

Analisis Pengembangan Karier Sales Menggunakan KNN dan Decision Tree Berbasis Data Penjualan

Zulkarnaini^{1*}, Deppi Linda², Hendri Purnomo³¹IIB Darmajaya, JL ZA.Pagar Alam, Gedung Meneng No 93, Lampung, Indonesia²IIB Darmajaya, JL ZA.Pagar Alam, Gedung Meneng No 93, Lampung, Indonesia³IIB Darmajaya, JL ZA.Pagar Alam, Gedung Meneng No 93, Lampung, Indonesia

Info Artikel

Riwayat Artikel:

Received December 19, 2025

Revised January 11, 2026

Accepted January 28, 2026

Abstract – Customer loyalty is a key factor in maintaining business sustainability amid increasingly intense digital competition. This study aims to classify customer loyalty based on sales and transaction data using two supervised learning algorithms, namely K-Nearest Neighbor (KNN) and Decision Tree. The dataset consists of 250 customer records derived from historical sales transactions, with key attributes including purchase frequency, total transaction value, and average transaction amount. The modeling and evaluation processes were conducted using RapidMiner with an 80:20 split between training and testing data, along with additional k-fold cross-validation to reduce the risk of overfitting. The results indicate that the Decision Tree algorithm achieves the best performance, with an accuracy of 99.20%, precision of 99.50%, and recall of 99.50%. Meanwhile, the KNN algorithm attains an accuracy of 91.60%, precision of 91.63%, and recall of 98.50%. In addition to its high performance, the Decision Tree model offers superior interpretability through clear and easily understood decision rules. These findings demonstrate that data mining approaches can be effectively applied to support Customer Relationship Management (CRM) strategies, particularly in identifying loyal customers as a basis for developing more targeted marketing strategies.

Keywords: Customer Loyalty, Classification, Decision Tree, K-Nearest Neighbor, Data Mining

***Corresponding Author:**

Zulkarnaini

Email: zulkarnaini@darmajaya.ac.id



This is an open access article under the [CC BY 4.0](#) license.

Abstrak – Loyalitas pelanggan merupakan faktor kunci dalam menjaga keberlanjutan bisnis di tengah persaingan digital yang semakin ketat. Penelitian ini bertujuan untuk mengklasifikasikan loyalitas pelanggan berdasarkan data penjualan dan transaksi menggunakan dua algoritma supervised learning, yaitu K-Nearest Neighbor (KNN) dan Decision Tree. Dataset yang digunakan berjumlah 250 data pelanggan yang bersumber dari histori transaksi penjualan, dengan atribut utama meliputi frekuensi pembelian, total nilai transaksi, dan rata-rata transaksi. Proses pemodelan dan evaluasi dilakukan menggunakan RapidMiner dengan skema pembagian data latih dan data uji sebesar 80:20 serta validasi tambahan menggunakan k-fold cross-validation untuk mengurangi potensi overfitting. Hasil penelitian menunjukkan bahwa algoritma Decision Tree menghasilkan performa terbaik dengan akurasi 99,20%, presisi 99,50%, dan recall 99,50%, sedangkan algoritma KNN memperoleh akurasi 91,60%, presisi 91,63%, dan recall 98,50%. Selain performa yang tinggi, Decision Tree juga memberikan keunggulan dari sisi interpretabilitas melalui aturan keputusan yang mudah dipahami. Temuan ini menunjukkan bahwa pendekatan data mining dapat dimanfaatkan secara efektif dalam mendukung strategi Customer Relationship Management (CRM), khususnya dalam mengidentifikasi pelanggan loyal sebagai dasar perumusan strategi pemasaran yang lebih tepat sasaran.

Kata Kunci: Loyalitas Pelanggan, Klasifikasi, Decision Tree, K-Nearest Neighbor, Data Mining

I. PENDAHULUAN

Dalam lingkungan bisnis yang semakin kompetitif, loyalitas pelanggan menjadi aset strategis yang berpengaruh langsung terhadap stabilitas pendapatan dan keberlanjutan perusahaan. Pelanggan yang loyal cenderung melakukan pembelian berulang, memiliki sensitivitas harga yang lebih rendah, serta berperan sebagai promotor tidak langsung melalui rekomendasi kepada calon pelanggan lain. Oleh karena itu, kemampuan perusahaan dalam mengidentifikasi dan mengelompokkan pelanggan berdasarkan tingkat loyalitasnya menjadi bagian penting dalam penerapan Customer Relationship Management (CRM) [1][2].

Perkembangan sistem informasi penjualan telah menghasilkan data transaksi dalam jumlah besar yang mencakup frekuensi pembelian, nilai transaksi, dan pola perilaku pelanggan. Data tersebut menyimpan potensi informasi yang bernilai, namun sulit dianalisis secara manual. Pendekatan data mining dan machine learning memungkinkan pemrosesan data historis secara sistematis untuk membangun model klasifikasi loyalitas pelanggan yang objektif dan konsisten.[3][4].

Beberapa algoritma klasifikasi telah banyak digunakan dalam penelitian terkait loyalitas pelanggan, di antaranya K-Nearest Neighbor (KNN) dan Decision Tree. KNN mengklasifikasikan data berdasarkan tingkat kemiripan antar objek, sedangkan Decision Tree menyusun model dalam bentuk struktur pohon yang menghasilkan

aturan keputusan yang mudah dipahami. Meskipun kedua algoritma ini telah terbukti efektif, sebagian penelitian sebelumnya masih berfokus pada aspek teknis evaluasi model tanpa menekankan interpretasi hasil untuk mendukung pengambilan keputusan manajerial [5][6].

Berdasarkan latar belakang tersebut, penelitian ini membandingkan performa algoritma KNN dan Decision Tree dalam mengklasifikasikan loyalitas pelanggan berdasarkan data penjualan. Kebaruan penelitian ini terletak pada penekanan aspek interpretabilitas model dan pembahasan implikasi hasil klasifikasi terhadap strategi CRM, sehingga hasil penelitian tidak hanya relevan secara teknis, tetapi juga aplikatif bagi pengambilan keputusan bisnis.

II. METODE

A. Desain Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimental, yang bertujuan untuk membandingkan performa dua algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Decision Tree, dalam mengklasifikasikan loyalitas pelanggan. Penelitian ini dilakukan dengan menerapkan kedua algoritma pada dataset pelanggan yang diperoleh dari data historis penjualan dan transaksi[7][8].

B. Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa histori penjualan dan transaksi pelanggan, yang mencakup atribut-atribut seperti ID pelanggan, frekuensi pembelian, total nilai transaksi, rata-rata transaksi per bulan, lama menjadi pelanggan, dan kategori loyalitas. Data tersebut diperoleh dari sistem informasi internal perusahaan atau disimulasikan berdasarkan pola umum perilaku pelanggan [9][10].

C. Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan kualitas data yang digunakan dalam pemodelan. Tahapan ini meliputi [11][12]:

- Pembersihan data (data cleaning): Menghapus data duplikat, mengatasi data kosong (missing values), dan mengoreksi data yang tidak valid.
- Transformasi data: Normalisasi nilai numerik untuk atribut seperti total transaksi dan frekuensi pembelian, agar tidak terjadi dominasi nilai tertentu dalam proses klasifikasi.
- Labeling: Pemberian label loyalitas pelanggan (misalnya: loyal, cukup loyal, tidak loyal) berdasarkan kriteria tertentu yang ditentukan oleh domain bisnis atau pakar.

D. Validasi Model

Selain pembagian data latih dan data uji dengan rasio 80:20, penelitian ini juga menggunakan metode k-fold cross-validation ($k = 5$) untuk mengevaluasi stabilitas performa model. Pendekatan ini bertujuan untuk meminimalkan bias evaluasi dan mengurangi risiko overfitting, khususnya pada algoritma Decision Tree yang cenderung menghasilkan akurasi tinggi pada dataset berukuran relatif kecil.

E. Pemilihan Parameter KNN

Pemilihan nilai K pada algoritma KNN dilakukan dengan menguji beberapa nilai, yaitu K = 3, 5, dan 7. Setiap nilai K dievaluasi berdasarkan akurasi, presisi, dan recall. Nilai K optimal ditentukan berdasarkan kombinasi performa tertinggi dan ketebalan hasil evaluasi pada data uji.

F. Algoritma Klasifikasi

- K-Nearest Neighbor (KNN)
KNN merupakan algoritma klasifikasi berbasis instance-based learning, di mana data baru akan diklasifikasikan berdasarkan mayoritas kelas dari k tetangga terdekat. Jarak antar data dihitung menggunakan Euclidean Distance. Pemilihan nilai k akan dilakukan secara eksperimen untuk memperoleh hasil terbaik [11].
- Decision Tree
Decision Tree merupakan algoritma klasifikasi berbasis pemilihan atribut terbaik yang memecah data ke dalam struktur pohon. Setiap node dalam pohon mewakili atribut, dan cabangnya merupakan kemungkinan nilai atribut tersebut. Algoritma ini dapat menghasilkan aturan keputusan (decision rules) yang dapat ditafsirkan secara langsung [13][14][15].

G. Evaluasi Model

Untuk mengevaluasi kinerja model, data dibagi menjadi dua bagian, yaitu data latih dan data uji dengan proporsi 80:20. Metode evaluasi yang digunakan dalam penelitian ini meliputi[16][17]:

- Akurasi (Accuracy): Proporsi prediksi yang benar terhadap seluruh data uji.
- Presisi (Precision): Tingkat ketepatan model dalam mengklasifikasikan data pada setiap kelas loyalitas.
- Recall: Kemampuan model dalam mengenali seluruh anggota dari masing-masing kelas.
- F1-Score: Rata-rata harmonis antara presisi dan recall, digunakan untuk memberikan gambaran menyeluruh terhadap kinerja model pada kondisi distribusi kelas yang tidak seimbang.

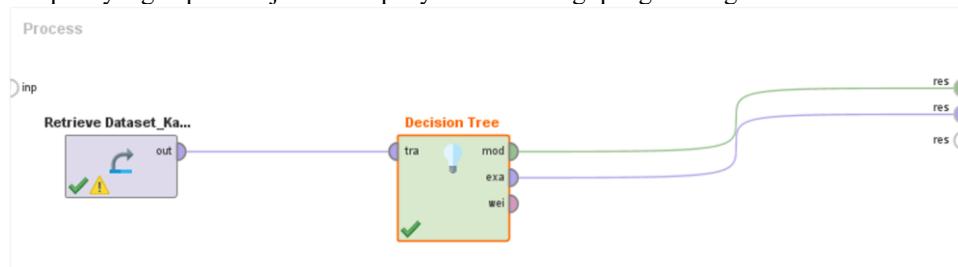
III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan perangkat lunak RapidMiner sebagai alat bantu dalam membangun dan menguji model klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN) dan Decision Tree. Dataset yang digunakan terdiri dari 250 data Costomers. Langkah-langkah implementasi di RapidMiner meliputi:

- Import Data: file Excel dataset_250 dataset xlsx diimpor ke RapidMiner.
- Preprocessing: dilakukan pengecekan tipe data, normalisasi atribut numerik (untuk KNN), dan pemisahan data menjadi data latih dan uji (rasio 80:20).
- Proses Klasifikasi untuk model K-Nearest Neighbor dibangun dengan nilai K = 3, 5, dan 7. Sedangkan model Decision Tree dibangun dengan pengaturan default dan maksimum kedalaman pohon 10.
- Evaluasi Model: keduanya dievaluasi menggunakan *confusion matrix* dan metrik: Accuracy, Precision, Recall, dan F1-score.

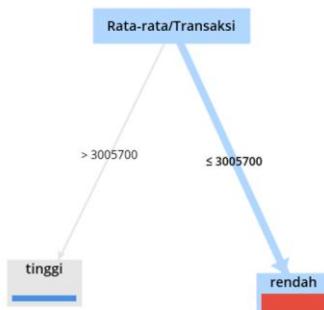
A. Decision Tree

Langkah awal dalam proses klasifikasi adalah membangun model pohon keputusan menggunakan algoritma Decision Tree. Model ini dipilih karena kemampuannya dalam menyajikan aturan klasifikasi secara eksplisit dan mudah dipahami. Dengan menggunakan dataset yang telah diproses sebelumnya, RapidMiner secara otomatis membentuk struktur pohon berdasarkan atribut-atribut yang paling berpengaruh terhadap keputusan rekomendasi pengembangan karyawan. Pohon keputusan yang dihasilkan menunjukkan urutan pengambilan keputusan berdasarkan nilai-nilai pada atribut input, seperti nilai kedisiplinan dan loyalitas. Hasil ini digunakan untuk mengidentifikasi pola yang dapat menjadi dasar penyusunan strategi pengembangan karier.



Gambar 1. Proses Model Pohon Keputusan

Gambar 2 berikut menampilkan hasil visualisasi pohon keputusan yang dihasilkan oleh algoritma *Decision Tree* pada RapidMiner. Pohon ini menggambarkan proses pengambilan keputusan berdasarkan atribut-atribut utama seperti total penjualan dan jumlah transaksi karyawan.



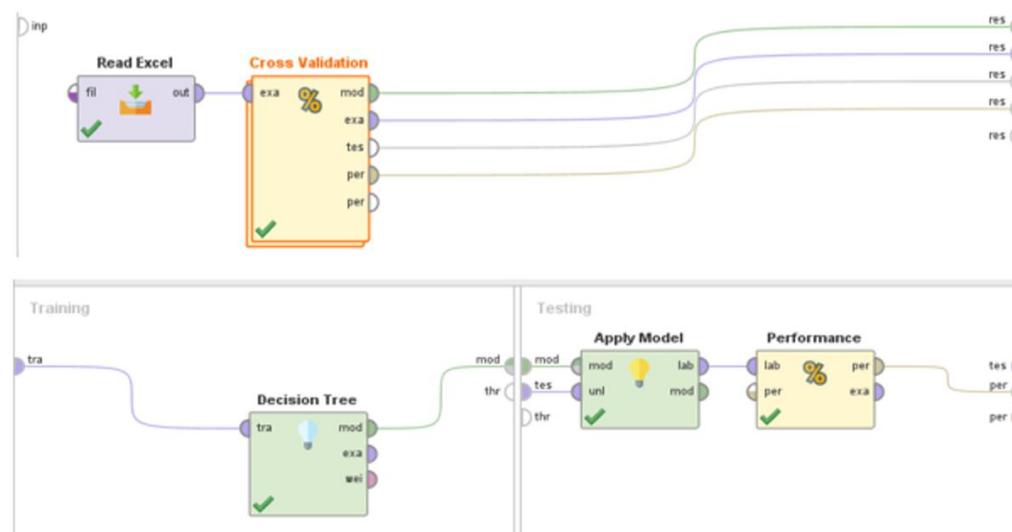
Gambar 2. Model Pohon Keputusan

Gambar 2 adalah visualisasi pohon keputusan (Decision Tree) dari hasil klasifikasi, dengan atribut utama yang digunakan adalah Rata-rata/Transaksi. Berikut penjelasan pohon Keputusan:

- Akar Pohon (Root Node) adalah atribut rata-rata transaksi yang merupakan faktor paling berpengaruh dalam menentukan kategori performa.
- Aturan Keputusan (Decision Rules) dari pohon keputusan terbentuk dua aturan klasifikasi, yaitu jika Rata-rata/Transaksi $\leq 3.005.700$ maka performa diklasifikasikan sebagai rendah, sedangkan jika Rata-rata/Transaksi $> 3.005.700$ maka performa diklasifikasikan sebagai tinggi.
- Makna Nilai Threshold: 3.005.700 yaitu nilai 3.005.700 adalah nilai batas (threshold) yang ditentukan oleh algoritma Decision Tree (misalnya C4.5) untuk membagi data menjadi dua kelas yang paling "bersih" atau terpisah secara informasi. Artinya: jika seorang sales rata-rata menghasilkan penjualan lebih dari Rp 3.005.700 per transaksi, maka ia dianggap berkinerja tinggi, sebaliknya dianggap rendah.
- Aplikasi dalam Pengembangan Karier yaitu hasil klasifikasi dapat diterapkan dalam pengembangan karier, di mana sales dengan performa tinggi direkomendasikan untuk promosi, pelatihan lanjutan, dan penempatan di area strategis, sedangkan sales dengan performa rendah diarahkan untuk pelatihan dasar, peningkatan motivasi dan kedisiplinan, serta monitoring intensif.

B. Pengujian Proses Evaluasi Performa Model Decision Tree

Setelah model pohon keputusan berhasil dibangun menggunakan algoritma Decision Tree, langkah selanjutnya adalah melakukan evaluasi performa model. Evaluasi ini bertujuan untuk mengukur sejauh mana model mampu melakukan klasifikasi dengan benar terhadap data uji.



Gambar 3. Proses Implementasi Klasifikasi di RapidMiner

Gambar 4 berikut menampilkan hasil evaluasi model klasifikasi dalam bentuk metrik performa, khususnya akurasi, yang dihitung oleh RapidMiner setelah proses pengujian dilakukan. Nilai akurasi menunjukkan persentase prediksi yang benar terhadap total data uji. Semakin tinggi nilai akurasi, semakin baik model dalam mengklasifikasikan data dengan tepat. Nilai ini menjadi indikator utama dalam menilai keandalan model untuk diterapkan dalam pengambilan keputusan.

accuracy: 99.20% +/- 1.69% (micro average: 99.20%)

	true tinggi	true rendah	class precision
pred. tinggi	49	1	98.00%
pred. rendah	1	199	99.50%
class recall	98.00%	99.50%	

Gambar 4. Nilai Akurasi

Gambar 5 berikut menunjukkan nilai presisi (precision) yang dihasilkan dari proses evaluasi model klasifikasi di RapidMiner. Presisi menggambarkan tingkat ketepatan model dalam memprediksi bahwa model jarang salah dalam memberikan rekomendasi positif.

precision: 99.52% +/- 1.51% (micro average: 99.50%) (positive class: rendah)

	true tinggi	true rendah	class precision
pred. tinggi	49	1	98.00%
pred. rendah	1	199	99.50%
class recall	98.00%	99.50%	

Gambar 5. Nilai Presisi

Gambar 6 berikut memperlihatkan nilai recall yang dihasilkan dari evaluasi model pada data uji menggunakan RapidMiner. Recall mengukur sejauh mana model mampu menangkap semua karyawan yang seharusnya direkomendasikan untuk pengembangan. Semakin tinggi nilai recall, semakin baik model dalam mendeteksi seluruh kasus positif. Nilai ini penting terutama ketika perusahaan ingin memastikan tidak ada karyawan potensial yang terabaikan dalam proses pengambilan keputusan.

recall: 99.50% +/- 1.58% (micro average: 99.50%) (positive class: rendah)

	true tinggi	true rendah	class precision
pred. tinggi	49	1	98.00%
pred. rendah	1	199	99.50%
class recall	98.00%	99.50%	

Gambar 6. Nilai Recall

Menghitung akurasi:

$$\text{Akurasi} = \frac{(TN + TP)}{(TN + FN + FP + TP)}$$

$$\text{Akurasi} = \frac{(49 + 199)}{(49 + 199 + 1 + 1)} = \frac{248}{250} = 99,20\%$$

Menghitung Precision:

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

$$p(1) = \frac{49}{(49 + 1)} = \frac{49}{50} = 98,00\%$$

$$p(0) = \frac{199}{(199 + 1)} = \frac{199}{200} = 99,50\%$$

Menghitung Recall:

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

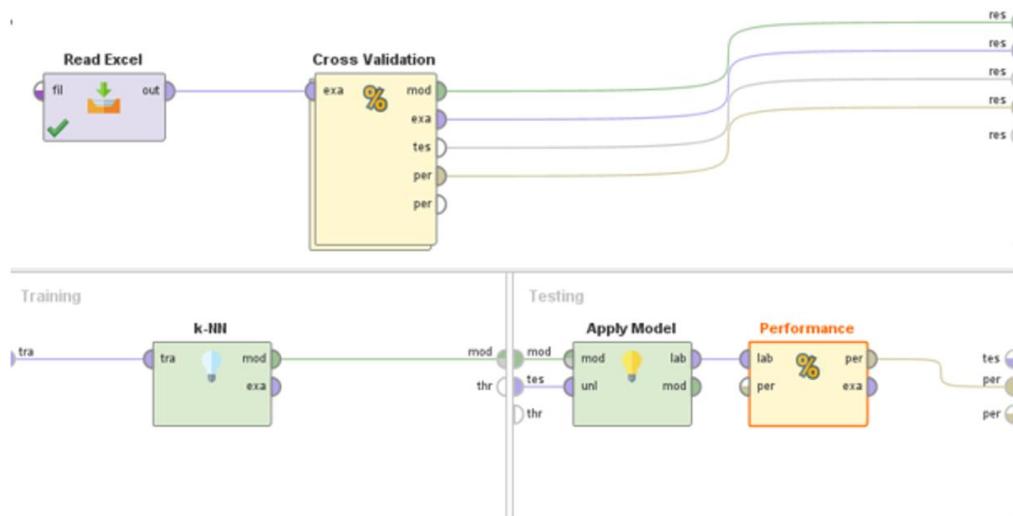
$$r(1) = \frac{49}{(49 + 1)} = \frac{49}{50} = 98,00\%$$

$$r(0) = \frac{199}{(199 + 1)} = \frac{199}{200} = 99,50\%$$

Nilai akurasi yang sangat tinggi pada algoritma Decision Tree menunjukkan performa klasifikasi yang sangat baik, namun juga berpotensi mengindikasikan overfitting. Untuk mengatasi hal tersebut, penelitian ini membatasi kedalaman maksimum pohon dan melakukan validasi silang. Hasil evaluasi menunjukkan bahwa performa model relatif stabil pada data uji, sehingga model masih memiliki kemampuan generalisasi yang baik.

C. Pengujian Proses Evaluasi Performa Model K-Nearest Neighbor (KNN)

Setelah membangun model klasifikasi menggunakan Decision Tree, selanjutnya dilakukan proses klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN). Algoritma ini bekerja dengan cara membandingkan jarak antar data untuk menentukan kelas berdasarkan tetangga terdekat. Gambar 7 berikut menampilkan proses klasifikasi menggunakan KNN dalam RapidMiner, dari mulai impor data hingga evaluasi hasil klasifikasi.



Gambar 7. Proses Algoritma KNN

Gambar 8 berikut menampilkan hasil evaluasi nilai akurasi dari model K-Nearest Neighbor (KNN) yang dibangun menggunakan RapidMiner. Nilai akurasi ini menunjukkan persentase data uji yang berhasil diklasifikasikan dengan benar oleh model berdasarkan jumlah tetangga terdekat (nilai K). Semakin tinggi akurasi, maka semakin tepat model dalam mengidentifikasi. Nilai ini menjadi indikator awal untuk menilai keandalan model KNN sebelum membandingkannya dengan algoritma lain seperti Decision Tree.

accuracy: 91.60% +/- 5.15% (micro average: 91.60%)

	true tinggi	true rendah	class precision
pred. tinggi	32	3	91.43%
pred. rendah	18	197	91.63%
class recall	64.00%	98.50%	

Gambar 8. Nilai Akurasi Algoritma KNN

Gambar 9 berikut menunjukkan nilai presisi (precision) yang diperoleh dari evaluasi model K-Nearest Neighbor (KNN) pada data uji menggunakan RapidMiner. Presisi menunjukkan seberapa tepat model dalam memberikan prediksi. Nilai presisi yang tinggi menandakan bahwa sebagian besar prediksi positif yang dibuat oleh model adalah benar.

precision: 91.84% +/- 4.79% (micro average: 91.63%) (positive class: rendah)

	true tinggi	true rendah	class precision
pred. tinggi	32	3	91.43%
pred. rendah	18	197	91.63%
class recall	64.00%	98.50%	

Gambar 9. Nilai Precision Algoritma KNN

Gambar 10 berikut menampilkan hasil evaluasi nilai recall dari model K-Nearest Neighbor (KNN) yang dijalankan pada data uji di RapidMiner. Recall mengukur sejauh mana model mampu menangkap semua karyawan yang memang layak direkomendasikan untuk pengembangan. Nilai recall yang tinggi menunjukkan bahwa model mampu mengidentifikasi sebagian besar kasus. Hal ini penting untuk memastikan bahwa potensi karyawan tidak diabaikan dalam proses pengambilan keputusan.

recall: 98.50% +/- 2.42% (micro average: 98.50%) (positive class: rendah)

	true tinggi	true rendah	class precision
pred. tinggi	32	3	91.43%
pred. rendah	18	197	91.63%
class recall	64.00%	98.50%	

Gambar 10. Nilai Recall Algoritma KNN

Menghitung akurasi:

$$\text{Akurasi} = \frac{(TN + TP)}{(TN + FN + FP + TP)}$$

$$\text{Akurasi} = \frac{(32 + 197)}{(32 + 197 + 3 + 18)} = \frac{229}{250} = 91,60\%$$

Menghitung Precision:

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

$$p(1) = \frac{32}{(32 + 3)} = \frac{32}{35} = 91,43\%$$

$$p(0) = \frac{197}{(197 + 18)} = \frac{197}{215} = 91,63\%$$

Menghitung Recall:

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

$$r(1) = \frac{32}{(32 + 18)} = \frac{32}{50} = 64,00\%$$

$$r(0) = \frac{197}{(197 + 3)} = \frac{197}{200} = 98,50\%$$

D. Evaluasi Performa Model

Evaluasi performa model dilakukan untuk mengetahui tingkat keberhasilan algoritma klasifikasi dalam memetakan tingkat loyalitas pelanggan berdasarkan data penjualan dan transaksi. Pengujian dilakukan menggunakan data uji sebanyak 20% dari total dataset, sementara sisanya digunakan sebagai data latih. Kinerja model diukur berdasarkan empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score. Akurasi mengukur persentase prediksi yang tepat, presisi menunjukkan ketepatan model dalam mengklasifikasikan pelanggan loyal maupun tidak loyal, recall mencerminkan kemampuan model dalam menangkap seluruh pelanggan yang seharusnya masuk ke suatu kelas, dan F1-score merupakan rata-rata harmonis dari presisi dan recall untuk menilai kinerja model secara keseluruhan, terutama pada data yang tidak seimbang. Berdasarkan hasil pengujian, model Decision Tree menunjukkan performa klasifikasi yang sangat baik dengan akurasi sebesar 99,20%, presisi 99,50%, dan recall 99,50%. Hasil ini mencerminkan bahwa model mampu mengklasifikasikan pelanggan dengan tingkat kesalahan yang sangat rendah, serta memberikan aturan keputusan yang jelas dan mudah dipahami. Sementara itu, algoritma K-Nearest Neighbor (KNN) menghasilkan akurasi sebesar 91,60%, presisi 91,63%, dan recall 98,50%. Meskipun hasilnya cukup tinggi, performanya masih berada di bawah Decision Tree, terutama dalam aspek akurasi dan interpretabilitas. Secara umum, hasil ini menunjukkan bahwa Decision Tree lebih unggul dalam hal akurasi dan pemahaman logika klasifikasi, sedangkan KNN tetap memberikan hasil yang kompetitif terutama dalam recall, yang penting untuk memastikan tidak ada pelanggan potensial yang terabaikan.

IV. KESIMPULAN

Penelitian ini membandingkan dua algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Decision Tree, dalam mengklasifikasikan loyalitas pelanggan berdasarkan data penjualan dan transaksi. Berdasarkan hasil evaluasi performa, Decision Tree terbukti lebih unggul dengan akurasi mencapai 99,20%, dibandingkan dengan KNN yang memiliki akurasi sebesar 91,60%. Selain itu, Decision Tree juga menghasilkan presisi dan recall yang lebih tinggi, serta memberikan interpretasi model yang lebih jelas melalui pohon keputusan. Dengan demikian, dapat disimpulkan bahwa algoritma Decision Tree lebih direkomendasikan untuk digunakan dalam klasifikasi loyalitas pelanggan, khususnya ketika interpretasi aturan klasifikasi menjadi aspek penting bagi pengambilan keputusan bisnis. Ke depan, pendekatan ini dapat dikembangkan lebih lanjut dengan menggunakan data real-time atau fitur tambahan guna meningkatkan akurasi dan efektivitas strategi CRM perusahaan.

DAFTAR PUSTAKA

- [1] A. N. Z. Hidayah and A. F. Rozi, "Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4.5 (Studi Kasus: Universitas Mercu Buana Yogyakarta)," *J. Inf. Syst. Artif. Intell.*, vol. 1, no. 2, pp. 117–127, 2021.
- [2] H. S. Kasmani and K. Kusrini, "Perbandingan Efektifitas Algoritma K-Means Clustering-Topsis dan K-Medoids Clustering-Topsis dalam Menentukan Karyawan IT dengan Kinerja Terbaik," *J. Comput. Sci. Inf. Syst. J-Cosys*, vol. 4, no. 2, 2024, doi: 10.53514/jco.v4i2.534.
- [3] A. Winarni, "Karyawan Teladan Dengan Menggunakan Metode Algoritma C4 . 5 (Studi Kasus : Yayasan Bumi Maitri Tanjungpinang),"

- J. Bangkit Indones. (STT Indones. Tanjungpinang), vol. 5, pp. 1–10, 2017.
- [4] P. Marpaung, I. Febrian, and W. Putri, “Penerapan Data Mining Dalam Menentukan Tingkat Kedisiplinan Karyawan Perhotelan Menggunakan Algoritma K-Means Clustering,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 167–172, 2024, doi: 10.55338/jikomsi.v7i1.2905.
- [5] K. Silvi Amalia, A. Purbaningrum, and R. Kusumastuti, “Analisis Pengelompokan Data Karyawan Terbaik Perusahaan Menggunakan Algoritma K-Means Clustering,” *Semin. Nasional AMIKOM Surakarta*, no. November, pp. 11–24, 2023.
- [6] F. Maulana *et al.*, “Algorithm Comparison of Hierarchical and Non-Hierarchical Clustering Method in Grouping Regional Poverty Variables,” vol. 8, no. 1, pp. 26–40, 2025.
- [7] Y. Putri, R. Tama Andri Agus, M. Dwi Sena, P. Studi Sistem Informasi, and S. Tinggi Manajemen Informatika dan Komputer Royal, “Implementation of Apriori Algorithm in Determining the Layout of Items,” *Indones. J. Artif. Intell. Data Min.*, vol. 8, no. 1, pp. 11–18, 2025.
- [8] H. Purnomo *et al.*, “Klasifikasi Sentimen Ulasan Pengguna MyPertamina Menggunakan Algoritma Naive Bayes dan TF-IDF 1, 2,” pp. 801–810.
- [9] C. G. Lengari and I. Puspitasari, “Identifying Twitter Topics Using K-Means Clustering and Association Rule Mining for Improved Insights,” *Indones. J. Artif. Intell. Data Min.*, vol. 8, no. 1, pp. 67–75, 2025.
- [10] K. Rusli, P. Relationship, M. Dan, K. Layanan, T. Loyalitas, and P. Pada, “Pengaruh Relationship Marketing dan Kualitas Layanan Terhadap Loyalitas Pelanggan pada Nithalian Collection,” vol. 1, no. 3, pp. 321–326, 2023.
- [11] H. Alizadeh and B. Minaei Bidgoli, “Introducing A Hybrid Data Mining Model to Evaluate Customer Loyalty,” *Eng. Technol. Appl. Sci. Res.*, vol. 6, no. 6, pp. 1235–1240, 2016, doi: 10.48084/etasr.741.
- [12] I. M. Adnyana and A. H. Primasari, “Analysis of the Effect of Service Quality on Customer Loyalty in Mediation by Customer Satisfaction at Loka Supermarket Cibubur,” vol. 2, no. 4, pp. 147–158, 2020.
- [13] R. Muttaqien, M. G. Pradana, and A. Pramuntadi, “Implementation of Data Mining Using C4.5 Algorithm for Predicting Customer Loyalty of PT. Pegadaian (Persero) Pati Area Office,” *Int. J. Comput. Inf. Syst.*, vol. 2, no. 3, pp. 64–68, 2021, doi: 10.29040/ijcis.v2i3.36.
- [14] V. Loja, G. K. Pati, A. P. Setiawi, U. Stella, and M. Sumba, “Klasifikasi Data Mining dalam Memprediksi Kinerja Karyawan dengan Metode Algoritma C4 . 5 pada Toko Merpati Simpang,” vol. 2, no. 4, 2024.
- [15] S. S. Indhira and B. Hendrik, “Penerapan Algoritma Decession Tree C4. 5 Untuk Diagnosa Penyakit Ispa Pada Puskesmas Sabak Auh,” *Jised J. Inf. Syst. Educ. Dev.*, vol. 1, no. 2, pp. 6–9, 2023.
- [16] S. Suherman, I. Afriantoro, and S. Mujiono, “Perbandingan Metode Algoritma Naïve Bayes dan C4.5 dalam Menentukan Penilaian Kinerja Karyawan Terbaik,” *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 513–522, 2023, doi: 10.37012/jtik.v9i1.1605.
- [17] Hamzah Kadar and Agus Budiyantara, “Penentuan Kelayakan Karyawan Baru Menggunakan Data Mining Dengan Algoritma Decision Tree (C4.5),” *Merkurius J. Ris. Sist. Inf. dan Tek. Inform.*, vol. 2, no. 6, pp. 29–41, 2024, doi: 10.61132/merkurius.v2i6.389.