

PENGUNAAN APLIKASI RAPID MINER UNTUK KLASIFIKASI KUALITAS APEL DENGAN ALGORITMA DECISION TREE

Riyan Dwi Saputra¹, Hasbi Firmansyah²

^{1,2} Program Studi S1 Informatika, Universitas Pancasakti Tegal, Kota Tegal

Corresponding author : riyandwisaputraa06@gmail.com

hasbifirmansyah@upstegal.ac.id

Abstrak

Industri pertanian dan pemasaran sangat bergantung pada kualitas apel, jadi diperlukan metode klasifikasi buah yang lebih baik. Menggunakan algoritma Decision Tree yang digunakan oleh perangkat lunak RapidMiner, penelitian ini mengklasifikasikan kualitas apel berdasarkan fitur seperti kemanisan, kesegaran, ukuran, dan berat. Dataset yang digunakan terdiri dari 4001 entri dengan 9 atribut yang diperoleh dari platform Kaggle. Untuk memastikan konsistensi dan akurasi model, tahap preprocessing data dilakukan. Model klasifikasi menghasilkan akurasi sebesar 52,76%, dengan recall yang tinggi untuk kelas "Good" tetapi rendah untuk kelas "Bad". Sebuah ketidakseimbangan data diidentifikasi sebagai komponen utama yang memengaruhi kinerja model. Hasil penelitian menunjukkan betapa pentingnya menyesuaikan parameter algoritma dan menyeimbangkan data untuk meningkatkan akurasi dan keandalan model prediksi..

Kata Kunci : klasifikasi kualitas apel, *decision tree*, rapid miner, data mining, *machine learning*

Abstract

The agricultural and marketing industries are highly dependent on the quality of apples, so better methods of fruit classification are needed. Using the Decision Tree algorithm used by RapidMiner software, the study classified the quality of apples based on features such as sweetness, freshness, size, and weight. The Dataset used consists of 4001 entries with 9 attributes obtained from the Kaggle platform. To ensure the consistency and accuracy of the model, the data preprocessing stage is performed. Classification Model produces accuracy of 52.76%, with high recall for the Class "Good" but low for the class "Bad". A data imbalance was identified as a major component affecting model performance. The results show how important it is to adjust algorithm parameters and balance the data to improve the accuracy and reliability of predictive models..

Keywords: Apple quality classification, decision tree, rapid miner, data mining, machine learning

PENDAHULUAN

Dalam industri pertanian dan pemasaran, kualitas buah apel sangat penting. Akibatnya, penelitian telah dimotivasi untuk mengembangkan metode yang lebih efektif untuk memprediksi dan meningkatkan kualitas ini. Beberapa metode konvensional, seperti pengukuran manual dan analisis laboratorium, telah digunakan, tetapi mereka cenderung memakan waktu dan tidak selalu memberikan hasil yang akurat. Ada potensi besar untuk meningkatkan prediksi kualitas buah apel dengan kemajuan pesat dalam teknologi, khususnya dalam konteks machine learning. Penggunaan teknik analisis prediktif berbasis machine learning dapat membantu menemukan pola kompleks dalam data yang sulit diakses oleh metode konvensional. Selain itu, menggunakan preprocessing data dalam konteks ini dapat membantu mengoptimalkan kualitas dataset, meningkatkan akurasi model, dan mengurangi bias yang mungkin terjadi. (Indu Indah Purnomo, 2024)

Data mining adalah metode yang digunakan untuk mengekstrak informasi dari kumpulan data yang belum ditemukan secara manual. Istilah ini mulai dikenal karena pentingnya pengolahan

data dalam berbagai bidang akademik, kesehatan, pendidikan, dan bahkan industri. Ilmuwan data mining menggabungkan kecerdasan buatan, statistik, dan penelitian basis data yang terus berkembang. (Ekin Adhi Guna, 2023)

Klasifikasi adalah proses mencari sekumpulan model, pola, atau fungsi yang menggambarkan dan membedakan objek data sehingga dapat dimasukkan ke dalam kelas tertentu dari banyak kelas yang tersedia. Proses klasifikasi melibatkan pencarian objek yang memiliki fitur tertentu, yang kemudian dimasukkan ke dalam salah satu kelas yang sudah diartikan sebelumnya. Setiap dataset yang digunakan harus memiliki label atau fitur yang dimaksudkan untuk diklasifikasikan. Untuk menghitung akurasi klasifikasi, persentase instance yang diberi label kelas yang tepat dihitung. Untuk dapat memperkirakan kelas objek yang belum diketahui, klasifikasi adalah tujuan. (Indu Indah Purnomo, 2024)

Decision Tree adalah struktur *flowchart* yang terdiri dari pohon, di mana setiap simpul internal menunjukkan suatu tes atribut. Setiap cabang adalah representasi distribusi kelas atau kelas. Untuk kasus ini, alur *decision tree* memegang prediksi kelas dari simpul ke akar hingga simpul daun. Keaturan klasifikasi dapat dengan mudah diubah menjadi pohon keputusan. Data ditampilkan dalam bentuk tabel dengan atribut dan catatan di pohon keputusan. (Indu Indah Purnomo, 2024)

Rapid Miner adalah platform pembelajaran mesin dan ilmu data yang kuat dengan banyak alat untuk pemodelan, evaluasi, implementasi, dan persiapan data. Dengan antarmuka yang mudah digunakan, perangkat lunak ini memungkinkan orang tanpa pengalaman pemrograman untuk membangun dan menguji berbagai model. *Rapid Miner* memungkinkan pengguna membuat alur kerja yang memungkinkan pemrosesan dan analisis data melalui antarmuka drag-and-drop. Perangkat lunak ini mendukung berbagai sumber data, termasuk basis data, file datar, dan platform big data seperti Hadoop dan Spark. Selain itu, perangkat lunak ini mencakup banyak operator yang sudah dibangun, yang merupakan blok bangunan dari alur kerja yang mencakup semua tahap proses penambangan data, seperti pembersihan data, pemilihan fitur, dan pemasangan algoritma. (Muhammad Rafi Nahjan, 2023)

Penelitian ini menggunakan algoritma *Decision Tree*, akan membuat klasifikasi dari kualitas apel. Data yang diolah akan mengklasifikasi “Good” dan “Bad” pada kualitas apel. Hasil penelitian ini akan menunjukkan bahwa metode *Decision Tree* adalah solusi yang efektif dan efisien untuk klasifikasi kualitas apel. Dengan ini, metode *Decision Tree* dapat digunakan oleh industri apel atau petani untuk meningkatkan produktivitas dan kualitas hasil panen mereka.

METODE PENELITIAN

1. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data kualitas apel yang diperoleh dari platform Kaggle (www.kaggle.com). Dataset ini terdiri dari 4001 entri dengan 9 atribut. Kumpulan data ini berisi informasi tentang berbagai atribut dari sekumpulan buah, yang memberikan wawasan tentang karakteristiknya. Kumpulan data tersebut mencakup detail seperti *ID* buah, ukuran, berat, kemanisan, kerenyahan, kesegaran, kematangan, keasaman, dan kualitas.

2. Preprocessing Data

Proses persiapan data termasuk :

- Pembersihan data untuk menghilangkan nilai-nilai yang hilang (*missing values*)
- Normalisasi atribut numerik untuk memastikan konsistensi
- Encoding atribut kategorikal menjadi format numerik

3. Implementasi Algoritma Decision Tree

Decision Tree adalah algoritma yang digunakan untuk membuat model prediktif berdasarkan aturan-aturan keputusan yang dihasilkan dari data pelatihan. *Rapid Miner* digunakan untuk membangun dan mengevaluasi model ini dengan Langkah-langkah sebagai berikut :

- Data dibagi menjadi subset pelatihan, dan pengujian biasanya menggunakan rekomendasi 70:30 atau 80:20 untuk membagi data
- Membangun model Decision Tree menggunakan operator "Decision Tree".
- Evaluasi kinerja model menggunakan metrik akurasi, presisi, dan recall.

HASIL

1. Performance

Dengan menggunakan perangkat lunak *Rapid Miner*, algoritma Decision Tree digunakan untuk mengklasifikasikan kualitas apel. Data yang digunakan terdiri dari kumpulan atribut seperti warna, tekstur, ukuran, dan tingkat kekerasan. Dalam pengujian ini, algoritma ini berhasil mencapai akurasi sebesar 52,76%, seperti yang ditunjukkan oleh tabel dibawah ini :

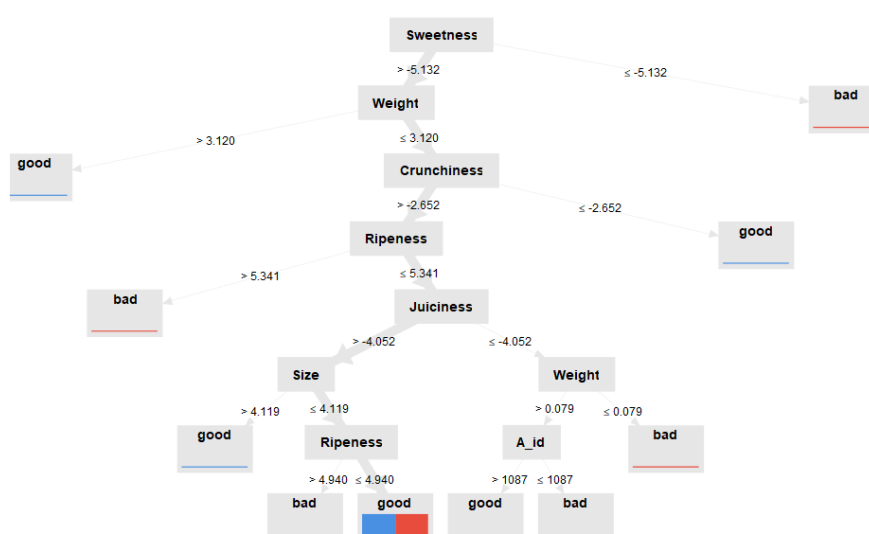
Accuracy : 52.76%

	<i>true good</i>	<i>true bad</i>	<i>class precision</i>
<i>pred.good</i>	1604	1512	51.48%
<i>pred.bad</i>	0	85	100.00%
<i>class recall</i>	100.00%	5.32%	

2. Graph View

Hasil klasifikasi dengan percabangan yang dapat menghasilkan kesimpulan ditunjukkan dalam tampilan pohon keputusan. Hasil dari *Decision Tree* dapat dilihat dibawah ini :

Gambar 1. *Decision Tree* Hasil Klasifikasi Kualitas Apel (Pohon Keputusan)



Sumber : Dokumentasi Sendiri

PEMBAHASAN

Hasil pengujian menunjukkan beberapa hal penting yang harus diperhatikan :

1. Tingkat Akurasi Rendah:

Model ini memiliki tingkat akurasi sebesar 52,76%, menunjukkan bahwa itu tidak mampu mengklasifikasikan data dengan baik. Kemungkinan penyebabnya adalah ketidakseimbangan data (*imbalance data*), yang terjadi pada dataset antara kelas "Good" dan "Bad". Seperti yang terlihat, model memprediksi sebagian besar data "Good" dengan benar, tetapi hanya dapat memprediksi sebagian kecil data "Bad".

2. Ketidakseimbangan Prediksi

Seperti yang ditunjukkan oleh tabel hasil akurasi, kelas "Good" memiliki recall sebesar 100,00%, yang menunjukkan bahwa semua data yang sebenarnya "Good" dapat dikenali dengan sukses oleh model. Sebaliknya, kelas "Bad" memiliki recall hanya 5,32%, yang menunjukkan bahwa model tidak dapat mengenali sebagian besar data "Bad".

3. Precision pada kelas "Good"

Precision untuk kelas "Good" hanya sebesar 51,48%, yang berarti hanya 51,48% dari seluruh prediksi "Good" benar. Ini menunjukkan bahwa model sering salah menafsirkan data "Bad" sebagai "Good".

4. Kemungkinan Perbaikan:

Ada beberapa cara untuk meningkatkan kinerja model. Salah satunya adalah penyeimbangan data (*Data Balancing*). Metode seperti oversampling pada kelas minoritas ("Bad") atau undersampling pada kelas mayoritas ("Good") dapat membantu mengurangi ketidakseimbangan data.

Tuning Parameter Model : Mengubah parameter algoritma Pilihan Pohon, seperti kedalaman maksimum (*max depth*) atau jumlah sampel minimum per node, dapat meningkatkan performa.

KESIMPULAN

Studi ini menemukan bahwa algoritma *Decision Tree* yang digunakan oleh perangkat lunak RapidMiner mampu mengklasifikasikan kualitas apel dengan akurasi sebesar 52,76%. Hasil ini menunjukkan bahwa model mengenali apel kelas "Good" lebih baik daripada apel kelas "Bad", yang mungkin karena ketidakseimbangan data dalam dataset. Namun, penelitian ini menunjukkan bahwa prosedur preprocessing data seperti pembersihan dan normalisasi sangat penting untuk meningkatkan kualitas dataset untuk mendukung analisis yang lebih baik.

Disarankan untuk menggunakan metode penyeimbangan data seperti oversampling atau undersampling dan mengatur parameter algoritma untuk mengoptimalkan akurasi klasifikasi untuk meningkatkan kinerja model. Selain itu, penelitian ini meningkatkan kemungkinan *machine learning*, khususnya algoritma *Decision Tree*, dapat membantu proses pengambilan keputusan berbasis data di sektor pertanian.

DAFTAR PUSTAKA

- Ekin Adhi Guna, M. D. (2023). Implementasi Algoritma Decision Tree untuk Klasifikasi Data Evaluation. *Jurnal Sistem Informasi dan Ilmu Komputer*, 1(4), 167-177.
- Indu Indah Purnomo, G. A. (2024). Analisis Prediktif Dan Preprocessing Untuk Kualitas Buah Apel. *Technologia : Jurnal Ilmiah* 15(4), 681-687.
- Muhammad Rafi Nahjan, N. H. (2023). IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101-104.