

## Analisis Data Mining untuk Klasifikasi Kafe Populer di Jakarta Menggunakan *Decision Tree* dan Visualisasi dengan *Tableau*

DOI: <http://dx.doi.org/10.35889/jutisi.v14i2.2660>

Creative Commons License 4.0 (CC BY – NC) 

**Siti Anisah<sup>1</sup>, Irwansyah<sup>2\*</sup>**

Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta Timur, Indonesia

\*e-mail *Corresponding Author*: [irwansyah@uhamka.ac.id](mailto:irwansyah@uhamka.ac.id)

### Abstract

*The rapid growth of the café business in Jakarta has increased competition in this sector. Cafes are now not only places to enjoy food and drinks, but also function as cross-generational social spaces. This condition encourages the need for research to understand the factors that influence the popularity of a cafe. This study aims to analyze popular cafes in Jakarta using the C4.5 Decision Tree algorithm and present the results visually through Tableau. The data used includes price, number of reviews, rating, population density, location, address and neighborhood. Using RapidMiner, a classification process was performed to identify the variables that have the most influence on the popularity of a cafe. The Decision Tree algorithm was chosen because it can produce models that are easy to understand and visualize. The results of the analysis showed an accuracy rate of 95.69%, with high precision and recall values in each of the popular and less popular classes.*

**Keyword:** *Classification; Decision Tree C4.5; Cafe Popularity; Data Visualisation; Tableau*

### Abstrak

Pesatnya pertumbuhan bisnis kafe di Jakarta meningkatkan persaingan yang semakin ketat di sektor ini. Kafe kini bukan hanya tempat menikmati makanan dan minuman, tetapi juga berfungsi sebagai ruang sosial lintas generasi. Kondisi ini mendorong perlunya penelitian untuk memahami faktor-faktor yang mempengaruhi popularitas sebuah kafe. Penelitian ini bertujuan untuk menganalisis kafe-kafe populer di Jakarta menggunakan algoritma *Decision Tree C4.5* dan menyajikan hasilnya secara visual melalui Tableau. Data yang digunakan meliputi harga, jumlah ulasan, rating, kepadatan penduduk, lokasi, alamat dan kelurahan. Dengan menggunakan *RapidMiner*, proses klasifikasi dilakukan untuk mengidentifikasi variabel yang paling berpengaruh terhadap popularitas suatu kafe. Algoritma *Decision Tree* dipilih karena mampu menghasilkan model yang mudah dipahami dan divisualisasikan. Hasil analisis menunjukkan tingkat akurasi model sebesar 95,69%, dengan nilai *precision* dan *recall* yang tinggi di masing-masing kelas populer dan kurang populer.

**Kata kunci:** *Klasifikasi; Decision Tree C4.5; Popularitas Kafe; Visualisasi Data; Tableau*

### 1. Pendahuluan

Pesatnya pertumbuhan industri makanan dan minuman di Jakarta, khususnya pada bisnis kafe [1], menunjukkan bahwa kafe tidak hanya menjadi tempat makan dan minum, tetapi juga sebagai ruang sosial di berbagai kalangan masyarakat [2]. Fenomena ini mendorong perlunya dilakukan penelitian untuk memahami faktor-faktor yang mempengaruhi popularitas kafe. Dengan meningkatnya persaingan antar bisnis, adanya kebutuhan dalam pendekatan ilmiah untuk menganalisis data dan menggali informasi strategis untuk membuat keputusan yang lebih tepat.

Saat ini, Jakarta adalah kota dengan jumlah kafe terbanyak di Indonesia, yaitu sekitar 48% dari total kafe di Indonesia [3]. Namun, tidak semua kafe ramai dikunjungi. Beberapa faktor seperti lokasi, pelayanan, harga, suasana, rating dan ulasan pelanggan menjadi faktor utama

dalam menentukan popularitas kafe [4], [5]. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2023, Jakarta memiliki kepadatan penduduk yang sangat tinggi, yaitu 16.165 jiwa/km<sup>2</sup>, yang juga mempengaruhi potensi bisnis [6]. Di sisi lain, perbedaan antara ulasan dan penilaian pelanggan menimbulkan kebingungan dalam menilai kualitas kafe [7]. Permasalahan ini dapat dianalisis dengan mengklasifikasikan popularitas kafe berdasarkan berbagai indikator yang telah ditentukan.

Sebagai solusinya, metode data mining seperti algoritma *Decision Tree* dapat digunakan untuk melakukan klasifikasi berdasarkan variabel-variabel yang relevan. Algoritma ini memiliki keunggulan dalam menghasilkan pola yang mudah dipahami dan diinterpretasikan [8]. Selain itu, untuk mendukung analisis, visualisasi data menggunakan *tableau* dapat membantu menyajikan hasil analisis secara interaktif [9], sehingga memperjelas pola yang ditemukan, dan memudahkan komunikasi data yang lebih efektif [10]. Dengan begitu, dapat mendukung proses pengambilan keputusan yang baik bagi pemilik bisnis.

Peneitian ini bertujuan untuk menganalisis dan mengklasifikasikan kafe populer di Jakarta dengan menggunakan Algoritma *Decision Tree C4.5* dan memvisualisasikan hasil analisis menggunakan *Tableau*. Hasil dari penelitian ini diharapkan dapat memberikan manfaat bagi para pemilik usaha dalam hal memahami preferensi pelanggan dan mendukung strategi bisnis berbasis data, serta membantu konsumen dalam mengambil keputusan berdasarkan informasi objektif dan terstruktur.

## 2. Tinjauan Pustaka

Penelitian pertama yang dilakukan oleh Siti Hinggit [11], penelitiannya yang berjudul “Klasifikasi Kepuasan Pelanggan *Coffee Shop* Paturupa Menggunakan *Decision Tree*” penelitiannya bertujuan untuk menganalisis kepuasan pelanggan dan memberikan rekomendasi peningkatan pelayanan. Studi ini melibatkan 150 responden yang dikategorikan dan menerapkan teknik SMOTE dan *NearMiss* untuk mengatasi ketidakseimbangan data. Hasil terbaik diperoleh dengan pendekatan *NearMiss* pada rasio data pelatihan dan pengujian 70:30, mencapai akurasi 96%, *recall* 100%, *presisi* 95%, dan spesifisitas 75%. Variabel paling signifikan yang memengaruhi kepuasan pelanggan adalah kesesuaian menu makanan dan minuman dengan daftar menu. Berbeda dengan penelitian sebelumnya, penelitian ini berfokus pada klasifikasi kafe populer di Jakarta menggunakan *Decision Tree* dan *Tableau* untuk memberikan wawasan lebih komprehensif dalam analisis tren kafe.

Penelitian kedua oleh Fihir [12], yang berjudul “Klasifikasi Tingkat Kepuasan Pelanggan Kopi Kenangan Menggunakan Metode *Decision Tree* Pada Aplikasi Kopi Kenangan”. Hasil penelitiannya menunjukkan bahwa algoritma *Decision Tree* dapat digunakan untuk mengklasifikasikan kepuasan pelanggan dengan tingkat akurasi 86.96%, akurasi 90% untuk kelas puas dan 66.67% untuk kelas tidak puas. Berbeda dengan penelitian tersebut, penelitian ini tidak berfokus pada analisis kepuasan pelanggan terhadap suatu merek atau aplikasi tertentu, melainkan pada klasifikasi kafe populer yang ada di Jakarta secara lebih luas, dengan mempertimbangkan berbagai variabel seperti harga, rating, ulasan, lokasi, dan kepadatan penduduk.

Penelitian ketiga dilakukan oleh Tanjung [13], dengan judul “Penerapan *Datamining* Klasifikasi Pada Faktor Pemilihan Café Bagi Anak Millineal”, menggunakan *Decision Tree C4.5* untuk menganalisis faktor-faktor (pelayanan, lokasi, wifi dan stop kontak, makanan dan minuman, serta biaya) yang memengaruhi preferensi generasi milenial terhadap kafe di Lapangan Sutomo, Pematangsiantar. Dengan menggunakan *RapidMiner*, menghasilkan model pohon keputusan dan 13 aturan, yang menunjukkan bahwa layanan adalah faktor terpenting dengan akurasi 83,33%. Berbeda dengan penelitian ini, fokus penelitian ini tidak membahas pilihan kafe untuk satu generasi, tetapi lebih mengklasifikasikan kafe-kafe yang ada di Jakarta secara umum tanpa membatasi pada kelompok generasi tertentu.

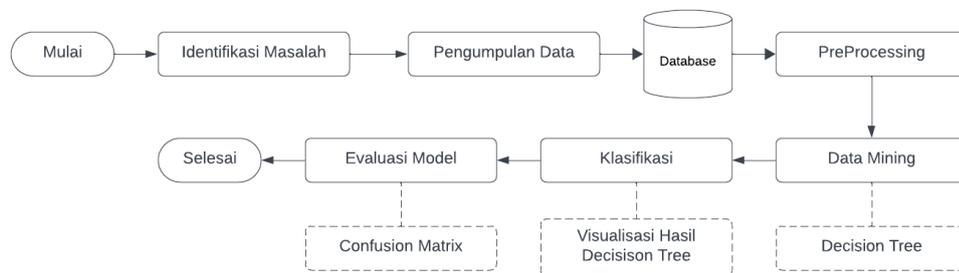
Penelitian keempat dilakukan oleh Iqbal dkk. [14], yang berjudul “Perbandingan Metode *Decision Tree* dan *Naive Bayes* pada Tingkat Penjualan Minuman Kopi di Kopi Pawon Nusantara”. Penelitian ini membandingkan dua algoritma klasifikasi untuk mengetahui metrik mana yang paling cocok, serta mengetahui pola pembelian berdasarkan data penjualan minuman kopi. Hasilnya menunjukkan bahwa algoritma *Decision Tree* merupakan algoritma terbaik yang memiliki akurasi sebesar 97,83% dan AUC 0,957, sementara *Naive Bayes* memiliki akurasi 94,35% dengan AUC 0,969. Berbeda dengan penelitian tersebut, penelitian ini tidak membandingkan dua metode klasifikasi, tetapi berfokus pada penggunaan satu algoritma yaitu,

*Decision Tree C4.5*, untuk mengklasifikasikan popularitas kafe berdasarkan berbagai faktor, serta dilengkapi dengan visualisasi hasil menggunakan *Tableau*.

Penelitian Kelima oleh Kenidy dkk. [15], dengan judul “Visualisasi Data Penjualan dengan *Tableau*”. Penelitian ini berfokus pada bagaimana membuat *dashboard* interaktif menggunakan *Tableau* untuk memvisualisasikan data penjualan harian. Tujuannya untuk membantu manajemen Cafe XYZ yang berada di Jakarta dalam mengidentifikasi item menu yang paling laris dan mengevaluasi kinerja penjualan. Perbedaannya, fokus utama penelitian yang sedang dilakukan bukan hanya pada visualisasi, tetapi pada proses klasifikasi popularitas kafe menggunakan Algoritma *Decision Tree C4.5* dan *Tableau* sebagai alat untuk menyajikan hasil klasifikasi secara interaktif.

### 3. Metodologi

Metode penelitian ini dilakukan menggunakan algoritma *Decision Tree C4.5*, yang bertujuan untuk melakukan klasifikasi pada data mining dan memvisualisasi hasilnya. Adapun proses penelitian melalui beberapa tahapan sistematis yang dimulai dari identifikasi masalah hingga evaluasi model. Setiap tahapannya dirancang untuk memastikan bahwa data yang digunakan telah melalui proses pra-pemrosesan yang tepat, sehingga menghasilkan model klasifikasi yang optimal. Berikut ini adalah alur penelitian secara keseluruhan:



Gambar 1. Alur Penelitian

Penelitian ini focus untuk menganalisis dan mengklasifikasikan kafe-kafe di Jakarta berdasarkan tingkat popularitasnya menggunakan Algoritma *Decision Tree C4.5*. Masalah yang diidentifikasi adalah kurangnya pemahaman terhadap factor-faktor yang menentukan kesuksesan sebuah kafe, di Tengah persaingan bisnis yang semakin kompetitif di Ibukota. Dataset yang digunakan berasal dari dua sumber, yaitu: pertama, melalui *ekstensions* bernama *Insta Data Scraper* dan kedua, dari buku Badan Pusat Statistik (BPS) yang di unduh melalui situs resmi BPS [16].

#### 3.1 Preprocessing

Adapun beberapa tahapan data mining sebelum melakukan klasifikasi data [17]. Dataset yang akan digunakan akan diolah terlebih dahulu, Proses *cleaning* data atau pembersihan data bertujuan untuk menghapus data yang kosong dengan cara membuang *missing value*, duplikasi data, dan memeriksa konsistensi data dan memperbaiki kesalahan data.

Tabel 1. Sebelum *Cleaning*

No	Nama Cafe	Rating	Jumlah Ulasan	Harga	Alamat	Lokasi
1	Seulawah Coffee	4.2	867	Rp 25-50K	Jl. Pesanggrahan No.8, RT.7/RW.5	Pinggir Jalan
2	Tantular Cafe Pesanggrahan	4.4	509	Rp 50-75	...	...
3	Xyz District	4.3	527	...	Jl. Surya Wijaya Raya No.8	Pinggir Jalan
4	...	...	...	...	...	...
584	Nafas Kopi	5.0	44	Rp 25-50K	1, Jl. Flamboyan No.1, RT.1/RW.6, Kb. Jeruk	Pinggir Jalan

Tabel 2. Setelah Cleaning

No	Nama Cafe	Rating	Jumlah Ulasan	Harga	Alamat	Lokasi
1	Seulawah Coffee	4.2	867	Rp 25-50K	Jl. Pesanggrahan No.8, RT.7/RW.5 2, Jl.	Pinggir Jalan
2	Tantular Cafe Pesanggrahan	4.4	509	Rp 50-75	Pesanggrahan No.10G, RT.1/RW.9, Kembangan Sel	Pinggir Jalan
3	Xyz District	4.3	527	Rp 50–100K	Jl. Surya Wijaya Raya No.8	Pinggir Jalan
4	...	...	...	...	...	...
584	Nafas Kopi	5.0	44	Rp 25–50K	1, Jl. Flamboyan No.1, RT.1/RW.6, Kb. Jeruk	Pinggir Jalan

Integrasi data, bertujuan untuk menggabungkan data dari beberapa sumber agar menjadi satu dataset yang terstruktur dan konsisten untuk memenuhi kriteria tertentu sesuai data yang diperlukan. Berikut data sampel yang digabungkan hasil dari web scraping dan data BPS.

Tabel 3. Integrasi Data

No	Nama Cafe	Rating	Jumlah Ulasan	Harga	Alamat	Lokasi	Kelurahan	Kepadatan Penduduk
1	Seulawah Coffee	4.2	867	Rp 25-50K	Jl. Pesanggrahan No.8, RT.7/RW.5 2, Jl.	Pinggir Jalan	Kembangan Selatan	7.510
2	Tantular Cafe Pesanggrahan	4.4	509	Rp 50-75	Pesanggrahan No.10G, RT.1/RW.9, Kembangan Sel	Pinggir Jalan	Kembangan Selatan	7.510
3	Xyz District	4.3	527	Rp 50–100K	Jl. Surya Wijaya Raya No.8	Pinggir Jalan	Kedoya Utara	19.480
4	...	...	...	...	...	...	...	...
584	Nafas Kopi	5.0	44	Rp 25–50K	1, Jl. Flamboyan No.1, RT.1/RW.6, Kb. Jeruk	Pinggir Jalan	Kebon Jeruk	17.860

Selanjutnya data *selection*, data yang telah didapatkan akan di seleksi untuk mengambil kolom yang akan diteliti. Pada penelitian ini mengambil sebanyak 6 kolom.

Tabel 4. Data Selection

No	Atribut Sebelum Selection	Atribut Sesudah Selection
1	Nama Café	Rating
2	Rating	Jumlah Ulasan
3	Jumlah Ulasan	Harga
4	Harga	Lokasi
5	Alamat	Kelurahan
6	Lokasi	Kepadatan Penduduk
7	Kelurahan	
8	Kepadatan Penduduk	

Setelah melalui beberapa tahapan maka didapatkan data siap olah dengan perhitungan menggunakan *tools* RapidMiner. Berikut tabel *output* yang dihasilkan setelah melakukan *preprocessing* data.

Tabel 5. Data Siap Olah

No	Rating	Jumlah Ulasan	Harga	Lokasi	Kelurahan	Kepadatan Penduduk
1	4.2	867	Rp 25-50K	Pinggir Jalan	Kembangan Selatan	7.510
2	4.4	509	Rp 50-75	Pinggir Jalan	Kembangan Selatan	7.510
3	4.3	527	Rp 50–100K	Pinggir Jalan	Kedoya Utara	19.480
4	...	...	...	...	...	...
584	5.0	44	Rp 25–50K	Pinggir Jalan	Kebon Jeruk	17.860

### 3.2 Proses Decision Tree

Decision Tree adalah salah satu algoritma *Machine Learning* yang digunakan untuk membangun model prediksi atau klasifikasi berbasis keputusan dengan struktur yang mirip dengan pohon. Pohon keputusan terdiri dari *node* yang mewakili keputusan, cabang yang mewakili hasil keputusan dan daun yang mewakili hasil klasifikasi akhir [18]. Terdapat banyak jenis algoritma *Decision Tree* yang telah dikembangkan, seperti ID3, CART, dan C4.5, yang masing-masing dengan pendekatan yang berbeda untuk pemilihan atribut dan pembentukan pohon keputusan.

Algoritma C4.5 dipilih untuk penelitian ini karena kemampuannya dalam menangani atribut *kontinu* dan *diskrit*, dan juga *missing values*. Algoritma C4.5 menentukan atribut yang optimal dengan menghitung *entropy* dan *gain ratio* untuk menghasilkan pohon keputusan yang lebih stabil dan akurat, terutama saat memiliki data yang kompleks [19]. Berikut rumus perhitungan *gain* dan *entropy*:

#### 1) Perhitungan Entropy

Perhitungan nilai *entropy* digunakan untuk mengukur ketidakpastian dalam sebuah sekumpulan data. Semakin tinggi nilai *entropy*, semakin tidak pasti hasilnya. Oleh karena itu, *entropy* dimanfaatkan untuk mencari atribut terbaik dalam proses pengambilan keputusan.

$$\text{Entropy}(s) = \sum_{i=1}^n p_i \times \log_2 p_i \quad (1)$$

Keterangan Rumus:

S = Himpunan Kasus

A = Fitur

n = Jumlah partisi S

Pi = Proporsi dari Si terhadap S

#### 2) Perhitungan Gain

Menghitung nilai *gain* yaitu mengukur seberapa besar ketidakpastian (*entropy*) pada sebuah *node* setelah data dibagi berdasarkan suatu atribut. Semakin besar nilai *Gain*, semakin efektif atribut tersebut dalam membagi data menjadi sub-*node* yang lebih teratur.

$$\text{Gain}(S, A) = \text{Entropy}(s) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{entropy}(S_i) \quad (2)$$

Keterangan Rumus:

S = Himpunan Kasus

A = Atribut

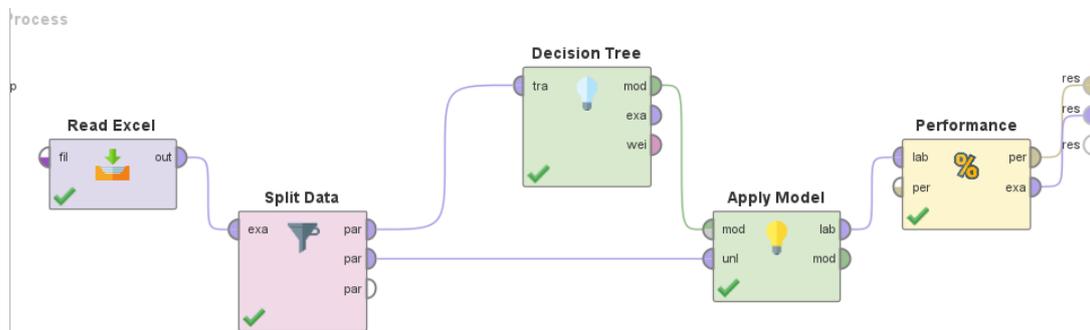
n = Jumlah partisi dari atribut A

|Si| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

### 3.3 Proses RapidMiner

Pada akhir tahap preprocessing, data yang telah dibersihkan dan dipersiapkan akan di proses menggunakan *tools* RapidMiner untuk membangun model klasifikasi menggunakan algoritma *decision tree C4.5*.



Gambar 2. Proses Rapid Miner

Gambar di atas menunjukkan proses klasifikasi kafe-kafe di wilayah Jakarta dengan algoritma *decision tree* menggunakan *tools* RapidMiner. Data yang diproses adalah 584 baris data yang dikumpulkan dari setiap wilayah di Jakarta. Data ini diperoleh dari proses pengumpulan dan pembersihan data menggunakan *Microsoft Excel* sebelum diimpor ke dalam RapidMiner untuk tahap analisis lebih lanjut. Langkah pertama yang dilakukan adalah memasukkan dataset yang akan digunakan dan menghubungkan operator yang telah ditentukan untuk memberi label pada data tersebut. Data kemudian dibagi menjadi dua bagian dengan menggunakan operator *split data*, yaitu 80% data latih dan 20% data uji. Pembagian ini bertujuan agar model dapat dilatih dan diuji secara seimbang untuk menghindari *overfitting*. Selanjutnya, algoritma *Decision Tree C4.5* diterapkan pada data latih untuk membangun model klasifikasi. Model yang telah terbentuk kemudian digunakan untuk menguji data uji menggunakan operator *Apply Model*. Terakhir, performa dievaluasi menggunakan operator *Performance* yang memberikan hasil mengenai akurasi dan model klasifikasi.

### 3.4 Proses Visualisasi Tableau



Gambar 3. Proses Visualisasi

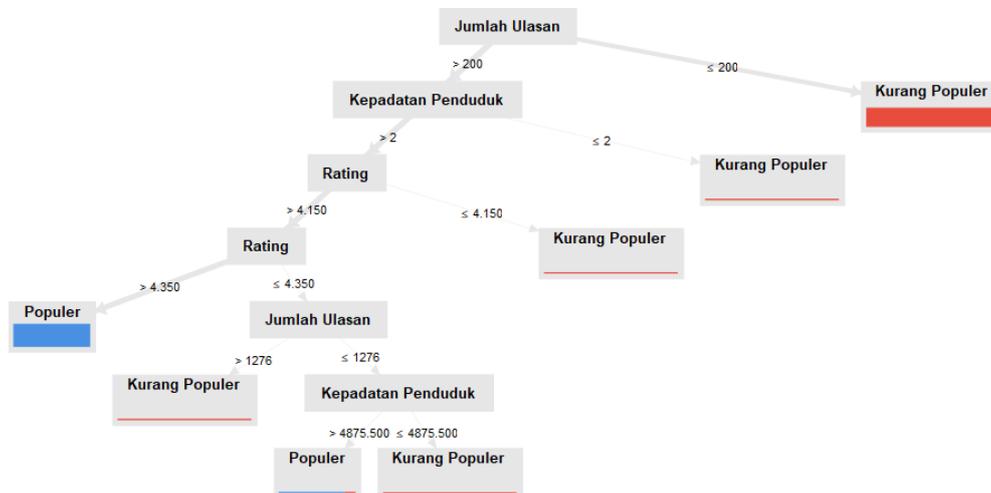
Proses visualisasi data dalam penelitian ini menggunakan *Tableau Public* melalui beberapa tahap. Pertama, dataset yang digunakan dalam bentuk file *Excel (.xlsx)*, yang sudah diproses dan siap untuk dianalisis. Data kemudian diimpor ke *tableau public*, untuk di olah. Setelah data berhasil dimuat, *tableau* menyediakan berbagai opsi visualisasi seperti diagram batang, diagram garis, peta geografis, dan grafik interaktif lainnya.

## 4. Hasil dan Pembahasan

### 4.1 Hasil Klasifikasi

Berdasarkan hasil dari pohon keputusan tersebut, dapat dilakukan analisis terhadap faktor-faktor yang mempengaruhi kualitas kafe dan klasifikasi kafe terpopuler di Jakarta. Pertama, jumlah ulasan dipilih sebagai simpul akar untuk memastikan bahwa ini adalah faktor yang paling penting untuk dipertimbangkan. Kafe dengan kurang dari 200 ulasan cenderung kurang populer. Hal ini dikarenakan semakin tinggi jumlah ulasan, semakin beragam umpan balik yang diterima yang mempengaruhi hasil evaluasi. Kedua, kepadatan penduduk adalah simpul kedua yang menentukan popularitas sebuah kafe. Oleh karena itu, kepadatan penduduk menjadi pertimbangan terpenting kedua setelah jumlah ulasan. Area dengan kepadatan penduduk lebih dari dua memiliki potensi yang lebih tinggi, yang mengindikasikan bahwa lokasi dan target pasar

sangat penting untuk kesuksesan sebuah kafe. Ketiga, simpul ketiga adalah rating. Rating di atas 4.150 dan 4.350 mengindikasikan tingkat kualitas layanan dan produk yang tinggi di kafe, yang berdampak signifikan pada popularitas kafe.



Gambar 4. Hasil Pohon Keputusan

Hasil dari klasifikasi dapat dianalisis bahwa, perhatikan jumlah ulasan yang optimal dan pahami faktor-faktor yang menentukan popularitas kafe. Pilihlah lokasi yang strategis dengan kepadatan penduduk yang sesuai. Selain itu, pertahankan rating 4.350 atau lebih tinggi dan berikan layanan yang baik. Pohon keputusan ini memiliki dampak yang signifikan terhadap peringkat popularitas kafe. Setelah memahami model pohon keputusan berdasarkan dataset yang digunakan, langkah selanjutnya adalah mengevaluasi kinerja model menggunakan RapidMiner.

Tabel 6. *Confusion Matrix* pada RapidMiner

Metrik	Kelas	True Positif	False Positif	Recall	Precision	Accuracy
Populer	Positif	59	2	95.16%	96.72%	-
Kurang Populer	Negatif	52	3	96.30%	94.55%	-
Keseluruhan	-	-	-	-	-	95.69%

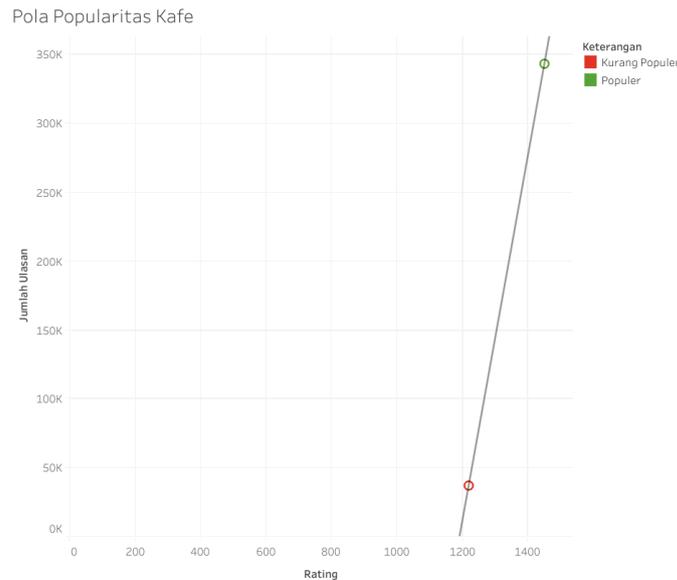
Tabel di atas menunjukkan hasil performa model *Decision Tree C4.5* berdasarkan dataset yang digunakan. Model ini memiliki akurasi keseluruhan sebesar 95,69%, dengan nilai presisi dan recall yang tinggi pada kedua kelas. Dengan hasil tersebut, model dapat dikatakan mampu mengklasifikasikan data secara efektif. Berikut adalah penjelasan perhitungan akurasi secara manual berdasarkan *confusion matrix*:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{59+52}{59+52+2+3} = \frac{111}{116} \\
 &= 0.9569 = 95.69\%
 \end{aligned}$$

**4.2 Hasil Visualisasi**

**1) Pola Popularitas Kafe**

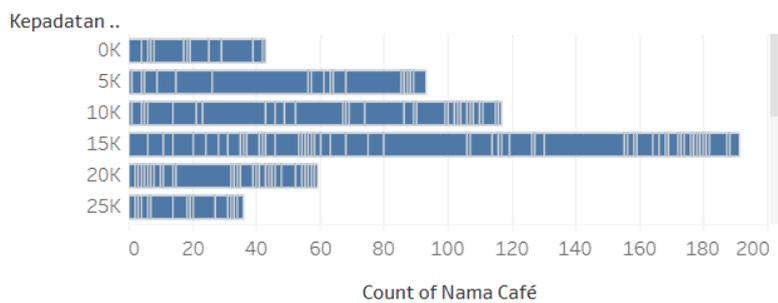
Hasil visualisasi pada Gambar 5 menunjukkan korelasi antara rating dan jumlah ulasan, mengindikasikan bahwa popularitas sebuah kafe tidak hanya bergantung pada rating yang tinggi, tetapi juga pada jumlah ulasan yang diberikan pelanggan. Semakin banyak pelanggan yang memberikan ulasan, semakin besar kemungkinan kafe tersebut dikenal luas dan populer.



Gambar 5. Grafik Popularitas Kafe

2) Distribusi Kafe Berdasarkan Lokasi

Distribusi Kafe Berdasarkan Lokasi



Gambar 6. Grafik Distribusi Kafe

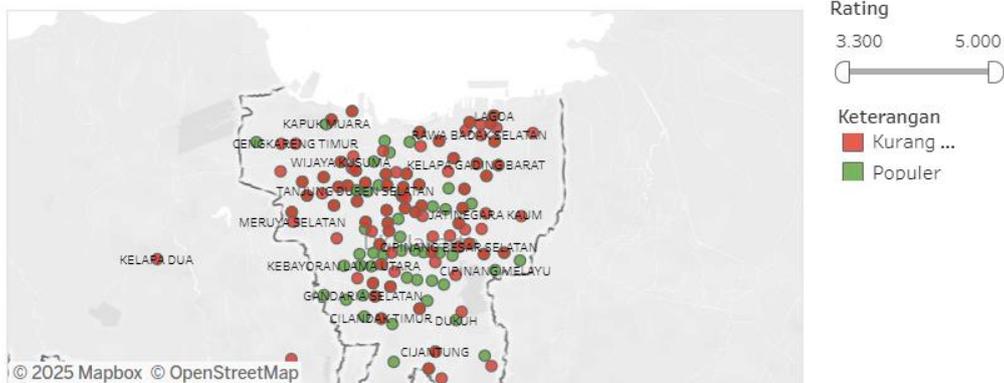
Grafik pada Gambar 6 menunjukkan distribusi menurut lokasinya bahwa jumlah kafe terbesar ditemukan di area dengan kepadatan penduduk 10-20 ribu orang. Hal ini menunjukkan bahwa kafe cenderung memilih Lokasi dengan keseimbangan antara kepadatan penduduk dan aksesibilitas, seperti area komersial atau jalan utama. Sebaliknya terdapat sedikit bar di daerah dengan kepadatan penduduk <5 ribu dan >40 ribu orang dapat disebabkan oleh keterbatasan permintaan di daerah terpencil atau tingginya biaya sewa di pusat kota. Hal ini sejalan dengan temuan penelitian sebelumnya yang menunjukkan bahwa lokasi dengan aksesibilitas tinggi dapat meningkatkan peluang kunjungan pelanggan dan daya Tarik bisnis kafe [20]. Oleh karena itu, memilih lokasi yang tepat tidak hanya mempengaruhi kepuasan pelanggan, tetapi juga merupakan faktor kunci dalam menciptakan keunggulan kompetitif dalam industri kafe.

3) Peta Sebaran Popularitas Kafe

Peta PADA Gambar 7 menunjukkan persebaran distribusi kafe-kafe populer di Jakarta, dengan warna hijau menunjukkan kafe yang populer dan warna merah menunjukkan kafe yang kurang populer. Selain itu, peta ini dilengkapi dengan filter informasi rating yang memungkinkan calon pengguna untuk mencari sesuai dengan rentang nilai rating tertentu, sehingga hanya kafe dengan rating sesuai kriteria yang ditampilkan di peta. Informasi harga juga disertakan dalam tampilan peta untuk memberikan gambaran rentang harga tiap kafe, sehingga pengguna dapat menyesuaikan pilihan berdasarkan preferensi dan anggaran. Sementara bagi pemilik bisnis, informasi ini dapat digunakan sebagai tolak ukur untuk meningkatkan kualitas layanan mereka dan dengan demikian meningkatkan daya saing mereka di pasar. Data ini juga dapat digunakan

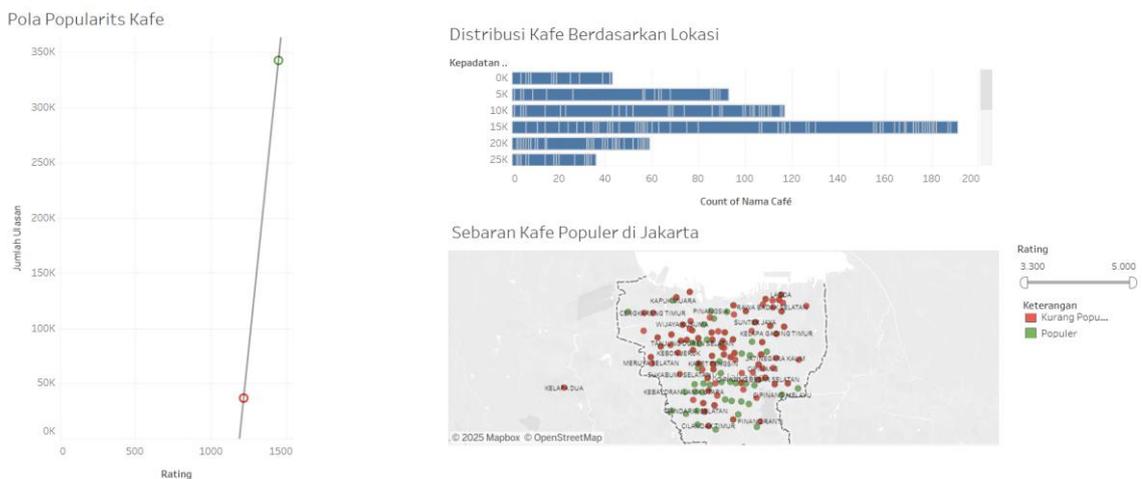
sebagai peluang bisnis, dengan sejumlah area bertitik merah yang menunjukkan potensi strategi berbasis data untuk meningkatkan daya tarik kafe. Dengan menganalisis faktor-faktor keberhasilan kafe yang populer, seperti lokasi yang strategis, harga yang kompetitif dan *customer feedback*. Para pengusaha dapat mengembangkan strategi untuk membuka kafe-kafe baru atau melakukan *rebranding* terhadap kafe yang kurang populer agar lebih menarik.

#### Sebaran Kafe Populer di Jakarta



Gambar 7. Peta Sebaran Popularitsa Kafe

#### 4) Dashboard Visualisasi



Gambar 8. Dasboard Visualisasi

Tampilan pada dashboard Gambar 8 menunjukkan popularitas kafe seluruh wilayah Jakarta. Dengan tiga bagian utama, yaitu *scatter plots*, *stacked bars*, dan *symbol maps*. Alat ini membantu menampilkan data secara visual dengan menggunakan data dan indikator tertentu. Dengan cara ini, pengguna dapat memahami informasi dengan lebih cepat dan akurat.

#### 4.3 Pembahasan

Penelitian ini bertujuan untuk memperkaya studi tentang penerapan algoritma *Decision Tree*, khususnya metode C4.5, dalam mengklasifikasikan tingkat popularitas kafe di Jakarta. Model yang dibangun menghasilkan akurasi sebesar 95,69%, dengan nilai presisi dan recall yang tinggi untuk kedua kelas Populer dan Kurang Populer. Temuan ini menunjukkan bahwa metode C4.5 efektif dalam menyelesaikan permasalahan klasifikasi yang diidentifikasi pada awal penelitian, yakni mendukung pengambilan keputusan bisnis berbasis data.

Selain itu, penelitian ini juga menghasilkan tiga visualisasi; Pola popularitas kafe untuk melihat tren popularitas berdasarkan ulasan dan rating, Distribusi Kafe berdasarkan Lokasi untuk menganalisis distribusi kafe di Jakarta, dan Peta Distribusi Sebaran Popularitas Kafe yang

menunjukkan area dengan popularitas tinggi dan rendah. Kehadiran visualisasi ini mempercepat proses interpretasi data dan meningkatkan akurasi dalam penyusunan strategi bisnis.

Temuan ini sejalan dengan penelitian terdahulu, seperti studi oleh Siti Hinggut [11] yang mencapai akurasi 96% dalam analisis kepuasan pelanggan, serta Fahir [12] dengan akurasi 86,96% pada aplikasi Kopi Kenangan. Meskipun objek penelitian berbeda, hasil-hasil tersebut bersama-sama menguatkan efektifitas algoritma *Decision Tree*, khususnya C4.5, dalam mengidentifikasi faktor-faktor penting dalam sektor layanan konsumen. Dengan demikian, penelitian ini memperkaya literatur terkait penerapan metode klasifikasi berbasis pohon keputusan dalam mendukung strategi bisnis berbasis data.

## 5. Simpulan

Berdasarkan hasil penelitian, metode klasifikasi *Decision Tree C4.5* berhasil diterapkan untuk mengidentifikasi faktor-faktor utama yang memengaruhi popularitas kafe. Model yang dihasilkan menunjukkan akurasi sebesar 95,69%, dengan jumlah ulasan sebagai faktor paling dominan, diikuti oleh kepadatan penduduk dan rating, dalam menentukan popularitas suatu kafe.

Visualisasi data menggunakan *Tableau* membantu menyajikan hasil klasifikasi secara intuitif melalui *dashboard* interaktif. Visualisasi ini dengan jelas menunjukkan bahwa kafe-kafe populer umumnya memiliki jumlah ulasan yang tinggi dan terletak di area padat penduduk dan di lokasi strategis seperti pinggir jalan utama atau area komersial yang ramai. Pola ini menunjukkan bahwa aksesibilitas dan visibilitas memiliki peran penting dalam menarik pelanggan.

Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi model klasifikasi lain guna membandingkan kinerja model dan menambahkan faktor eksternal, seperti tren media sosial dan preferensi demografis pelanggan, untuk memperkaya analisis dan meningkatkan akurasi prediksi.

## Daftar Referensi

- [1] A. Nisa, M. R. Widhiasti, and E. P. Dewi, "Indoor to outdoor: Transformation of Coffee Shops in Jakarta," *Int. J. Built Environ. Sci. Res.*, vol. 6, no. 1, pp. 17–30, 2022, doi: 10.24853/ijbesr.6.1.17-32.
- [2] M. Bitutomo and D. R. S. Ambarwati, "Konsep Sustainable Design Pada Interior Loma Kafe Prambanan," *Sungging*, vol. 2, no. 1, pp. 67–79, 2023, doi: 10.21831/sungging.v2i1.60661.
- [3] F. M. Ayu Riska Salsabillah, Habib Bahari K, "Product Quality, Service Quality, Price, and Location Influence Towards Coffee Shop Customer's Satisfaction," *Manag. Stud. Entrep. J.*, vol. 5, no. 1, pp. 1408–1423, 2024, [Online]. Available: <http://journal.yrpioku.com/index.php/msej>
- [4] R. N. Hikmah, S. S. Putra, and K. Digdowiseiso, "The Influence Of Product Quality, Price Perception, And Service Quality On Consumer Satisfaction At Inspirasi Coffee Jatinegara," *Manag. Stud. Entrep. J.*, vol. 4, no. 6, pp. 9027–9034, 2023, [Online]. Available: <http://journal.yrpioku.com/index.php/msej>
- [5] L. T. Halila Titin Hariyanto, "Analisis Pengaruh Online Customer Review, Online Customer Rating, dan Star Seller terhadap Kepercayaan Pelanggan Hingga Keputusan Pembelian pada Toko Online di Shopee," vol. 9, no. 2, pp. 71–78, 2020, doi: 10.1145/1133890.1133898.
- [6] B. P. Statistik, "Provinsi DKI Jakarta dalam Angka 2024," vol. 54, 2024.
- [7] K. Safitri and E. Widiati, "Pengaruh Online Customer Review, Rating, dan Kepercayaan terhadap Niat Beli Konsumen pada Aplikasi Tokopedia," *Manajerial J. Manaj. dan Sist. Inf.*, vol. 21, no. 2, pp. 153–160, 2022, [Online]. Available: <http://ejournal.upi.edu/index.php/manajerial/>
- [8] B. Aktavera and H. O. L. Wijaya, "Klasifikasi Produk Menggunakan Algoritma Decision Tree," *J. Teknol. Inf. Mura*, vol. 15, no. 1, pp. 24–29, 2024, doi: 10.32767/jti.v15i1.2264.
- [9] S. Nuraeni, "Analisis Visualisasi Pendapatan Perusahaan Sektor Industri Manufaktur Menggunakan Tableau Public," vol. 3, pp. 11494–11505, 2024.
- [10] R. M. Parthe, "Comparative Analysis of Data Visualization Tools: Power BI and Tableau," *Interantional J. Sci. Res. Eng. Manag.*, vol. 07, no. 10, pp. 1–11, 2023, doi: 10.55041/ijrsrem26272.
- [11] S. Hinggut, "Klasifikasi Kepuasan Pelanggan Coffee Shop Paturupa Menggunakan Decision Tree," In *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*,

- 2023, Vol. 4, No. 2, pp. 510-518.
- [12] U. H. Muhammad Fahir, Martanto, "Klasifikasi Tingkat Kepuasan Pelanggan Kopi Kenangan Menggunakan Metode Decision Tree Pada Aplikasi Kopi Kenangan," vol. 7, no. 6, pp. 3830–3833, 2023.
- [13] F. Dwi, P. Tanjung, A. P. Windarto, and E. Irawan, "Penerapan Datamining Klasifikasi Pada Faktor Pemilihan Café Bagi Anak Millineal," *Bull. Data Sci.*, vol. 1, no. 3, pp. 90–98, 2022.
- [14] M. Iqbal, S. Miskiyah, S. L. Sham, S. Anwar, and M. H. 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, doi: 10.31294/jinsan.v4i1.3682.
- [15] N. T. Kenidy, T. Sutrisno, and I. Lewenusa, "Visualisasi Data Penjualan dengan Tableau," vol. 14, no. 2, pp. 143–147, 2024, [Online]. Available: <https://jurnal.ubl.ac.id/index.php/expert/article/download/4050/2870>
- [16] Dinas Kependudukan dan Pencatatan Sipil DKI Jakarta, "Data Kepadatan Penduduk DKI Jakarta," 2024. [Online]. Available: [https://satudata.jakarta.go.id/open-data/detail?kategori=dataset&page\\_url=data-kepadatan-penduduk-provinsi-dki-jakarta&data\\_no=1](https://satudata.jakarta.go.id/open-data/detail?kategori=dataset&page_url=data-kepadatan-penduduk-provinsi-dki-jakarta&data_no=1)
- [17] I. N. Dwi Vernanda, Tri Herdiawan, Erika Rachmawati, "Analisis Tingkat Pengunduran Diri Dan Strategi Peningkatan Partisipasi Mahasiswa Menggunakan Metode Decision Tree," vol. 9, no. 2, pp. 145–154, 2023.
- [18] S. S. A. Larasati, E. N. K. Dewi, B. H. Farhansyah, F. A. Bachtiar, and F. Pradana, "Penerapan Decision Tree dan Random Forest dalam Deteksi Tingkat Stres Manusia Berdasarkan Kondisi Tidur," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 7, pp. 1503–1510, 2023, doi: 10.25126/jtiik.1077993.
- [19] E. Prasetyo and B. Prasetyo, "Peningkatan Akurasi Klasifikasi Algoritma C 4.5 Menggunakan Teknik Bagging pada Diagnosis Penyakit Jantung," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 1035–1040, 2020, doi: 10.25126/jtiik.2020752379.
- [20] D. Lie, M. Butarbutar, S. Simatupang, E. Efendi, H. A. Damanik, and M. F. Silaen, "Pengaruh Lokasi Terhadap Kepuasan Konsumen Pada Kedai Kopi Baravi Pematangsiantar," *Strateg. J. Manag. Sci.*, vol. 1, no. 2, pp. 32-41, 2021, doi: 10.37403/strategic.v1i2.19.