

KOMPARASI C4.5 DAN NAÏVE BAYES UNTUK KLASIFIKASI KETERJUALAN PRODUK PETSHOP BAYBAY SIDOARJO

COMPARISON OF C4.5 AND NAÏVE BAYES FOR PRODUCT SALES CLASSIFICATION AT PETSHOP BAYBAY SIDOARJO

Wahyu Ajeng Shintia Ningrum^{1*}, Yudha Herlambang Cahya Pratama²

*E-mail: wshintianingrum@gmail.com

^{1,2} Sistem Informasi, Fakultas Teknik dan Desain, Universitas Hayam Wuruk Perbanas

Abstrak

Penelitian ini bertujuan untuk membandingkan kinerja algoritma Decision Tree (C4.5) dan Naïve Bayes dalam mengklasifikasikan produk terlaris pada Petshop Bay Bay Sidoarjo berdasarkan data histori penjualan. Klasifikasi produk dilakukan untuk mengelompokkan tingkat keterjualan produk ke dalam tiga kelas, yaitu tinggi, sedang, dan rendah, sehingga dapat membantu pihak petshop dalam pengambilan keputusan terkait pengelolaan stok dan strategi penjualan. Data yang digunakan dalam penelitian ini merupakan data histori penjualan produk hewan peliharaan periode Januari 2022 hingga Desember 2024 yang diperoleh langsung dari pihak Petshop Bay Bay Sidoarjo. Proses pengolahan dan pengujian data dilakukan menggunakan perangkat lunak RapidMiner dengan tahapan praproses data, pembentukan model klasifikasi, serta evaluasi performa model. Pengukuran kinerja algoritma dilakukan menggunakan metrik accuracy, weighted mean precision, dan weighted mean recall. Hasil penelitian menunjukkan bahwa algoritma Decision Tree (C4.5) memiliki kinerja yang lebih baik dibandingkan dengan algoritma Naïve Bayes berdasarkan nilai evaluasi yang diperoleh. Dengan demikian, algoritma Decision Tree (C4.5) dinilai lebih efektif untuk digunakan dalam proses klasifikasi produk terlaris pada Petshop Bay Bay Sidoarjo. Hasil penelitian ini diharapkan dapat menjadi referensi bagi pelaku usaha dalam menerapkan teknik data mining untuk mendukung pengambilan keputusan berbasis data.

Kata kunci: data mining, decision tree (C4.5), naïve bayes, klasifikasi produk, rapidminer.

Abstract

This study aims to compare the performance of the Decision Tree (C4.5) and Naïve Bayes algorithms in classifying best-selling products at Petshop Bay Bay Sidoarjo based on historical sales data. Product classification is conducted to group the level of product sales into three classes, namely high, medium, and low, in order to support decision-making related to inventory management and sales strategies. The data used in this study consist of historical sales data of pet products from January 2022 to December 2024, obtained directly from Petshop Bay Bay Sidoarjo. Data processing and model testing were carried out using the RapidMiner software, including data preprocessing, model construction, and performance evaluation. The performance of the algorithms was measured using accuracy, weighted mean precision, and weighted mean recall metrics. The results of this study indicate that the Decision Tree (C4.5) algorithm demonstrates better performance than the Naïve Bayes algorithm based on the evaluation results obtained. Therefore, the Decision Tree (C4.5) algorithm is considered more effective for classifying best-selling products at Petshop Bay Bay Sidoarjo. The findings of this study

are expected to serve as a reference for business practitioners in implementing data mining techniques to support data-driven decision-making.

Keywords: *data mining, decision tree (C4.5), naïve bayes, product classification, rapidminer.*

1. PENDAHULUAN

Sejumlah kemajuan teknologi informasi telah mendorong perkembangan *data mining*, khususnya dalam pengolahan dan analisis data penjualan. Di era digital, volume transaksi terus meningkat seiring penggunaan sistem informasi dan platform digital dalam kegiatan bisnis. *Data mining* merupakan teknik eksplorasi data untuk menemukan pola dan relasi dalam data berukuran besar guna mendukung pengambilan keputusan [1]. Dalam konteks bisnis ritel, penerapan *data mining* membantu pelaku usaha memahami perilaku konsumen dan mengidentifikasi produk dengan tingkat permintaan tertinggi [2]. Oleh karena itu, penggunaan metode klasifikasi menjadi penting untuk membantu pengelolaan data penjualan secara lebih efektif dan efisien.

Perkembangan digitalisasi juga dirasakan oleh sektor UMKM di Indonesia, termasuk usaha penjualan produk hewan peliharaan seperti Petshop Bay Bay Sidoarjo. Usaha ini memiliki volume penjualan yang tinggi dengan berbagai jenis produk yang dijual setiap hari. Permasalahan yang dihadapi adalah kesulitan menentukan produk terlaris secara konsisten tanpa bantuan analisis berbasis data. Analisis penjualan menggunakan *data mining* dapat membantu pengelolaan stok dan penyusunan strategi pemasaran yang lebih tepat [3]. Dengan adanya analisis tersebut, pelaku usaha dapat mengambil keputusan bisnis berdasarkan data transaksi yang dimiliki. Hal ini menunjukkan pentingnya penerapan metode klasifikasi dalam mendukung pengembangan UMKM berbasis digital.

Beberapa penelitian sebelumnya telah membandingkan algoritma *Decision Tree* dan *Naïve Bayes* dalam klasifikasi data penjualan. Algoritma *decision Tree C4.5* memiliki akurasi yang lebih stabil dibandingkan *Naïve Bayes* pada analisis sentimen produk [4]. Penelitian lain menunjukkan bahwa kedua algoritma memiliki tingkat akurasi yang relatif seimbang, namun *Decision Tree* lebih mudah dipahami karena menghasilkan aturan klasifikasi yang jelas [5]. Selain itu, setiap algoritma memiliki keunggulan sesuai dengan karakteristik data yang digunakan [6]. Berdasarkan hasil penelitian tersebut, penulis tertarik membandingkan kedua algoritma pada data penjualan UMKM untuk menentukan metode yang paling optimal dalam mengklasifikasikan produk terlaris. Penelitian ini menggunakan data penjualan dari Petshop Bay Bay Sidoarjo sebagai objek penelitian.

Penelitian ini memanfaatkan aplikasi *RapidMiner* untuk mengimplementasikan algoritma *Decision Tree (C4.5)* dan *Naïve Bayes*. *RapidMiner* merupakan perangkat lunak *open-source* yang mudah digunakan dalam proses data mining tanpa memerlukan kemampuan pemrograman yang kompleks [7]. *RapidMiner* juga mampu memberikan visualisasi pola data penjualan dengan lebih jelas pada pengujian algoritma *Decision Tree* [8]. Selain itu, penelitian lain menunjukkan bahwa *Naïve Bayes* cenderung memberikan hasil baik pada data berukuran besar, sedangkan *Decision Tree* lebih sesuai untuk data kategorikal [9]. Evaluasi penelitian dilakukan menggunakan *accuracy*, *precision*, dan *recall* untuk mengetahui performa masing-masing algoritma [10]. Dengan demikian, hasil penelitian diharapkan dapat menunjukkan metode klasifikasi yang paling sesuai untuk data penjualan yang digunakan.

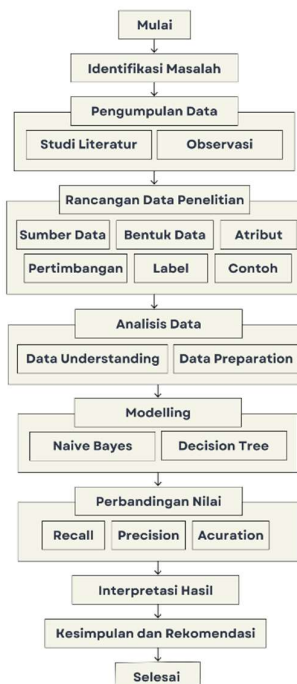
Klasifikasi produk terlaris berbasis *data mining* memiliki manfaat besar dalam mendukung pengelolaan stok dan strategi bisnis yang lebih efektif. Analisis data penjualan dapat membantu pelaku usaha mengidentifikasi produk dengan performa terbaik secara lebih cepat dan akurat [11]. Hasil kualifikasi juga dapat digunakan untuk mengetahui tren pembelian konsumen sehingga strategi pemasaran dapat disesuaikan dengan kebutuhan pasar. Penelitian ini berfokus pada implementasi, perhitungan, analisis, dan perbandingan hasil klasifikasi menggunakan

algoritma *Decision Tree* dan *Naïve Bayes* dengan bantuan *RapidMiner*. Diharapkan penelitian ini mampu memberikan manfaat teoritis maupun praktis bagi penulis, peneliti selanjutnya, dan Petshop Bay Bay Sidoarjo. Selain itu, hasil penelitian diharapkan dapat menjadi dasar pengambilan keputusan bisnis berbasis data yang lebih tepat dan berkelanjutan.

2. METODOLOGI

Penelitian ini dilakukan melalui beberapa tahapan yang disusun secara sistematis untuk mencapai tujuan klasifikasi produk terlaris menggunakan algoritma *Decision Tree (C4.5)* dan *Naïve Bayes*. Permasalahan utama yang diangkat adalah belum adanya analisis berbasis data dalam menentukan produk terlaris pada Petshop Bay Bay Sidoarjo. Selama ini proses penentuan produk terlaris masih dilakukan secara manual berdasarkan pengamatan penjualan harian. Kondisi tersebut berpotensi menyebabkan ketidaktepatan dalam pengelolaan stok dan pengambilan keputusan bisnis. Oleh karena itu, penelitian ini bertujuan membangun model klasifikasi yang mampu mengidentifikasi kategori produk terlaris secara objektif dan lebih akurat. Dengan adanya model klasifikasi, pengelola toko diharapkan dapat memahami pola penjualan dan menyusun strategi bisnis berbasis data secara lebih efektif.

Tahapan penelitian dimulai dari proses identifikasi masalah. Setiap tahapan dilakukan secara berurutan agar proses penelitian dapat berjalan secara sistematis dan menghasilkan model klasifikasi yang optimal. Pada tahap awal, peneliti melakukan identifikasi terhadap permasalahan penjualan produk yang terjadi pada Petshop Bay Bay Sidoarjo. Selanjutnya dilakukan proses pengumpulan dan pengolahan data sebelum masuk ke tahap pemodelan menggunakan algoritma klasifikasi. Tahapan penelitian secara keseluruhan dapat dilihat pada diagram alur penelitian berikut.



Gambar 1. Tahapan Penelitian

Pengumpulan data dilakukan melalui dua metode, yaitu studi literatur dan observasi. Studi literatur dilakukan dengan menelaah jurnal, buku, prosiding, dan penelitian terdahulu yang berkaitan dengan algoritma *Decision Tree (C4.5)*, *Naïve Bayes*, serta analisis data penjualan.

Tahap ini bertujuan memperoleh landasan teori yang kuat mengenai metode yang digunakan dalam penelitian. Selain itu, observasi dilakukan dengan mempelajari data penjualan yang diperoleh langsung dari pihak Petshop Bay Bay dalam bentuk file digital. Data tersebut digunakan untuk memahami struktur data, atribut yang tersedia, serta pola penjualan produk. Kedua metode pengumpulan data ini mendukung proses penelitian agar sesuai dengan kondisi nyata yang terjadi di lapangan.

Rancangan data penelitian disusun berdasarkan kebutuhan klasifikasi tingkat keterjualan produk menggunakan data histori penjualan. Dataset berasal dari transaksi penjualan Petshop Bay Bay yang telah terdokumentasi selama beberapa tahun dan disajikan dalam bentuk tabel *spreadsheet*. Setiap baris pada *dataset* merepresentasikan transaksi penjualan, sedangkan setiap kolom menunjukkan atribut pendukung analisis. Atribut yang digunakan meliputi tanggal transaksi, nama produk, kategori produk, jumlah terjual, dan harga jual. Atribut tersebut dipilih karena memiliki pengaruh langsung terhadap tingkat keterjualan produk. *Output* dari penelitian ini berupa kategori tingkat penjualan, yaitu tinggi, sedang, dan rendah. Contoh struktur *dataset* penelitian dapat dilihat pada tabel berikut.

Table 1. Contoh Dataset Penelitian

Tahun	Bulan	Nama Produk	Kategori Produk	Harga Jual	Jumlah Terjual	Total penj.	Ket
2022	1	Cat Litter Basic	Kebutuhan	Rp 2000	3	Rp 6000	Tinggi
...	...	...	...	...	...	...	...

Tahap analisis data dilakukan berdasarkan metode *CRISP-DM* yang mencakup data *understanding* dan data *preparation*. Pada tahap data *understanding*, peneliti melakukan eksplorasi awal terhadap *dataset* untuk memahami karakteristik, kelengkapan, dan distribusi data. Proses ini bertujuan memastikan bahwa data yang digunakan relevan dengan kebutuhan penelitian. Selain itu, dilakukan identifikasi terhadap atribut penting serta kemungkinan adanya data kosong atau tidak konsisten. Selanjutnya, tahap data *preparation* dilakukan dengan membersihkan data, memilih atribut yang relevan, dan melakukan transformasi data. Tahapan ini penting agar dataset siap digunakan dalam proses pemodelan algoritma klasifikasi.

Pembagian data pada penelitian ini menggunakan metode *k-Fold Cross Validation* dengan nilai $k = 10$. *Dataset* dibagi menjadi sepuluh bagian dengan ukuran yang relatif sama, sehingga setiap data memperoleh kesempatan yang sama sebagai data training maupun data testing. Dalam setiap iterasi, 90% data digunakan sebagai *training* data dan 10% digunakan sebagai *testing* data. Proses pengujian dilakukan sebanyak sepuluh kali agar hasil evaluasi lebih objektif dan stabil. Data *training* digunakan untuk membangun model klasifikasi, sedangkan data *testing* digunakan untuk mengukur kemampuan model dalam mengklasifikasikan data baru. Hasil evaluasi dari seluruh iterasi kemudian dirata-ratakan untuk memperoleh nilai *accuracy*, *precision*, dan *recall*.

Tahap modelling dilakukan menggunakan algoritma *Naïve Bayes* dan *Decision Tree* (C4.5) melalui perangkat lunak *RapidMiner*. *Naïve Bayes* merupakan metode klasifikasi probabilistik yang menggunakan Teorema Bayes dengan asumsi antar atribut bersifat independen. Algoritma ini dikenal memiliki efisiensi komputasi yang tinggi dan cocok digunakan pada data berukuran besar. Sementara itu, *Decision Tree* (C4.5) bekerja dengan membentuk struktur pohon keputusan berdasarkan nilai *information gain* tertinggi dari setiap atribut. Pohon keputusan yang dihasilkan memudahkan pengguna dalam memahami pola hubungan antar atribut data penjualan. Kedua algoritma tersebut diterapkan untuk mengetahui metode yang paling optimal dalam mengklasifikasikan produk terlaris.

Evaluasi model dilakukan menggunakan tiga parameter utama, yaitu *accuracy*, *precision*, dan *recall*. *Accuracy* digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan data secara keseluruhan. *Precision* digunakan untuk menilai ketepatan hasil prediksi positif yang dihasilkan model, sedangkan *recall* digunakan untuk mengukur kemampuan model dalam mengenali seluruh data positif yang sebenarnya. Dalam penelitian ini, ketiga parameter dihitung menggunakan pendekatan *weighted mean* karena klasifikasi dilakukan pada tiga kelas, yaitu tinggi, sedang, dan rendah. Hasil evaluasi dari algoritma *Naïve Bayes* dan *Decision Tree (C4.5)* kemudian dibandingkan untuk menentukan performa terbaik. Perbandingan dilakukan menggunakan *confusion matrix* yang menunjukkan hubungan antara data aktual dan hasil prediksi model.

Interpretasi hasil dilakukan untuk memahami makna dari nilai *accuracy*, *precision*, dan *recall* yang diperoleh dari kedua algoritma. *Decision Tree (C4.5)* memiliki keunggulan dalam menampilkan aturan keputusan dalam bentuk pohon yang mudah dipahami dan dianalisis. Sementara itu, *Naïve Bayes* unggul dalam kecepatan komputasi dan efisiensi pengolahan data dengan jumlah atribut besar. Hasil penelitian menunjukkan bahwa setiap algoritma memiliki kelebihan yang berbeda sesuai dengan karakteristik data yang digunakan. Oleh karena itu, pemilihan algoritma terbaik perlu disesuaikan dengan kebutuhan analisis dan tujuan penelitian. Dalam konteks penelitian ini, hasil klasifikasi diharapkan dapat membantu Petshop Bay Bay Sidoarjo dalam menentukan strategi penjualan dan pengelolaan stok produk secara lebih efektif dan berbasis data.

3. HASIL DAN PEMBAHASAN

Hasil penelitian disajikan dalam bentuk proses pengolahan data, implementasi algoritma, serta evaluasi performa model menggunakan metrik *accuracy*, *weighted mean precision*, dan *weighted mean recall*. Selain itu, pada bab ini juga dilakukan pembahasan mengenai perbandingan performa kedua algoritma dalam mengklasifikasikan produk terlaris.

3.1 Hasil

Data yang digunakan dalam penelitian ini merupakan data historis penjualan produk dari Petshop Bay Bay yang digunakan untuk proses klasifikasi produk terlaris menggunakan algoritma *Naïve Bayes* dan *Decision Tree (C4.5)*. *Dataset* terdiri dari 360 data penjualan bulanan selama periode Januari 2022 sampai Desember 2024. Atribut yang digunakan meliputi tahun, bulan, nama produk, kategori produk, harga jual, jumlah terjual, total penjualan, dan keterangan tingkat keterjualan. Untuk memberikan gambaran struktur dataset penelitian, contoh data penjualan disajikan pada tabel berikut.

Table 2. Data Penjualan Petshop Bay Bay

Tahun	Bulan	Nama Produk	Kategori Produk	Harga Jual	Jumlah Terjual	Total penj.	Ket
2022	1	Bola Mainan Hewan	Aksesoris	Rp 4000	3	Rp 12000	Rendah
...	...	...	...	...	...	...	...

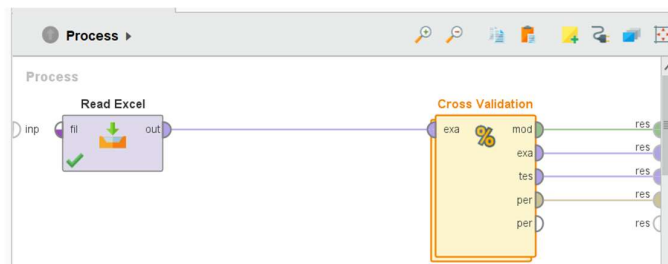
Kemudian dilakukan tahapan data *understanding* dan data *preparation*. Pada tahap data *understanding* dilakukan identifikasi terhadap kondisi awal *dataset*. Hasil analisis menunjukkan masih terdapat format data yang belum seragam, nilai kosong, serta beberapa atribut yang kurang relevan terhadap proses klasifikasi. Tahap data *preparation* dilakukan dengan membersihkan dan mentransformasi data agar lebih konsisten dan siap digunakan pada proses pemodelan. Beberapa kolom administratif dan data rekapitulasi dihapus karena tidak diperlukan dalam klasifikasi. Selain itu, atribut bulan diubah ke format teks dan nama produk disederhanakan menjadi kode

numerik untuk mempermudah pengolahan data pada *RapidMiner*. Hasil akhir tahap ini berupa *dataset* yang lebih terstruktur dan siap digunakan pada proses implementasi algoritma. Transformasi yang menunjukkan data siap digunakan pada tahap pemilihan atribut dan pemodelan dapat dilihat pada tabel berikut.

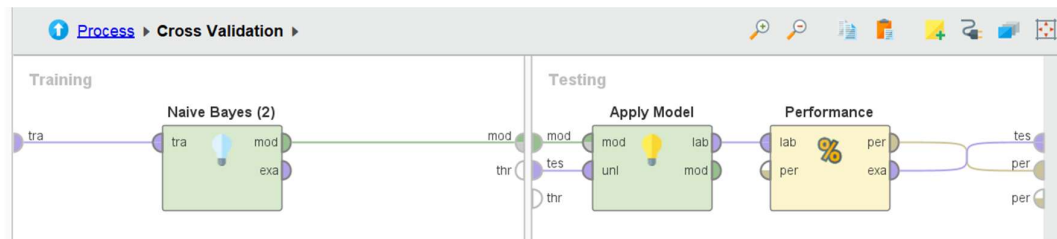
Table 3. Dataset Hasil Data Preparation

tahun	bulan	produk	kategori	harga	keterangan
2022	jan	1	aksesoris	4000	rendah
2022	Jan	2	aksesoris	19900	sedang
2022	jan	3	aksesoris	9000	rendah
...	...	...	...	...	...

Implementasi algoritma *Naïve Bayes* dilakukan menggunakan aplikasi *RapidMiner*. *Dataset* diimpor menggunakan operator *Read Excel*, kemudian dilakukan pengaturan tipe data dan penentuan atribut keterangan sebagai label kelas. Proses pengujian menggunakan metode *Cross Validation* untuk membagi data *training* dan *testing* secara otomatis. Pada tahap *training*, algoritma *Naïve Bayes* digunakan untuk membangun model klasifikasi berdasarkan data latih. Selanjutnya, model diuji menggunakan data testing dan dievaluasi menggunakan operator *Performance* untuk menghasilkan nilai *accuracy*, *weighted mean precision*, dan *weighted mean recall*.



Gambar 2. Penghubungan operator Read Excel dengan Cross Validation pada RapidMiner



Gambar 3. Struktur training dan testing algoritma Naïve Bayes

Hasil pengujian menunjukkan bahwa algoritma *Naïve Bayes* memperoleh nilai *accuracy* sebesar 55,28%, *weighted mean precision* sebesar 56,89%, dan *weighted mean recall* sebesar 57,49%. Berdasarkan hasil tersebut, model *Naïve Bayes* sudah mampu melakukan klasifikasi data penjualan, namun masih terdapat kesalahan klasifikasi pada beberapa kelas keterjualan. Model ini tetap mampu mengenali pola dasar pada data penjualan yang digunakan dalam penelitian. Selain itu, hasil evaluasi menunjukkan bahwa algoritma *Naïve Bayes* dapat digunakan sebagai pembandingan dalam menentukan algoritma dengan performa terbaik.

accuracy: 55.28% +/- 8.92% (micro average: 55.28%)

	true rendah	true sedang	true tinggi	class precision
pred. rendah	75	28	12	65.22%
pred. sedang	36	64	12	57.14%
pred. tinggi	32	41	60	45.11%
class recall	52.45%	48.12%	71.43%	

Gambar 4. Accuracy Naïve Bayes

weighted_mean_precision: 56.89% +/- 9.90% (micro average: 55.82%), weights: 1, 1, 1

	true rendah	true sedang	true tinggi	class precision
pred. rendah	75	28	12	65.22%
pred. sedang	36	64	12	57.14%
pred. tinggi	32	41	60	45.11%
class recall	52.45%	48.12%	71.43%	

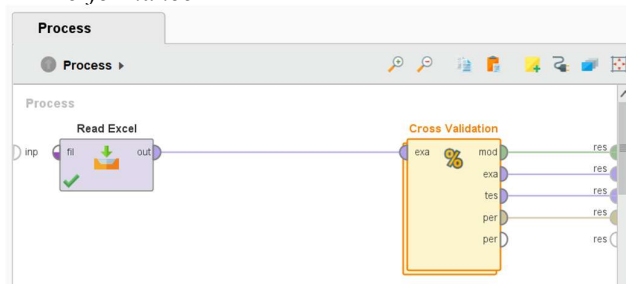
Gambar 5. Weighted Mean Precision Naïve Bayes

weighted_mean_recall: 57.49% +/- 10.19% (micro average: 57.33%), weights: 1, 1, 1

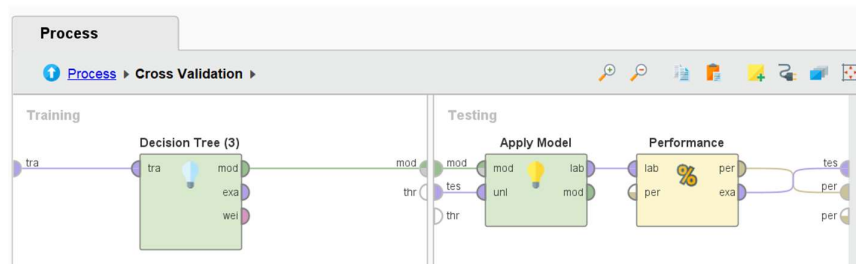
	true rendah	true sedang	true tinggi	class precision
pred. rendah	75	28	12	65.22%
pred. sedang	36	64	12	57.14%
pred. tinggi	32	41	60	45.11%
class recall	52.45%	48.12%	71.43%	

Gambar 6. Weighted Mean Recall Naïve Bayes

Implementasi algoritma *Decision Tree (C4.5)* dilakukan menggunakan *dataset* yang sama dengan algoritma *Naïve Bayes* agar hasil pengujian dapat dibandingkan secara objektif. *Dataset* diimpor ke dalam *RapidMiner* menggunakan operator *Read Excel*, kemudian dilakukan pengaturan atribut dan penentuan label kelas sebelum proses klasifikasi dilakukan. Proses *training* dan *testing* dilakukan menggunakan metode *Cross Validation*. Pada tahap *training* digunakan operator *Decision Tree (C4.5)* untuk membentuk model klasifikasi berupa pohon keputusan. Model yang dihasilkan kemudian diuji menggunakan data testing dan dievaluasi menggunakan operator *Performance*.



Gambar 7. Proses Cross Validation pada algoritma Decision Tree



Gambar 8. Struktur training dan testing algoritma Decision Tree (C4.5)

Hasil evaluasi menunjukkan bahwa algoritma *Decision Tree (C4.5)* memperoleh nilai *accuracy* sebesar 77,22%, *weighted mean precision* sebesar 80,96%, dan *weighted mean recall* sebesar 77,63%. Nilai tersebut menunjukkan bahwa *Decision Tree* mampu melakukan klasifikasi data penjualan dengan tingkat ketepatan yang lebih baik dibandingkan *Naïve Bayes*. Hasil ini menunjukkan bahwa algoritma *Decision Tree* lebih mampu mengenali pola hubungan antar atribut pada data penjualan. Struktur pohon keputusan yang dihasilkan juga mempermudah proses interpretasi hasil klasifikasi. Selain itu, model mampu membagi data ke dalam kategori keterjualan dengan lebih terstruktur dan jelas. Tingginya nilai *precision* menunjukkan bahwa

kesalahan prediksi kelas relatif lebih sedikit dibandingkan algoritma sebelumnya. Sementara itu, nilai *recall* yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali data aktual pada setiap kelas keterampilan produk.

accuracy: 77.22% +/- 4.10% (micro average: 77.22%)

	true rendah	true sedang	true tinggi	class precision
pred. rendah	120	41	3	73.17%
pred. sedang	22	89	12	72.36%
pred. tinggi	1	3	69	94.52%
class recall	83.92%	66.92%	82.14%	

Gambar 9. Accuracy Decision Tree

weighted_mean_precision: 80.96% +/- 3.55% (micro average: 80.02%), weights: 1, 1, 1

	true rendah	true sedang	true tinggi	class precision
pred. rendah	120	41	3	73.17%
pred. sedang	22	89	12	72.36%
pred. tinggi	1	3	69	94.52%
class recall	83.92%	66.92%	82.14%	

Gambar 10. Weighted Mean Precision Decision Tree

weighted_mean_recall: 77.63% +/- 4.91% (micro average: 77.66%), weights: 1, 1, 1

	true rendah	true sedang	true tinggi	class precision
pred. rendah	120	41	3	73.17%
pred. sedang	22	89	12	72.36%
pred. tinggi	1	3	69	94.52%
class recall	83.92%	66.92%	82.14%	

Gambar 11. Weighted Mean Recall Decision Tree

Berdasarkan hasil pengujian yang telah dilakukan, algoritma *Decision Tree (C4.5)* memiliki performa yang lebih baik dibandingkan *Naïve Bayes* pada seluruh metrik evaluasi. *Decision Tree* memperoleh nilai *accuracy*, *precision*, dan *recall* yang lebih tinggi sehingga dinilai lebih optimal dalam mengklasifikasikan tingkat keterampilan produk pada data penjualan Petshop Bay Bay.

Table 4. Perbandingan Hasil Evaluasi Algoritma

Algoritma	Accuracy	Weighted Mean Precision	Weighted Mean Recall
Naïve Bayes	55.28%	56.89%	57.49%
Decision Tree (C4.5)	77.22%	80.96%	77.63%

3.2 Pembahasan

Hasil pengujian menunjukkan bahwa algoritma *Naïve Bayes* mampu melakukan klasifikasi data penjualan dengan kategori cukup. Namun, performa model masih belum optimal karena beberapa atribut pada data penjualan memiliki hubungan satu sama lain, sedangkan *Naïve Bayes* mengasumsikan setiap atribut bersifat independen. Akibatnya, masih terdapat kesalahan klasifikasi pada beberapa kategori produk sehingga mempengaruhi nilai *accuracy*, *precision*, dan *recall* yang diperoleh.

Interpretasi hasil dilakukan untuk menjelaskan perbedaan kinerja algoritma *Naïve Bayes* dan *Decision Tree (C4.5)* berdasarkan tiga metrik evaluasi, yaitu *accuracy*, *weighted mean precision*, dan *weighted mean recall*. Interpretasi ini bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai kemampuan masing-masing algoritma dalam melakukan klasifikasi terhadap data yang digunakan dalam penelitian.

Berdasarkan hasil evaluasi menggunakan metrik *accuracy*, algoritma *Naïve Bayes* menghasilkan nilai *accuracy* sebesar 55,28%, sedangkan algoritma *Decision Tree (C4.5)* memperoleh nilai *accuracy* sebesar 77,22%. Perbedaan nilai ini menunjukkan bahwa *Decision Tree (C4.5)* memiliki tingkat ketepatan klasifikasi yang lebih tinggi dibandingkan dengan *Naïve Bayes*. Nilai *accuracy* yang lebih rendah pada algoritma *Naïve Bayes* disebabkan oleh karakteristik algoritma tersebut yang mengasumsikan bahwa setiap atribut bersifat saling independen. Pada data penjualan, atribut seperti kategori produk, harga, dan periode penjualan memiliki keterkaitan satu sama lain. Ketika asumsi independensi ini tidak terpenuhi, kemampuan *Naïve Bayes* dalam mengenali pola data menjadi terbatas, sehingga meningkatkan jumlah kesalahan klasifikasi dan menurunkan nilai *accuracy* secara keseluruhan.

Sementara itu, algoritma *Decision Tree (C4.5)* mampu menghasilkan nilai *accuracy* yang lebih tinggi karena dapat memodelkan hubungan antar atribut secara langsung melalui pembentukan aturan keputusan. Algoritma ini memilih atribut dengan nilai *information gain* tertinggi sebagai dasar pemisahan data, sehingga mampu menangkap pola keterjualan produk secara lebih efektif. Dengan struktur pohon keputusan, *Decision Tree* dapat menyesuaikan diri terhadap karakteristik data penjualan yang bersifat kompleks dan tidak independen, sehingga menghasilkan klasifikasi yang lebih akurat.

Berdasarkan hasil evaluasi menggunakan metrik *Weighted Mean Precision*, algoritma *Naïve Bayes* menghasilkan nilai sebesar 56,89%, sedangkan algoritma *Decision Tree (C4.5)* memperoleh nilai *Weighted Mean Precision* sebesar 80,96%. Nilai *precision* yang lebih tinggi pada *Decision Tree* menunjukkan bahwa algoritma ini lebih mampu meminimalkan kesalahan prediksi kelas, sehingga hasil klasifikasi yang dihasilkan menjadi lebih tepat. Sebaliknya, rendahnya nilai *precision* pada *Naïve Bayes* menunjukkan masih adanya kesalahan prediksi kelas akibat keterbatasan model dalam merepresentasikan hubungan antar atribut.

Berdasarkan hasil evaluasi menggunakan metrik *Weighted Mean Recall*, algoritma *Naïve Bayes* menghasilkan nilai sebesar 57,49%, sedangkan algoritma *Decision Tree (C4.5)* memperoleh nilai *Weighted Mean Recall* sebesar 77,63%. Nilai *recall* yang lebih tinggi pada *Decision Tree* menunjukkan bahwa algoritma ini lebih efektif dalam mengenali data aktual dari setiap kelas, meskipun jumlah data antar kelas tidak sama. Hal ini disebabkan karena *Decision Tree* mampu membagi data ke dalam kelompok yang lebih representatif melalui struktur pohon keputusan dan pemilihan atribut yang paling berpengaruh.

Hasil evaluasi menunjukkan bahwa algoritma *Decision Tree (C4.5)* lebih efektif dalam mengenali data aktual pada setiap kelas secara seimbang. Keunggulan ini menunjukkan bahwa *Decision Tree (C4.5)* memiliki kemampuan yang lebih baik dalam mempelajari hubungan antar atribut dan membentuk aturan klasifikasi yang representatif terhadap karakteristik data penjualan. Secara keseluruhan, berdasarkan ketiga metrik evaluasi yang digunakan, algoritma *Decision Tree (C4.5)* menunjukkan performa yang lebih baik dibandingkan dengan *Naïve Bayes*, sehingga lebih sesuai digunakan pada data penelitian ini yang memiliki keterkaitan antar atribut dan kompleksitas pola yang cukup tinggi.

4. KESIMPULAN DAN SARAN

Algoritma *Decision Tree (C4.5)* dan *Naïve Bayes* berhasil diimplementasikan menggunakan perangkat lunak *RapidMiner* untuk proses klasifikasi produk terlaris di Petshop Bay Bay Sidoarjo. Implementasi dilakukan melalui tahapan pengolahan data, penentuan atribut dan kelas, serta pengujian model menggunakan metode *Cross Validation*. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree (C4.5)* memiliki performa yang lebih baik dibandingkan *Naïve Bayes* berdasarkan nilai *accuracy*, *weighted mean precision*, dan *weighted mean recall*. Selain memiliki tingkat akurasi yang lebih tinggi, *Decision Tree (C4.5)* juga lebih mudah dipahami karena menghasilkan struktur pohon keputusan yang jelas. Algoritma ini mampu menangani data kategorikal maupun numerik serta lebih toleran terhadap hubungan antaratribut yang kompleks. Penggunaan *RapidMiner* juga membantu proses analisis data menjadi lebih sistematis, efisien, dan mudah diinterpretasikan. Dengan demikian, *Decision Tree (C4.5)* dinilai lebih efektif digunakan dalam klasifikasi produk terlaris pada data penjualan Petshop Bay Bay Sidoarjo.

Berdasarkan hasil penelitian, pihak Petshop Bay Bay Sidoarjo disarankan menggunakan algoritma *Decision Tree (C4.5)* sebagai metode utama dalam analisis data penjualan. Penggunaan algoritma ini diharapkan mampu membantu proses pengambilan keputusan terkait strategi penjualan dan pengelolaan stok produk secara lebih tepat. Selain itu, penelitian selanjutnya disarankan untuk membandingkan *Decision Tree (C4.5)* dengan algoritma lain seperti *K-Nearest*

Neighbor (KNN), Support Vector Machine (SVM), atau Random Forest. Perbandingan tersebut bertujuan untuk mengetahui metode yang paling optimal dalam klasifikasi produk terlaris. Penelitian berikutnya juga dapat menggunakan *dataset* yang lebih besar dan lebih beragam agar hasil klasifikasi menjadi lebih akurat dan representatif. Dengan adanya pengembangan penelitian lebih lanjut, diharapkan metode klasifikasi yang digunakan dapat memberikan hasil yang lebih stabil dan sesuai dengan kebutuhan analisis data penjualan UMKM.

5. DAFTAR RUJUKAN

- [1] Aras, Z. and Sarjono., 2016. Analisis data mining untuk menentukan kelompok prioritas penerima bantuan bedah rumah menggunakan metode clustering K-Means (Studi Kasus: Kantor Kecamatan Bahar Utara). *Jurnal Manajemen Sistem Informasi*, 1(2), pp.159–170.
- [2] Sulistyio, G.B., Hasan, N., Kiswati, S., Natalia, F. and Muningsih, E., 2025. Segmentasi pelanggan dan optimalisasi penjualan pada data retail online berbasis model RFM. *Computer Network Technology*, 5(1), pp.16–22.
- [3] Wardani, N.W., Nugraha, P.G.S.C. and Mahendra, G.S., 2023. Implementasi Naïve Bayes pada data mining untuk mengklasifikasikan penjualan barang terlaris pada perusahaan ritel. *JST (Jurnal Sains dan Teknologi)*, 12(3), pp.656–668.
- [4] Hidayat, R.H., 2024. Perbandingan metode Naïve Bayes dan Decision Tree C4.5 untuk analisis sentimen produk Es Teh Indonesia di media sosial Twitter. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 7(2), pp.88–98.
- [5] Lubis, F.R., Harahap, M.K. and Husein, A.M., 2021. Analisis prediktif untuk keputusan bisnis: peramalan penjualan. pp.32–40.
- [6] Maranto, A.R.K., Damayanti, L. and Ramadika, I.R., 2024. Perbandingan algoritma C4.5 dengan Naïve Bayes untuk menduga loyalitas pelanggan pada perusahaan Internet Service Provider. *bit-Tech*, 7(2), pp.396–405.
- [7] Ainurrohman., 2021. Akurasi algoritma klasifikasi pada software RapidMiner dan Weka. *Prisma Prosiding Seminar Nasional Matematika*, 4, pp.493–499.
- [8] Faisal, H.V. and Dhika, H., 2021. Penerapan algoritma Decision Tree dalam penjualan handphone. *Jurnal Informatika dan Sistem Informasi*, 2(2), pp.413–424.
- [9] Bamba, O.T., Vadila, N., Fitrawati, S., Tedang, V.W. and Asrawati., 2025. Naive Bayes dan Decision Tree: studi kasus klasifikasi kepuasan pelanggan e-commerce. *Simkom*, 10(2), pp.254–262.
- [10] Eryana, H.G.D.A., Deivivi, F., Baehaqi, M.I., Wicaksono, A. and Puspita, Z.A., Studi komparatif Naive Bayes dan Decision Tree pada dataset bunga iris: evaluasi akurasi dan efisiensi. *Jurnal Rekayasa Informasi Swadharma*, 53, p.160.
- [11] Hapsari, A.T. and Fauzani, M.M., 2025. Implementasi data mining untuk prediksi minat pembelian produk roti menggunakan metode Apriori pada Toko Roti Dapur Bunda. *Jurnal Ekonomi dan Pembangunan Indonesia*, 3(1), pp.290–300.