Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 2 (2026) pp: 18693-18703
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Penentuan Outlet Potensial PT XYZ dengan Menggunakan Metode *Decision Tree*

Muhammad Firza[1], Sari Hartini[2]
[1]Program Studi Teknik Informatika, Universitas Bina Sarana Informatika
[2]Program Studi Sistem Informasi, Universitas Bina Sarana Informatika
[1]muhammadfirza389@gmail.com*, [2]sari.shi@bsi.ac.id

Abstrak

Penelitian ini bertujuan untuk menentukan outlet potensial PT XYZ menggunakan metode Decision Tree sebagai pendekatan data mining untuk mendukung pengambilan keputusan dalam penyusunan strategi distribusi dan pemasaran yang lebih efektif. Penelitian menerapkan tahapan Knowledge Discovery in Database (KDD) yang meliputi seleksi data, pembersihan data, transformasi data, penambangan data, serta evaluasi hasil. Dataset yang digunakan terdiri atas 1.069 data outlet yang berasal dari wilayah Kabupaten Bekasi dan Karawang, kemudian dibagi menjadi 748 data pelatihan dan 321 data pengujian menggunakan rasio 70:30. Proses pemodelan dilakukan dengan perangkat lunak RapidMiner menggunakan algoritma Decision Tree untuk mengklasifikasikan outlet ke dalam lima kategori, yaitu Sangat Tidak Potensial, Cukup Potensial, Potensial, Prioritas Potensial, dan Sangat Potensial. Hasil penelitian menunjukkan bahwa atribut total penjualan merupakan faktor yang paling berpengaruh dalam pembentukan pohon keputusan. Evaluasi model menggunakan confusion matrix menghasilkan nilai akurasi, presisi, dan recall masing-masing sebesar 100% pada seluruh kelas, yang menunjukkan bahwa model mampu mengklasifikasikan data secara sangat baik berdasarkan dataset yang digunakan. Hasil penelitian ini membuktikan bahwa metode Decision Tree mampu menghasilkan model klasifikasi yang efektif dalam mengidentifikasi tingkat potensi outlet sehingga dapat digunakan sebagai dasar penentuan prioritas kunjungan tim penjualan, penyusunan strategi distribusi yang lebih tepat sasaran, serta mendukung proses pengambilan keputusan perusahaan secara lebih cepat, objektif, dan berbasis data.

Kata kunci: Decision Tree, Penentuan Outlet Potensial, PT XYZ, Outlet Potensial, Klasifikasi Outlet.

1. Latar Belakang

Perkembangan teknologi informasi di era digital telah menghasilkan volume data yang sangat besar pada berbagai sektor, termasuk penjualan dan distribusi. Peningkatan volume transaksi menuntut perusahaan mampu mengelola data tersebut menjadi informasi yang bernilai guna mendukung pengambilan keputusan strategis [1]. Pemanfaatan teknik klasifikasi berbasis data historis penjualan menjadi salah satu pendekatan yang banyak diadopsi perusahaan untuk memetakan outlet ke dalam kategori-kategori keputusan secara objektif dan terukur [2].

PT XYZ merupakan perusahaan nasional yang bergerak di bidang industri farmasi. Perusahaan ini menjalankan aktivitas distribusi melalui jaringan outlet yang tersebar di berbagai wilayah, termasuk Kabupaten Bekasi dan Karawang, dengan karakteristik dan tingkat penjualan yang berbeda-beda antar outlet. Penentuan prioritas kunjungan outlet selama ini melibatkan tim penjualan dan supervisor wilayah, namun tidak semua outlet memberikan kontribusi penjualan yang optimal, sementara proses analisis yang masih dilakukan secara manual menyebabkan keterlambatan pengambilan keputusan dan berpotensi menimbulkan kesalahan dalam penentuan strategi distribusi. Kondisi inilah yang mendasari perlunya penerapan metode klasifikasi *data mining*, khususnya *Decision Tree*, karena metode ini mampu menghasilkan model berbentuk pohon keputusan yang terstruktur dan mudah diinterpretasikan, sehingga penentuan outlet potensial dapat dilakukan secara lebih cepat, konsisten, dan objektif dibandingkan proses manual yang selama ini berjalan.

Penelitian-penelitian terdahulu telah membuktikan efektivitas *Decision Tree* pada berbagai domain klasifikasi, sebagaimana dirangkum pada Tabel 1, dengan tingkat akurasi yang bervariasi antara 50% hingga 99,64% bergantung pada karakteristik data dan atribut yang digunakan [3]–[6].

Tabel 1. Penelitian Terdahulu

No	Peneliti (Tahun)	Judul Penelitian	Metode	Akurasi
1	Aini dkk. (2025)	Penerapan Metode *Decision Tree* dalam Penentuan Jurusan Siswa	*Decision Tree*	86,59%
2	Zacksavira dkk. (2025)	Implementasi Algoritma *Decision Tree* untuk Deteksi Depresi pada Pelajar	*Decision Tree*	89%
3	Syafii dkk. (2024)	Analisis Prediksi Customer Repeat Order Menggunakan Algoritma *Decision Tree* pada Perusahaan Transportasi	*Decision Tree*	83,33%
4	Shafarindu dkk. (2021)	Klasifikasi Data Penjualan pada Supermarket dengan Metode *Decision Tree*	*Decision Tree*	50%
5	Suriani (2026)	Penerapan *Data Mining* untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma *Decision Tree* C4.5	*Decision Tree* C4.5	99,64%

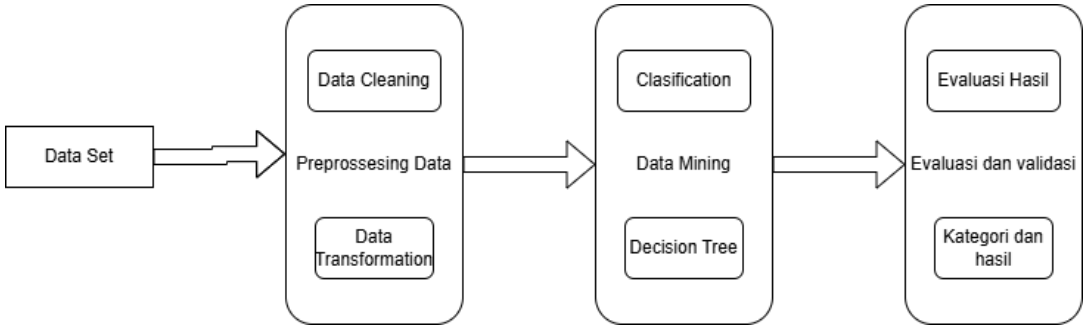
Sumber: diolah oleh penulis, 2026

Penelitian-penelitian tersebut umumnya berfokus pada bidang pendidikan, kesehatan mental, atau transportasi, serta hanya menghasilkan klasifikasi biner atau tiga kelas. Berbeda dari itu, penelitian ini secara khusus menerapkan *Decision Tree* pada sektor distribusi produk farmasi dan konsumen, dengan menghasilkan klasifikasi outlet yang lebih rinci ke dalam lima kategori, yaitu Sangat Tidak Potensial, Cukup Potensial, Potensial, Prioritas Potensial, dan Sangat Potensial. Kebaruan penelitian ini terletak pada kombinasi atribut data penjualan, jenis outlet, dan lokasi outlet sebagai dasar klasifikasi, sehingga hasilnya dapat langsung digunakan sebagai pedoman operasional bagi tim penjualan dalam menentukan prioritas kunjungan outlet dan penyusunan strategi distribusi PT XYZ, dibandingkan hanya membedakan outlet ke dalam kategori potensial atau tidak potensial secara sederhana sebagaimana pendekatan pada penelitian-penelitian terdahulu.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk: (1) menerapkan metode *Decision Tree* dalam mengklasifikasikan outlet potensial berdasarkan data penjualan PT XYZ; (2) mengidentifikasi atribut yang paling berpengaruh dalam penentuan kategori outlet; dan (3) mengukur tingkat akurasi model klasifikasi yang dihasilkan sebagai dasar pengambilan keputusan strategi distribusi perusahaan.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan data penjualan outlet PT XYZ yang bersifat numerik, diolah menggunakan teknik *data mining* agar menghasilkan kesimpulan yang objektif dan terukur [7]. Kerangka analisis mengikuti tahapan *Knowledge Discovery in Database* (KDD), yaitu suatu proses penambangan data untuk menemukan, memeriksa, dan mengidentifikasi pola pada kumpulan data yang telah melalui seleksi dan analisis [6], sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Tahapan *Knowledge Discovery in Database* (KDD)

2.1 Objek dan Sumber Data

Objek penelitian adalah data penjualan outlet PT XYZ, sebuah perusahaan farmasi dan produk konsumen dengan jaringan distribusi di wilayah Kabupaten Bekasi dan Karawang. Data yang digunakan berjumlah 1.069 catatan outlet dengan atribut sebagaimana disajikan pada Tabel 2.

DOI: https://doi.org/10.31004/riggs.v5i2.12294

Tabel 2. Atribut Dataset Penjualan Outlet PT XYZ

No	Variabel	Tipe	Keterangan
1	Total Penjualan	Integer	Nilai penjualan outlet
2	Lokasi Outlet	Kategorikal	Wilayah outlet
3	Jenis Outlet	Kategorikal	Tipe outlet
4	Nama Outlet	Real	Nama outlet
5	Outlet Potensial	Kategorikal	Label (target klasifikasi)
6	Kategori Outlet	Kategorikal	Kategori penjualan

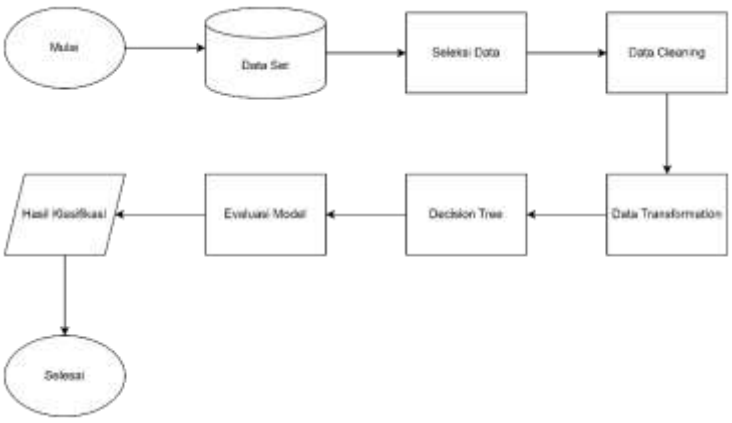
Sumber: diolah oleh penulis, 2026

Wilayah Kabupaten Bekasi dan Karawang dipilih sebagai lokasi penelitian karena merupakan salah satu kawasan distribusi utama PT XYZdengan kepadatan outlet yang cukup tinggi, mencakup berbagai jenis outlet seperti toko obat, apotek, minimarket, dan warung kelontong. Karakteristik outlet pada kedua wilayah ini bervariasi baik dari sisi lokasi, misalnya outlet yang berada di kawasan industri dengan mobilitas pekerja tinggi maupun outlet di kawasan permukiman padat penduduk, sehingga pola penjualan yang terbentuk juga beragam. Keberagaman karakteristik outlet inilah yang mendasari perlunya klasifikasi berbasis data untuk membantu tim penjualan mengidentifikasi outlet yang layak diprioritaskan tanpa harus melakukan survei manual pada seluruh outlet satu per satu.

Selain kepadatan outlet, wilayah Bekasi dan Karawang juga memiliki keunggulan berupa akses distribusi yang relatif mudah dijangkau dari pusat distribusi PT XYZ dibandingkan wilayah lain, sehingga hasil klasifikasi outlet pada penelitian ini berpotensi untuk diterapkan secara langsung dalam penyusunan rute kunjungan tim penjualan tanpa memerlukan penyesuaian logistik yang signifikan.

2.2 Tahapan Pengolahan Data

Tahap *data cleaning* dilakukan untuk mengatasi data yang tidak lengkap atau tidak konsisten, salah satunya melalui imputasi nilai yang hilang menggunakan nilai rata-rata [8]. Selanjutnya, tahap *data transformation* mengubah atribut total penjualan yang semula berupa angka menjadi bentuk kategorikal (rendah, sedang, tinggi) agar dapat diproses oleh algoritma *Decision Tree* [6]. Proses keseluruhan alur *data mining* yang diterapkan pada penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Alur Proses *Data Mining* pada PT XYZ

Setelah data siap, dataset dibagi menggunakan teknik *split data* dengan rasio 70:30 untuk memisahkan fase pelatihan dan pengujian model, sehingga evaluasi kinerja model menjadi lebih objektif dan representatif [9]. Dari 1.069 data outlet, sebanyak 748 data digunakan sebagai data pelatihan (*training*) untuk membangun model *Decision Tree*, sedangkan 321 data digunakan sebagai data pengujian (*testing*) untuk menilai kinerja model menggunakan RapidMiner.

DOI: https://doi.org/10.31004/riggs.v5i2.12294

Selain proses *data cleaning* dan *data transformation*, penelitian ini juga menerapkan teknik validasi untuk memastikan bahwa dataset yang telah diproses layak digunakan pada tahap pemodelan. Teknik pengumpulan dan pengujian data yang diterapkan meliputi validasi holdout, yaitu pembagian dataset menjadi subset pelatihan dan subset pengujian secara tetap, serta prinsip validasi silang (*cross validation*) sebagai bahan pertimbangan untuk penelitian lanjutan guna menguji stabilitas model pada subset data yang berbeda-beda [3]. Penerapan validasi holdout dengan rasio 70:30 pada penelitian ini dipilih karena kesederhanaannya dan kesesuaiannya dengan ukuran dataset yang tersedia, yaitu 1.069 data outlet, sehingga proses pelatihan dan pengujian model dapat dilakukan secara efisien tanpa mengorbankan representativitas data.

Selain pembagian data, tahap reduksi data turut dilakukan dengan menghilangkan atribut yang tidak relevan terhadap proses klasifikasi, seperti nama outlet yang bersifat unik untuk setiap catatan dan tidak memberikan pola yang dapat digeneralisasi. Dengan demikian, atribut yang digunakan dalam pembentukan model dibatasi pada total penjualan, lokasi outlet, jenis outlet, dan kategori outlet, sehingga model yang dihasilkan lebih ringkas dan fokus pada atribut yang benar-benar berpengaruh terhadap penentuan kategori outlet potensial [10].

2.3 Algoritma Decision Tree

Decision Tree merupakan representasi berbentuk diagram alir menyerupai pohon, dengan setiap simpul menunjukkan pengujian suatu atribut, cabang menunjukkan hasil pengujian, dan simpul daun menunjukkan kelas akhir dari data yang dianalisis [11]. Pembentukan pohon keputusan pada penelitian ini menggunakan algoritma C4.5, yang memiliki keunggulan mampu menangani data tidak lengkap (*missing value*), mengolah data numerik maupun kategorikal, serta tetap memberikan hasil yang baik meskipun terdapat *noise* pada data [12]. Setiap pemilihan atribut pemecah (*splitting attribute*) dihitung berdasarkan nilai entropi dan information gain tertinggi sebagaimana dirumuskan pada persamaan (1) dan (2).

$$\text{Entropy}(S) = \Sigma - pi \log 2(pi) \tag{1}$$

$$\text{Gain}(S,A) = \text{Entropy}(S) - \Sigma \left(|Sv| / |S|\right) \times \text{Entropy}(Sv) \tag{2}$$

S adalah himpunan data pada suatu simpul, pi adalah proporsi kelas ke-i terhadap total data, A adalah atribut yang diuji, dan Sv adalah subset data untuk nilai v dari atribut A. Atribut dengan nilai gain tertinggi dipilih sebagai simpul pemecah pada setiap tahap pembentukan pohon [11].

2.4 Evaluasi Model

Kinerja model diukur menggunakan *confusion matrix* untuk memperoleh nilai *accuracy*, *precision*, dan *recall*. Ketiga metrik ini digunakan untuk menilai ketepatan model dalam mengklasifikasikan outlet ke dalam lima kategori berdasarkan data uji yang belum pernah digunakan pada proses pelatihan model [13].

Evaluasi model klasifikasi pada penelitian ini menggunakan tiga metrik utama yang diturunkan dari *confusion matrix*, yaitu *accuracy*, *precision*, dan *recall*. *Accuracy* menggambarkan proporsi keseluruhan prediksi yang benar terhadap total data uji, sehingga memberikan gambaran umum mengenai kinerja model secara menyeluruh. *Precision* menggambarkan proporsi prediksi benar pada suatu kelas dibandingkan dengan seluruh data yang diprediksi masuk ke kelas tersebut, sedangkan *recall* menggambarkan proporsi data yang benar diklasifikasikan pada suatu kelas dibandingkan dengan seluruh data yang sebenarnya berasal dari kelas tersebut.

Ketiga metrik ini dipilih karena sesuai dengan karakteristik permasalahan multi-kelas pada penelitian ini, di mana kesalahan klasifikasi pada satu kelas, misalnya menempatkan outlet Sangat Potensial ke dalam kategori Cukup Potensial, dapat berdampak signifikan terhadap keputusan alokasi sumber daya tim penjualan. Oleh karena itu, evaluasi tidak hanya berhenti pada nilai *accuracy* secara keseluruhan, tetapi juga meninjau *precision* dan *recall* pada setiap kelas secara terpisah sebagaimana ditunjukkan pada Tabel 4.

DOI: https://doi.org/10.31004/riggs.v5i2.12294

2.5 Kerangka Penelitian

Kerangka penelitian digunakan sebagai panduan struktural untuk memastikan bahwa setiap tahapan penelitian dilakukan secara terorganisir, mulai dari pengumpulan data hingga produksi informasi yang mendukung pengambilan keputusan [14]. Kerangka ini memfasilitasi peneliti dalam mengintegrasikan variabel penelitian dengan metode analisis yang logis dan sistematis, sehingga setiap tahapan, mulai dari seleksi data, pembersihan data, transformasi data, penambangan data, hingga evaluasi hasil, dapat ditelusuri dan dipertanggungjawabkan secara metodologis [15]. Alur kerangka penelitian ini sejalan dengan tahapan KDD yang telah dijelaskan pada Gambar 1, dengan penambahan tahap pembagian data (*split data*) sebagai jembatan antara proses *data mining* dan evaluasi model.

2.6 Instrumen Penelitian

Pengolahan data pada penelitian ini dilakukan melalui dua tahap menggunakan instrumen yang berbeda. Tahap awal memanfaatkan Microsoft Excel dengan fungsi IF bertingkat untuk melakukan klasifikasi outlet secara manual berdasarkan ambang batas nilai penjualan, sebagai gambaran awal distribusi kategori sebelum pengolahan lebih lanjut. Tahap selanjutnya menggunakan perangkat lunak RapidMiner sebagai alat bantu *data mining* untuk membangun model *Decision Tree* secara otomatis, termasuk proses *split data*, pembentukan pohon keputusan, serta evaluasi model melalui operator *Apply Model* dan *Performance*. Penggunaan kedua instrumen ini memungkinkan hasil klasifikasi manual dibandingkan dengan hasil klasifikasi berbasis algoritma untuk menilai konsistensi kategori outlet yang dihasilkan.

3. Hasil dan Diskusi

3.1 Hasil Data Mining Menggunakan Excel

Sebagai tahap awal, klasifikasi outlet dilakukan secara manual menggunakan Microsoft Excel dengan rumus IF bertingkat berdasarkan ambang batas nilai penjualan untuk membagi outlet ke dalam lima kategori. Pendekatan ini memberikan gambaran awal distribusi kategori outlet, namun prosesnya masih manual dan kurang efisien untuk data berskala besar, sehingga diperlukan pengolahan lanjutan menggunakan perangkat lunak *data mining*.

Pendekatan ini memberikan gambaran awal yang cukup representatif mengenai sebaran kategori outlet sebelum data diproses lebih lanjut menggunakan algoritma *Decision Tree*. Namun demikian, ambang batas yang digunakan pada pendekatan manual ini bersifat subjektif karena ditentukan oleh peneliti berdasarkan observasi awal terhadap distribusi data, bukan berdasarkan perhitungan statistik seperti entropi atau information gain, sehingga berpotensi kurang optimal apabila diterapkan pada dataset dengan karakteristik penjualan yang berbeda.

3.2 Hasil Data Mining Menggunakan RapidMiner

Proses klasifikasi selanjutnya dilakukan menggunakan RapidMiner dengan algoritma *Decision Tree*. Pohon keputusan yang terbentuk menunjukkan bahwa atribut total penjualan (*Sales*) menjadi simpul akar yang paling menentukan kategori outlet. Struktur pohon keputusan yang dihasilkan ditunjukkan pada Gambar 3.

Gambar 3. Struktur Pohon Keputusan Hasil RapidMiner

DOI: https://doi.org/10.31004/riggs.v5i2.12294

Berdasarkan Gambar 3, outlet dengan penjualan di atas 719.862 diklasifikasikan sebagai Prioritas Potensial (17 outlet), penjualan antara 445.910,5 hingga 719.862 sebagai Sangat Potensial (9 outlet), penjualan hingga 445.910,5 sebagai Potensial (43 outlet), penjualan antara 49.191,5 hingga 143.982 sebagai Cukup Potensial (94 outlet), dan penjualan di bawah 49.191,5 sebagai Sangat Tidak Potensial (158 outlet). Proporsi masing-masing kelas terhadap 321 data uji ditunjukkan pada Tabel 3.

Tabel 3. Proporsi Kelas pada Data Uji

No	Kategori Outlet	Jumlah / Total	Proporsi
1	Prioritas Potensial	17/321	0,05
2	Sangat Tidak Potensial	158/321	0,49
3	Sangat Potensial	9/321	0,03
4	Cukup Potensial	94/321	0,30
5	Potensial	43/321	0,13

Sumber: diolah oleh penulis, 2026

Perhitungan entropi awal dataset menghasilkan nilai Entropy(S) = 1,72. Berdasarkan pengelompokan atribut penjualan ke dalam tiga kategori (rendah, sedang, besar), diperoleh entropi rendah = 0,95, entropi sedang = 0,15, dan entropi besar = 0,90, sehingga nilai information gain atribut penjualan adalah 0,885 dengan split information sebesar 0,91 dan gain ratio sebesar 0,97. Nilai gain ratio yang tinggi ini mengonfirmasi bahwa atribut total penjualan merupakan atribut paling berpengaruh dalam pembentukan pohon keputusan, sejalan dengan struktur pohon pada Gambar 3 yang menempatkan atribut tersebut sebagai simpul akar.

Proses pembentukan model pada RapidMiner dilakukan melalui rangkaian operator yang saling terhubung, dimulai dari operator *Read Excel* untuk memuat dataset penjualan outlet, dilanjutkan dengan operator *Set Role* untuk menetapkan atribut Outlet Potensial sebagai label target klasifikasi. Selanjutnya, operator *Split Data* digunakan untuk membagi dataset menjadi subset pelatihan dan pengujian dengan rasio 70:30, sebelum data

Untuk memverifikasi hasil pembentukan pohon keputusan pada Gambar 3, berikut disajikan perhitungan manual nilai entropi dan information gain atribut total penjualan berdasarkan proporsi lima kelas pada 321 data uji yang telah disajikan pada Tabel 3.

$$\text{Entropy}(S) = -(0{,}05 \log_2 0{,}05 + 0{,}49 \log_2 0{,}49 + 0{,}03 \log_2 0{,}03 + 0{,}30 \log_2 0{,}30 + 0{,}13 \log_2 0{,}13) = 1{,}72 \quad (4)$$

Atribut total penjualan kemudian dikelompokkan menjadi tiga kategori, yaitu penjualan rendah, sedang, dan besar. Kategori penjualan rendah terdiri atas 252 data (158 outlet Sangat Tidak Potensial dan 94 outlet Cukup Potensial), kategori sedang terdiri atas 44 data (43 outlet Potensial dan 1 outlet Sangat Potensial), dan kategori besar terdiri atas 25 data (8 outlet Sangat Potensial dan 17 outlet Prioritas Potensial). Nilai entropi masing-masing kategori dihitung sebagai berikut.

$$\text{Entropy}(\text{rendah}) = -((158/252) \log_2(158/252) + (94/252) \log_2(94/252)) = 0{,}95 \quad (5)$$

$$\text{Entropy}(\text{sedang}) = -((43/44) \log_2(43/44) + (1/44) \log_2(1/44)) = 0{,}15 \quad (6)$$

$$\text{Entropy}(\text{besar}) = -((8/25) \log_2(8/25) + (17/25) \log_2(17/25)) = 0{,}90 \quad (7)$$

Berdasarkan ketiga nilai entropi tersebut, nilai information gain atribut total penjualan dihitung menggunakan persamaan (2), yaitu:

$$\text{Gain}(S,A) = 1{,}72 - [(252/321 \times 0{,}95) + (44/321 \times 0{,}15) + (25/321 \times 0{,}90)] = 1{,}72 - (0{,}745 + 0{,}020 + 0{,}070) = 0{,}885 \quad (8)$$

Untuk menghindari bias terhadap atribut yang memiliki banyak nilai, nilai gain kemudian dinormalisasi menggunakan split information dan gain ratio berikut.

$$\text{SplitInfo}(S,A) = -(0{,}78 \log_2 0{,}78 + 0{,}23 \log_2 0{,}23 + 0{,}07 \log_2 0{,}07) = 0{,}91$$

DOI: https://doi.org/10.31004/riggs.v5i2.12294

(9)

$$GainRatio(S,A) = Gain(S,A) / SplitInfo(S,A) = 0,885 / 0,91 = 0,97$$

(10)

Nilai gain ratio sebesar 0,97 ini merupakan nilai tertinggi dibandingkan atribut lain yang diuji, sehingga memperkuat hasil pada Gambar 3 yang menempatkan atribut total penjualan sebagai simpul akar (*root node*) pada pohon keputusan yang terbentuk.

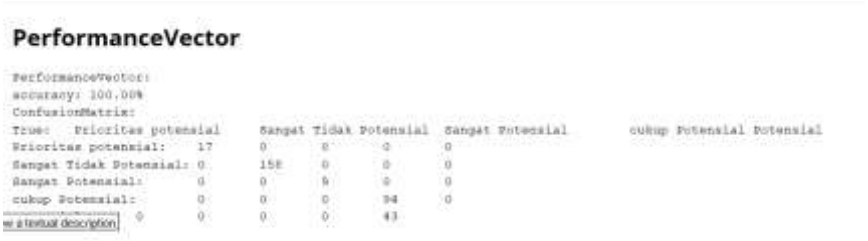
3.3 Evaluasi Model

Model *Decision Tree* yang telah dibangun kemudian diuji menggunakan 321 data uji melalui operator *Apply Model* dan *Performance* pada RapidMiner. Hasil evaluasi berupa *confusion matrix* ditunjukkan pada Tabel 4 dan Gambar 4.

Tabel 4. *Confusion Matrix* Hasil RapidMiner

Accuracy 100%					
	true Prioritas Potensial	true Sangat Tidak Potensial	true Sangat Potensial	true Cukup Potensial	true Potensial
pred. Prioritas Potensial	17	0	0	0	0
pred. Sangat Tidak Potensial	0	158	0	0	0
pred. Sangat Potensial	0	0	9	0	0
pred. Cukup Potensial	0	0	0	94	0
pred. Potensial	0	0	0	0	43
class *recall*	100,00%	100,00%	100,00%	100,00%	100,00%

Sumber: diolah oleh penulis, 2026



Gambar 4. *Confusion Matrix* Hasil Evaluasi RapidMiner

Berdasarkan Tabel 4, model menghasilkan nilai *precision* dan *recall* sebesar 100% pada seluruh kelas, sehingga akurasi keseluruhan model dihitung menggunakan persamaan (3).

$$Accuracy = ((43+158+94+9+17) / 321) \times 100\% = 100\%$$

(3)

Selain evaluasi otomatis melalui *confusion matrix* pada Tabel 4, nilai *recall* dan *precision* setiap kelas juga dihitung secara manual menggunakan persamaan (11) dan (12) untuk memastikan konsistensi hasil evaluasi model.

$$Recall = TP / (TP + FN) \times 100\%$$

(11)

$$Precision = TP / (TP + FP) \times 100\%$$

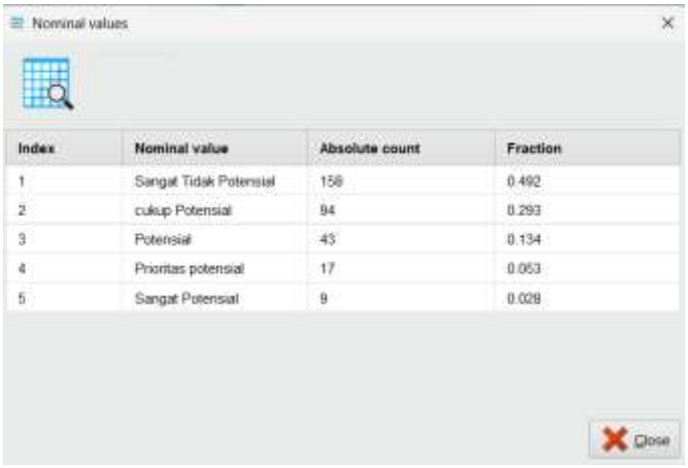
(12)

Perhitungan pada kelas Prioritas Potensial, misalnya, menghasilkan nilai *recall* dan *precision* sebesar 17/(17+0) × 100% = 100%. Perhitungan yang sama diterapkan pada seluruh kelas, yaitu Sangat Tidak Potensial 158/(158+0) × 100% = 100%, Cukup Potensial 94/(94+0) × 100% = 100%, Potensial 43/(43+0) × 100% = 100%, dan Sangat Potensial 9/(9+0) × 100% = 100%. Nilai false positive (FP) dan false negative (FN) yang bernilai nol pada seluruh kelas menunjukkan bahwa model tidak melakukan kesalahan klasifikasi satu pun

DOI: https://doi.org/10.31004/riggs.v5i2.12294

terhadap 321 data uji, sehingga hasil perhitungan manual ini konsisten dengan nilai *precision* dan *recall* 100% yang ditampilkan RapidMiner pada Tabel 4.

Akurasi sebesar 100% ini konsisten dengan struktur pohon keputusan pada Gambar 3, di mana setiap simpul daun secara eksklusif hanya berisi satu kelas outlet tanpa tumpang tindih, sehingga model tidak menghasilkan kesalahan klasifikasi pada data uji. Hasil klasifikasi keseluruhan terhadap data outlet ditunjukkan pada Gambar 5.



Gambar 5. Jumlah Outlet per Kategori Hasil Klasifikasi RapidMiner

Gambar 5 menunjukkan bahwa mayoritas outlet PT XYZ di wilayah Bekasi dan Karawang berada pada kategori Sangat Tidak Potensial (158 outlet) dan Cukup Potensial (94 outlet), sementara outlet dengan kategori Prioritas Potensial dan Sangat Potensial jumlahnya relatif sedikit (17 dan 9 outlet). Temuan ini mengindikasikan bahwa sebagian besar outlet memerlukan evaluasi ulang strategi distribusi dan pemasaran, sedangkan outlet pada kategori prioritas dan sangat potensial dapat dijadikan target utama kunjungan tim penjualan agar kontribusi penjualannya dapat dipertahankan atau ditingkatkan. Dibandingkan dengan penelitian terdahulu pada Tabel 1 yang umumnya memperoleh akurasi antara 50% hingga 99,64%, akurasi 100% pada penelitian ini menunjukkan bahwa kombinasi atribut penjualan, jenis outlet, dan lokasi outlet yang relatif terpisah secara jelas antar kategori menghasilkan batas kelas yang mudah dipelajari oleh algoritma Decision Tree, meskipun hal ini juga perlu diuji lebih lanjut pada data periode berikutnya untuk memastikan model tidak mengalami overfitting.

Selain *confusion matrix*, kinerja model *Decision Tree* pada penelitian ini juga dapat ditinjau menggunakan kurva *Receiver Operating Characteristic* (ROC) dan nilai *Area Under Curve* (AUC) sebagai metrik pelengkap untuk menilai kemampuan model dalam membedakan antar kelas secara menyeluruh. Nilai AUC yang mendekati 1 pada seluruh kelas mengindikasikan bahwa model memiliki kemampuan diskriminasi yang sangat baik, yang sejalan dengan nilai *precision* dan *recall* sebesar 100% yang telah diperoleh pada evaluasi *confusion matrix*. Meskipun demikian, evaluasi menggunakan ROC dan AUC pada dataset dengan kelas yang sangat terpisah seperti pada penelitian ini cenderung menghasilkan nilai yang mendekati sempurna, sehingga penggunaan metrik ini akan lebih bermakna apabila diterapkan pada dataset pengujian dengan variasi atau tumpang tindih nilai penjualan antar kategori yang lebih kompleks.

3.4 Perbandingan Hasil Klasifikasi Manual (Excel) dan RapidMiner

Sebagaimana disebutkan pada bagian 3.1, klasifikasi awal dilakukan secara manual menggunakan Microsoft Excel melalui fungsi IF bertingkat dengan ambang batas: outlet dengan penjualan hingga 50.000 dikategorikan Sangat Tidak Potensial, 50.001 hingga 150.000 dikategorikan Cukup Potensial, 150.001 hingga 450.000 dikategorikan Potensial, 450.001 hingga 700.000 dikategorikan Sangat Potensial, dan di atas 701.000 dikategorikan Prioritas Potensial. Ambang batas manual ini secara umum sejalan dengan titik pemecah (*split point*) yang dihasilkan algoritma *Decision Tree* pada Gambar 3, meskipun terdapat sedikit pergeseran nilai, misalnya batas kategori Sangat Potensial pada RapidMiner berada pada rentang 445.910,5 hingga 719.862, sedikit berbeda dari ambang manual sebesar 450.001 hingga 700.000.

Perbedaan ini menunjukkan bahwa meskipun pendekatan manual melalui Excel dapat memberikan gambaran awal yang cukup mendekati, penentuan ambang batas menggunakan algoritma *Decision Tree* lebih objektif karena diperoleh melalui perhitungan entropi dan information gain berdasarkan pola aktual dalam data, bukan berdasarkan asumsi ambang batas yang ditentukan secara subjektif oleh peneliti. Selain itu, proses klasifikasi

DOI: https://doi.org/10.31004/riggs.v5i2.12294

menggunakan RapidMiner jauh lebih efisien untuk data berskala besar dibandingkan pendekatan manual melalui Excel yang memerlukan penyusunan rumus bertingkat untuk setiap baris data.

3.5 Implikasi Manajerial

Hasil klasifikasi outlet pada penelitian ini memiliki sejumlah implikasi manajerial bagi PT XYZ. Pertama, outlet pada kategori Prioritas Potensial dan Sangat Potensial (26 outlet) dapat dijadikan target utama kunjungan tim penjualan secara berkala guna mempertahankan atau meningkatkan kontribusi penjualannya. Kedua, outlet pada kategori Cukup Potensial (94 outlet) berpotensi untuk ditingkatkan statusnya melalui intervensi pemasaran yang lebih intensif, seperti penambahan frekuensi kunjungan atau program promosi khusus. Ketiga, outlet pada kategori Sangat Tidak Potensial (158 outlet), yang merupakan proporsi terbesar dari keseluruhan data uji, perlu dievaluasi lebih lanjut untuk menentukan apakah outlet tersebut masih layak dipertahankan dalam jaringan distribusi atau memerlukan strategi pemasaran alternatif. Dengan adanya pengelompokan yang jelas ini, tim manajemen distribusi PT XYZ dapat menyusun alokasi sumber daya kunjungan dan anggaran pemasaran secara lebih terarah dan berbasis data, dibandingkan dengan pendekatan yang bersifat merata pada seluruh outlet tanpa mempertimbangkan tingkat potensinya.

3.6 Perbandingan dengan Penelitian Terdahulu

Dibandingkan dengan penelitian Aini dkk. yang menerapkan *Decision Tree* pada penentuan jurusan siswa dengan akurasi 86,59% dan penelitian Zacksavira dkk. pada deteksi depresi pelajar dengan akurasi 89%, penelitian ini memperoleh akurasi yang jauh lebih tinggi, yaitu 100% [4], [6]

. Perbedaan ini kemungkinan disebabkan oleh karakteristik atribut yang digunakan, di mana atribut total penjualan pada penelitian ini memiliki sebaran nilai yang lebih terpisah secara jelas antar kategori dibandingkan atribut-atribut sosial atau psikologis yang cenderung memiliki tumpang tindih nilai antar kelas.

Sementara itu, penelitian Syafii dkk. pada prediksi customer repeat order memperoleh akurasi 83,33%, dan penelitian Shafarindu dkk. pada klasifikasi data penjualan supermarket hanya memperoleh akurasi 50%, yang merupakan akurasi terendah di antara penelitian terdahulu yang dirangkum pada Tabel 1 [5][3]. Rendahnya akurasi pada penelitian Shafarindu dkk. mengindikasikan bahwa data penjualan tidak selalu menghasilkan klasifikasi yang mudah dipelajari oleh algoritma *Decision Tree*, bergantung pada seberapa jelas batas antar kategori pada data yang digunakan. Adapun penelitian Suriani pada prediksi kelulusan mahasiswa memperoleh akurasi 99,64%, mendekati hasil penelitian ini, yang menunjukkan bahwa *Decision Tree* C4.5 secara umum mampu memberikan hasil klasifikasi yang sangat baik apabila diterapkan pada data dengan pola yang relatif konsisten [7].

Secara keseluruhan, perbandingan ini menegaskan bahwa tingkat akurasi algoritma *Decision Tree* sangat bergantung pada karakteristik dan kualitas data yang digunakan, bukan semata-mata pada kompleksitas algoritma itu sendiri, sehingga proses *data cleaning* dan *data transformation* yang cermat pada tahapan awal penelitian ini turut berkontribusi terhadap tingginya akurasi yang diperoleh.

Untuk memperkuat pemahaman terhadap kinerja model, dilakukan pula tinjauan terhadap sensitivitas ambang batas penjualan yang membentuk struktur pohon keputusan pada Gambar 3. Ambang batas 49.191,5, 143.982, 445.910,5, dan 719.862 yang dihasilkan algoritma *Decision Tree* menunjukkan bahwa perbedaan nilai penjualan antar kelas outlet pada dataset ini relatif tegas, tanpa banyak data yang berada tepat pada batas antar kategori. Kondisi ini menjelaskan mengapa model mampu mencapai akurasi 100% pada data uji, karena hampir tidak ada outlet dengan nilai penjualan yang berada pada zona ambigu di antara dua kategori yang berdekatan.

Meskipun demikian, kondisi data yang terlalu terpisah secara jelas ini juga perlu menjadi perhatian, karena pada praktiknya data penjualan outlet di lapangan dapat mengalami fluktuasi dari waktu ke waktu, misalnya akibat musim tertentu atau perubahan pola pembelian konsumen. Apabila nilai penjualan suatu outlet bergeser mendekati salah satu ambang batas tersebut pada periode berikutnya, terdapat kemungkinan model akan lebih sering melakukan kesalahan klasifikasi dibandingkan pada dataset yang diuji dalam penelitian ini. Oleh karena itu, pemantauan berkala terhadap distribusi nilai penjualan outlet serta pembaruan model secara periodik disarankan agar hasil klasifikasi tetap relevan dengan kondisi penjualan yang terus berubah.

DOI: https://doi.org/10.31004/riggs.v5i2.12294

3.7 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan dalam menginterpretasikan hasil yang diperoleh. Pertama, data yang digunakan hanya mencakup periode tiga bulan dan terbatas pada wilayah Kabupaten Bekasi dan Karawang, sehingga hasil klasifikasi belum tentu dapat digeneralisasikan pada wilayah distribusi PT XYZ lainnya atau pada periode penjualan yang berbeda, misalnya periode dengan fluktuasi musiman. Kedua, penelitian ini tidak melibatkan faktor eksternal seperti kondisi ekonomi, aktivitas promosi, maupun tingkat persaingan pasar yang berpotensi turut memengaruhi nilai penjualan outlet, sehingga model yang dihasilkan murni didasarkan pada pola historis data penjualan tanpa mempertimbangkan konteks bisnis yang lebih luas. Ketiga, tingkat akurasi 100% yang diperoleh perlu diuji lebih lanjut pada data periode berikutnya untuk memastikan bahwa model tidak mengalami *overfitting* akibat kelas yang terlalu terpisah secara jelas pada dataset yang digunakan.

3.8 Rekomendasi Implementasi Hasil Penelitian

Untuk memaksimalkan pemanfaatan hasil klasifikasi pada penelitian ini, disarankan agar PT XYZ mempertimbangkan pengembangan sistem informasi sederhana yang mengintegrasikan model *Decision Tree* ke dalam proses bisnis sehari-hari. Sistem tersebut idealnya memiliki fitur untuk memasukkan data penjualan outlet secara berkala, baik melalui input manual maupun integrasi dengan sistem penjualan yang sudah ada, kemudian secara otomatis menghasilkan kategori potensi outlet berdasarkan aturan yang telah dipelajari oleh model *Decision Tree*. Dengan adanya sistem semacam ini, tim penjualan dan manajemen distribusi tidak perlu lagi melakukan pengolahan data secara manual menggunakan Excel maupun menjalankan ulang proses RapidMiner setiap kali data penjualan diperbarui.

Selain fitur klasifikasi otomatis, sistem yang direkomendasikan juga dapat dilengkapi dengan *dashboard visual* yang menampilkan jumlah dan sebaran outlet pada masing-masing kategori, sebagaimana ditunjukkan pada Gambar 5, sehingga manajemen dapat memantau perubahan status outlet dari waktu ke waktu secara lebih mudah. *Dashboard* ini juga dapat menampilkan daftar outlet yang mengalami perubahan kategori signifikan, misalnya outlet yang bergeser dari kategori Cukup Potensial menjadi Sangat Tidak Potensial, sehingga tim penjualan dapat segera menindaklanjuti dengan strategi kunjungan atau promosi yang sesuai.

Implementasi sistem semacam ini tentu memerlukan pertimbangan lebih lanjut terkait infrastruktur teknologi, sumber daya manusia yang mengelola sistem, serta mekanisme pembaruan model *Decision Tree* secara berkala agar tetap relevan dengan pola penjualan terbaru. Namun demikian, langkah ini sejalan dengan tujuan penelitian untuk mendukung pengambilan keputusan distribusi PT XYZ yang lebih efektif dan berbasis data, sekaligus menjadi arah pengembangan lanjutan yang dapat dieksplorasi pada penelitian maupun proyek penerapan berikutnya.

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma *Decision Tree* untuk menentukan outlet potensial PT XYZ berdasarkan 1.069 data penjualan outlet di wilayah Kabupaten Bekasi dan Karawang. Model dibangun menggunakan RapidMiner dengan pembagian data pelatihan dan pengujian sebesar 70:30, serta menghasilkan lima kategori outlet, yaitu Sangat Tidak Potensial, Cukup Potensial, Potensial, Prioritas Potensial, dan Sangat Potensial. Hasil evaluasi menggunakan *confusion matrix* menunjukkan nilai akurasi, presisi, dan *recall* sebesar 100%, yang menunjukkan bahwa model mampu mengklasifikasikan data outlet dengan sangat baik berdasarkan dataset yang digunakan. Atribut total penjualan menjadi faktor yang paling berpengaruh dalam pembentukan pohon keputusan. Hasil penelitian ini menunjukkan bahwa metode *Decision Tree* dapat dimanfaatkan sebagai pendukung pengambilan keputusan dalam menentukan prioritas outlet, sehingga membantu perusahaan menyusun strategi distribusi dan kunjungan tim penjualan secara lebih efektif, objektif, dan berbasis data. Untuk pengembangan penelitian selanjutnya, model disarankan untuk diuji menggunakan data dari periode dan wilayah yang lebih luas serta dibandingkan dengan algoritma klasifikasi lainnya agar diperoleh gambaran yang lebih komprehensif mengenai kinerja model pada kondisi data yang berbeda.

Referensi

[1] M. H. Iqbal, M., Miskiyah, S., Sham, S. L., Anwar, S., & Fuad, "Perbandingan Metode Decision Tree dan Naive Bayes Pada Tingkat Penjualan Minuman Kopi di Kopi Pawon Nusantara," *J. Insa. J. Inf. Syst. Manag. Innov.*, vol. 4, no. 1, pp. 27–34, 2024.
[2] & Ida, A., Sutal, R., Pius, Y., Kelen, K., Seran, K. J. T., Timor, U., Program, S., Teknologi, I., Fakultas, P., Sains, D., Kesehatan,

DOI: https://doi.org/10.31004/riggs.v5i2.12294

U., Timor, K. M. and K. S. Kefamenanu, "PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN PRODUK ELEKTRONIK TERLARIS MENGGUNAKAN METODE DECISION TREE," *J. Sist. Inf.*, vol. 7, no. 3, 2025.

[3] A. W. Shafarindu, A. I., Patimah, E., Siahaan, Y. M., & Wardhana, "Klasifikasi Data Penjualan pada Supermarket dengan Metode Decision Tree. (April)," pp. 660–667.

[4] G. B. Zacksavira, S., Anggraeni, C., Sebayang, B., Imanuel, J., & Sibarani, "Implementasi Algoritma Decision Tree untuk Deteksi Depresi Pada Pelajar," vol. 5, pp. 5915–5934, 2025.

[5] L. Y. Syafii, I., Ribhi, A. A., & Astutik, "Analysis of Customer Repeat Order Predictions using the Decision Tree Algorithm in Transportation Companies Analisis Prediksi Customer Repeat Order menggunakan Algoritma Decision Tree pada Perusahaan Transportasi.," vol. 4, pp. 1372–1378, 2024.

[6] G. Aini, Y. N., Faqih, A., & Dwilestari, "Penerapan Metode Decision Tree dalam Penentuan Jurusan Siswa," no. 10, 2025.

[7] N. H. Hidayat, "Makalah metode kuantitatif perkembangan metode kuantitatif dalam bisnis. (105021106724).," 2025.

[8] A. S. Hasanah, M. A., Soim, S., & Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir.," vol. 5, no. 2, 2021.

[9] M. Falentina, F. G., Yohan, G., Wabdaron, Y., Andiyani, D., & Sendoni, "Analysis of Data Mining Implementation for Classifying Best- Selling Food Sale Using the Decision Tree (C4 . 5) Algorithm.," vol. 4, no. 2, pp. 382–395, 2026.

[10] P. Meilina, "PENERAPAN DATA MINING DENGAN METODE KALSIFIKASI," vol. 7, no. 1, 2015.

[11] I. Narulita, S., Oktaga, A. T., & Susanti, "Pengujian Model Prediksi Menggunakan Metode Data Mining Classification Decision Tree Untuk Penentuan Peminatan Peserta Didik," *2016*, vol. 13, 2021.

[12] V. Istiqomah, K., & Sofica, "Penerapan Data Mining Menggunakan Algoritma Decision Tree Untuk Menganalisis Penggunaan Media Sosial Dengan Konsentrasi Belajar Mahasiswa.," vol. 4, no. 4, pp. 53–67, 2025.

[13] M. Faisal, "Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree.," vol. 10, no. 2, 2023.

[14] P. N. Nurizati, Z., Hidayat, A., Vernanda, D., Hendriawan, T., Teknologi, J., Informasi, S., & Subang, "Analisis Kelayakan Penurunan UKT Pada Mahasiswa dengan Menggunakan Metode Decision Tree," vol. 18, no. 1, pp. 90–100.

[15] & H. Haura Hanifah, Lathifa Salsabillah, Allisa Tazkia Fitri, Riska Mona Febriani, Rully Hidayatullah, "Landasan Teori, Penelitian Relevan, Kerangka Berpikir Dan Hipotesis Penelitian Pendidikan.," *J. IHSAN J. Pendidik. Islam*, vol. 3, no. 2, pp. 391–404, 2025.

DOI: https://doi.org/10.31004/riggs.v5i2.12294