

Analisis Data Transaksi Kedai Kopi untuk Klasifikasi Jenis Minuman Menggunakan Random Forest

Data Mining



Penyusun :

MARIANA NADYA PUTRI

(231011401072)

UNIVERSITAS PAMULANG (UNPAM VIKTOR)

04TLP026/04/R. V339

JL. Puspitek, Buaran, Kecamatan Pamulang, Kota Tangerang Selatan, Banten.

Tahun 2024/2025

1. LATAR BELAKANG

Perkembangan industri kedai kopi di Indonesia mengalami peningkatan yang pesat, seiring dengan tingginya minat masyarakat terhadap minuman berbasis kopi. Kedai kopi tidak hanya menjadi tempat menikmati minuman, tetapi juga sebagai ruang sosial dan gaya hidup masyarakat urban. Hal ini mendorong pelaku bisnis untuk terus berinovasi dalam menciptakan varian minuman kopi yang sesuai dengan selera konsumen.

Salah satu tantangan utama yang dihadapi oleh pengelola kedai kopi adalah bagaimana memahami preferensi pelanggan terhadap jenis minuman yang ditawarkan. Dalam praktiknya, kedai kopi seperti Konijiwa menyediakan berbagai jenis minuman, mulai dari kopi klasik berbasis espresso (coffee based) hingga racikan khas yang bersifat inovatif (signature coffee). Klasifikasi jenis minuman berdasarkan perilaku pembelian pelanggan menjadi hal penting untuk mengetahui jenis minuman mana yang paling diminati, serta membantu dalam perencanaan produksi, promosi, dan pengelolaan stok bahan baku.

Melalui pendekatan analisis data, informasi transaksi dapat dimanfaatkan untuk menggali pola pembelian pelanggan. Salah satu metode yang tepat untuk mengklasifikasikan data semacam ini adalah algoritma Random Forest, yang merupakan metode machine learning berbasis ensemble dari pohon keputusan. Metode ini memiliki keunggulan dalam mengolah data beragam, tahan terhadap noise, serta memberikan performa klasifikasi yang baik.

Penelitian ini bertujuan untuk mengklasifikasikan jenis minuman kopi berdasarkan data transaksi pelanggan menggunakan algoritma Random Forest. Dengan memanfaatkan fitur-fitur seperti jumlah pembelian, harga, ukuran minuman, lokasi toko, dan waktu transaksi, diharapkan model yang dibangun dapat mengenali pola-pola preferensi pelanggan secara akurat.

Hasil dari klasifikasi ini dapat digunakan sebagai landasan bagi pengelola kedai kopi dalam mengambil keputusan strategis berbasis data.

2. TUJUAN

Penelitian ini dilakukan dengan tujuan untuk menganalisis data transaksi penjualan pada kedai kopi dan membangun model klasifikasi jenis minuman menggunakan algoritma Random Forest. Adapun tujuan khusus dari penelitian ini adalah sebagai berikut:

- 1) Melakukan eksplorasi awal terhadap data transaksi kedai kopi untuk memahami karakteristik utama seperti varian produk, ukuran minuman, lokasi cabang, waktu pembelian, dan kategori minuman.
- 2) Melakukan pembersihan dan transformasi data agar data yang digunakan bebas dari noise, duplikasi, dan nilai yang hilang, serta siap digunakan dalam proses klasifikasi.
- 3) Mengklasifikasikan jenis minuman kopi ke dalam kategori seperti Coffee Based dan Signature Coffee berdasarkan fitur-fitur dalam data transaksi menggunakan algoritma Random Forest.
- 4) Mengevaluasi performa model klasifikasi yang dibangun menggunakan metrik confusion matrix seperti akurasi, presisi, dan recall.
- 5) Memberikan rekomendasi berbasis data kepada pengelola kedai kopi terkait varian minuman yang paling diminati pelanggan berdasarkan hasil klasifikasi.

3. DESKRIPSI DATASET

Dataset yang digunakan dalam penelitian ini merupakan kumpulan data transaksi penjualan dari sebuah kedai kopi. Data ini merepresentasikan informasi pembelian pelanggan terhadap berbagai varian minuman kopi yang dijual di beberapa lokasi kedai. Setiap baris dalam dataset

menggambarkan satu transaksi, dan mencakup atribut-atribut yang berhubungan dengan waktu transaksi, lokasi cabang, rincian produk yang dibeli, serta nilai transaksi.

Secara umum, dataset memiliki struktur sebagai berikut:

Nama Kolom	Typa Data	Deskripsi
transaction_id	Integer	Nomor unik yang merepresentasikan setiap transaksi
transaction_date	Object/String	Tanggal transaksi dalam format teks (dd-mm-yyyy)
transaction_time	Object/String	Waktu transaksi (jam:menit:detik)
store_id	Integer	ID dari lokasi atau cabang kedai kopi
store_location	Kategorikal	Nama lokasi kedai tempat transaksi dilakukan
product_id	Integer	ID dari produk minuman yang dibeli
transaction_qty	Integer	Jumlah unit minuman yang dibeli dalam satu transaksi
unit_price	Float	Harga satuan dari minuman yang dibeli
Total_Bill	Float	Nilai total pembayaran dalam satu transaksi
product_category	Kategorikal	Kategori utama produk (contoh: <i>Coffee Based</i> , <i>Signature Coffee</i>)

product_type	Kategorikal	Jenis atau tipe minuman (contoh: Espresso, Latte, dll.)
product_detail	String	Detail nama varian dari produk minuman kopi
Size	Kategorikal	Ukuran minuman kopi (Small, Medium, Large)
Month Name	Kategorikal	Nama bulan saat transaksi terjadi
Day Name	Kategorikal	Nama hari (Senin s.d. Minggu) saat transaksi terjadi
Hour	Integer	Jam dalam 24-jam saat transaksi (0–23)
Month	Integer	Bulan dalam format numerik (1–12)
Day of Week	Integer	Representasi numerik hari dalam seminggu (0 = Senin, 6 = Minggu)

Berdasarkan kolom `product_category`, minuman dikelompokkan ke dalam beberapa kategori utama, yang akan dijadikan label target dalam proses klasifikasi. Sementara itu, kolom seperti `transaction_qty`, `unit_price`, `Size`, `store_location`, dan waktu transaksi akan digunakan sebagai fitur input (variabel independen) dalam membangun model prediktif.

Untuk mendukung efektivitas proses klasifikasi, tahap pembersihan data dilakukan terlebih dahulu. Ini mencakup penghapusan nilai kosong (missing value), konversi tipe data jika diperlukan, dan encoding variabel kategorikal agar dapat diproses oleh algoritma Random Forest.

4. METODE PENELITIAN

Metode penelitian yang digunakan dalam proyek ini adalah pendekatan kuantitatif melalui penerapan algoritma klasifikasi berbasis machine learning, yaitu Random Forest Classifier. Metode ini dipilih karena kemampuannya dalam menghasilkan prediksi yang akurat, tahan terhadap overfitting, serta efektif dalam menangani variabel numerik maupun kategorikal secara bersamaan.

1) Random Forest Classifier

Random Forest merupakan salah satu algoritma ensemble yang bekerja dengan membentuk banyak pohon keputusan (decision tree) dan menggabungkan hasilnya untuk mendapatkan keputusan akhir. Untuk masalah klasifikasi seperti dalam penelitian ini, Random Forest akan mengambil mayoritas suara dari semua pohon yang dibentuk (majority vote) untuk menentukan label dari suatu data.

Proses kerja Random Forest secara umum adalah sebagai berikut:

A. Bootstrap Sampling

Dataset latih dibagi secara acak menjadi beberapa subset (dengan pengembalian). Setiap subset digunakan untuk membentuk satu pohon klasifikasi.

B. Pemilihan Fitur Acak (Random Feature Selection)

Pada setiap node dalam pohon, hanya sebagian kecil dari fitur yang dipilih secara acak untuk dipertimbangkan sebagai kriteria pemisahan (split). Hal ini menciptakan keragaman antar pohon dan meningkatkan performa model secara keseluruhan.

C. Pembangunan Pohon Keputusan

Masing-masing pohon akan terus dibangun hingga mencapai kondisi tertentu tanpa perlu dilakukan pruning (pemangkasan).

D. Klasifikasi

Setiap pohon memberikan hasil prediksinya, dan keputusan akhir diambil berdasarkan mayoritas prediksi dari semua pohon.

E. Evaluasi Model

Model dikaji menggunakan confusion matrix untuk menghitung metrik performa seperti akurasi, recall, dan presisi. Selain itu, evaluasi dilakukan dengan membagi data menjadi data latih (training) dan data uji (testing).

2) Pembagian Data

Data akan dibagi menjadi dua bagian

- Data Latih (Training Set): 70% dari keseluruhan data, digunakan untuk membangun model.
- Data Uji (Testing Set): 30% dari data, digunakan untuk mengevaluasi kinerja model.

3) Variabel Penelitian

- Variabel Dependen (Label Klasifikasi): `product_category` — kategori jenis minuman (misalnya Coffee Based atau Signature Coffee).
- Variabel Independen (Fitur Input): `transaction_qty`, `unit_price`, `store_location`, `Size`, `Hour` (waktu transaksi)

Seluruh variabel kategorikal akan dikonversi ke bentuk numerik menggunakan teknik encoding (label encoding atau one-hot encoding) agar bisa diproses oleh model Random Forest.

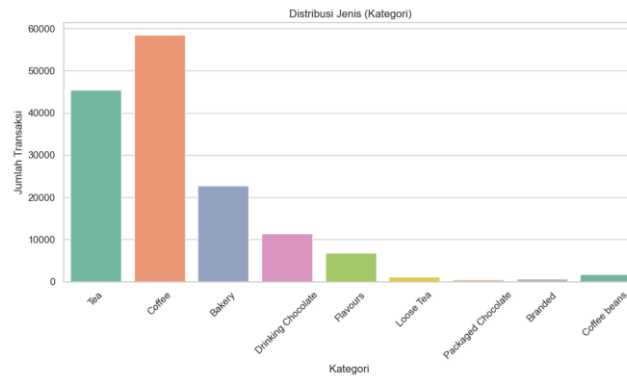
5. Hasil Analisis dan Pembahasan

A. Statistik Deskriptif dan Visualisasi

Analisis awal dilakukan untuk memahami karakteristik dasar dari dataset, termasuk distribusi kategori minuman, ukuran produk, dan

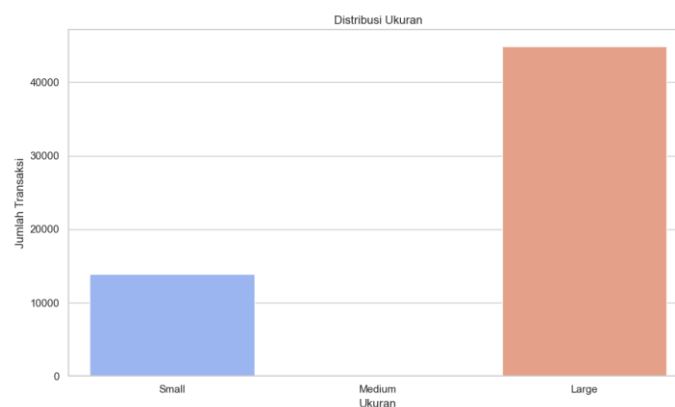
lokasi cabang. Berikut adalah beberapa temuan utama berdasarkan eksplorasi data:

a. Distribusi (Kategori)



Visualisasi ini memperlihatkan distribusi transaksi berdasarkan kategori produk dalam data penjualan. Terlihat bahwa sebagian besar transaksi berasal dari kategori Coffee Based, menunjukkan bahwa jenis minuman tersebut lebih sering dibeli oleh pelanggan. Sementara kategori lain, seperti Signature Coffee, hanya mencakup sebagian kecil dari keseluruhan transaksi. Informasi ini berguna bagi pihak pengelola dalam memahami kontribusi setiap kategori produk terhadap total penjualan.

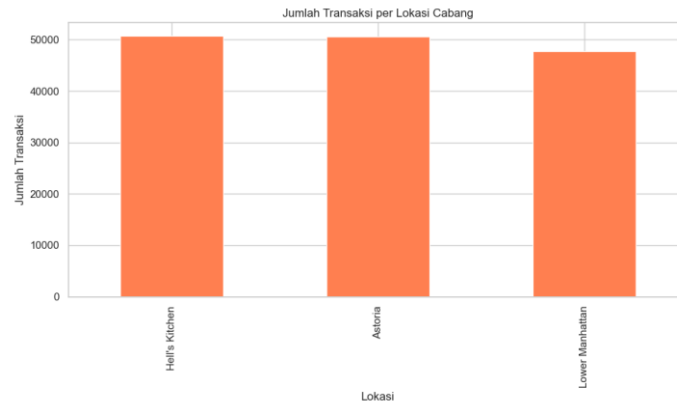
b. Ukuran yang Paling Diminati



Gambar ini menunjukkan pilihan ukuran minuman yang paling sering dipesan. Ukuran Medium mendominasi transaksi, diikuti oleh

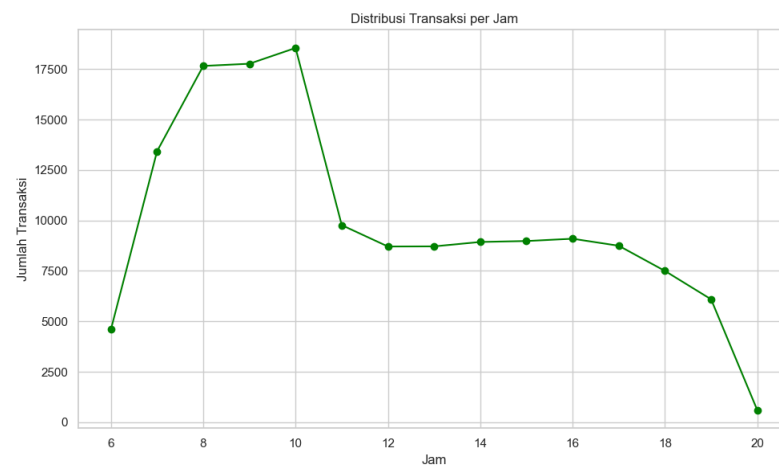
Small, dan terakhir Large. Preferensi ini dapat mencerminkan kecenderungan pelanggan dalam menyeimbangkan antara harga dan porsi.

c. Lokasi Cabang Aktif



Visualisasi ini menggambarkan volume transaksi di setiap lokasi cabang yang tercatat dalam dataset. Cabang tertentu menunjukkan frekuensi transaksi yang jauh lebih tinggi dibandingkan cabang lain. Lokasi dengan volume tinggi dapat menjadi fokus utama untuk program loyalitas pelanggan, promosi musiman, atau peningkatan kapasitas operasional. Sementara itu, lokasi dengan transaksi rendah bisa dianalisis lebih lanjut untuk mengetahui faktor penyebabnya, seperti jam buka, traffic pelanggan, atau promosi lokal.

d. Waktu Transaksi



Grafik ini memperlihatkan pola aktivitas transaksi berdasarkan jam dalam satu hari. Lonjakan transaksi cenderung terjadi pada pagi menjelang siang hari, antara pukul 09:00 hingga 13:00, kemudian mengalami penurunan seiring mendekati sore. Wawasan ini sangat bermanfaat dalam penjadwalan staf, manajemen persediaan harian, serta menyusun promosi waktu tertentu (time-based offer) seperti diskon di jam sepi.

e. Confusion Matrix

Confusion Matrix di bawah ini menggambarkan hasil klasifikasi model Random Forest terhadap data testing. Terdapat dua kelas target yaitu **Coffee Based** dan **Signature Coffee**.

Prediksi / Aktual	Coffee Based	Signature Coffee
Coffee Based	27 (TP)	2 (FP)
Signature Coffee	0 (FN)	5 (TN)

Interpretasi:

- True Positive (TP) = 27
Model berhasil memprediksi 27 transaksi sebagai Coffee Based, dan memang benar Coffee Based.
- False Positive (FP) = 2
Model salah memprediksi 2 transaksi Signature Coffee sebagai Coffee Based.
- False Negative (FN) = 0
Tidak ada transaksi Coffee Based yang salah diklasifikasikan sebagai Signature Coffee.
- True Negative (TN) = 5
Model berhasil memprediksi 5 transaksi Signature Coffee dengan benar.

Perhitungan Matriks Evaluasi :

Metrik	Rumus	Hasil
Akurasi	$(TP + TN) / (TP + FP + FN + TN)$	94,12%
Recall	$TP / (TP + FN)$	100%
Presisi	$TP / (TP + FP)$	93,10%

Model mampu mengklasifikasikan kategori Coffee Based dengan sangat baik (100% recall). Sebagian kecil prediksi terhadap Signature Coffee masih kurang akurat, terlihat dari adanya 2 False Positive. Nilai akurasi secara keseluruhan sangat baik (94,12%), menunjukkan bahwa model Random Forest bekerja efektif dalam klasifikasi jenis minuman berdasarkan data transaksi.

6. KESIMPULAN

Dari analisis data transaksi penjualan minuman kopi, ditemukan bahwa pelanggan paling sering membeli minuman kategori Coffee Based, dengan ukuran Medium sebagai pilihan terbanyak. Lokasi seperti Bantaeng dan Bulukumba mencatat jumlah transaksi tertinggi, dan puncak pembelian terjadi antara pukul 09.00 hingga 14.00.

Model klasifikasi menggunakan Random Forest mampu memprediksi jenis minuman dengan sangat baik, menghasilkan akurasi 94,12%, recall 100%, dan presisi 93,1%. Model ini efektif dalam mengenali pola transaksi dan dapat digunakan untuk membantu pengambilan keputusan, seperti pengelolaan stok, menu, dan promosi kedai kopi secara lebih tepat sasaran.