah and M. Murinto, "Klasifikasi Penyakit Tanaman Padi Berbasis Citra Daun Menggunakan Convolutional Neural Network (CNN)," *JSTIE (Jurnal Sarj. Tek. Inform.*, vol. 12, no. 1, p. 18, 2024, doi: 10.12928/jstie.v12i1.27314.

[13] J. M. H. Y. Al-afghoni *et al.*, "KLASIFIKASI JENIS BENIH KACANG MENGGUNAKAN SMOTE DAN DECISION TREE C4 . 5," vol. 9, no. 1, pp. 462–469, 2025.

[14] W. Agustiono *et al.*, "MODEL PENERJEMAH BAHASA ISYARAT HURUF HIJAIYAH," vol. 8, no. 6, pp. 12335–12342, 2024.

[15] I. Kurniawan, D. C. P. Buani, A. Abdussomad, W. Apriliah, and R. A. Saputra, "Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.20231026225.