

PENERAPAN MODEL PREDIKSI PENJUALAN PADA USAHA RUMAH MAKAN MENGGUNAKAN ALGORITMA RANDOM FOREST

Farah Nur Farida ¹, Ahmad Faqih ², Sandy Eka Permana ³

^{1,2} Teknik Informatika, STMIK IKMI Cirebon, Indonesia

³ Manajemen Informatika, STMIK IKMI Cirebon, Indonesia

Jl. Perjuangan No.10B, Kota Cirebon, Indonesia

qfarahnurfarida1@gmail.com

ABSTRAK

Dalam dunia bisnis kuliner, prediksi penjualan yang akurat sangat penting untuk memastikan efisiensi operasional, pengelolaan persediaan yang optimal, serta strategi promosi yang tepat. Penelitian ini bertujuan untuk membangun model prediksi penjualan di usaha rumah makan menggunakan algoritma Random Forest, yang dikenal karena kemampuannya dalam menangani data kompleks dan menghasilkan prediksi yang akurat. Penelitian ini menggunakan pendekatan kuantitatif dengan teknik *machine learning*, di mana data penjualan dikumpulkan selama sepuluh bulan (Januari hingga Oktober 2024). Setelah data dikumpulkan, dilakukan tahap preprocessing untuk membersihkan dan mempersiapkan data agar dapat digunakan oleh model. Algoritma *Random Forest* kemudian diterapkan untuk membangun model prediksi penjualan. Hasil evaluasi menunjukkan bahwa model ini mencapai akurasi 97% pada data uji, yang menandakan bahwa model ini sangat efektif dalam mengklasifikasikan kategori penjualan dengan tingkat kesalahan yang minimal. Selain itu, model ini juga menunjukkan keseimbangan yang baik antara precision dan recall, yang mengindikasikan kemampuannya dalam meminimalkan kesalahan klasifikasi. Model prediksi ini terbukti efektif untuk membantu usaha rumah makan dalam pengelolaan operasional dan strategi pemasaran.

Kata kunci : *Prediksi Penjualan, Random Forest, Promosi, Usaha Rumah Makan*

1. PENDAHULUAN

Cara Perkembangan pesat di bidang informatika telah membawa dampak signifikan pada hampir semua aspek kehidupan, termasuk sektor bisnis dan pendidikan. Penerapan teknologi informatika memungkinkan perusahaan mengotomatisasi berbagai proses bisnis, meningkatkan efisiensi, serta mempercepat pengambilan keputusan melalui analisis data yang lebih akurat [1].

Di dunia pendidikan, teknologi memberikan akses yang lebih inklusif terhadap pembelajaran melalui platform daring dan sistem evaluasi otomatis [2].

Kemajuan teknologi kecerdasan buatan (AI) dan *machine learning* semakin memperluas penerapan informatika, terutama dalam prediksi penjualan dan analisis preferensi pasar [3].

Teknologi ini juga memberikan kontribusi penting dalam sektor kesehatan dan keuangan, memungkinkan analisis data medis untuk diagnosis yang lebih cepat [4], serta pengelolaan risiko dan prediksi kredit dalam sektor keuangan [5].

Secara keseluruhan, informatika berperan sebagai pendorong utama transformasi digital di berbagai sektor.

Dalam industri kuliner, khususnya usaha rumah makan, fluktuasi penjualan sering kali terjadi akibat berbagai faktor seperti tren musiman, promosi, hari libur, dan acara khusus. Ketidakpastian ini menyulitkan pelaku usaha dalam mengelola stok, merancang strategi promosi, dan melakukan perencanaan operasional yang efisien. Padahal, pengambilan keputusan yang tepat berdasarkan data

historis penjualan dapat meningkatkan efisiensi operasional dan profitabilitas.

Salah satu pendekatan yang menjanjikan dalam mengatasi permasalahan tersebut adalah penggunaan teknologi prediktif berbasis *machine learning*. Namun, di tengah potensi besar tersebut, usaha kecil seperti rumah makan masih menghadapi tantangan dalam mengelola data yang kompleks serta keterbatasan sumber daya untuk menerapkan teknologi canggih. Selain itu, isu keamanan data dan kerentanan terhadap serangan *siber* turut menjadi hambatan dalam adopsi teknologi ini secara lebih luas [1].

Beberapa studi sebelumnya menunjukkan bahwa algoritma seperti *Random Forest* dan *Support Vector Machines* telah berhasil digunakan untuk berbagai tujuan prediksi, seperti deteksi penipuan kartu kredit [2] dan prediksi retensi [3].

Namun, kajian yang secara khusus menerapkan algoritma ini untuk memprediksi penjualan pada usaha rumah makan masih terbatas, meskipun kebutuhan akan prediksi yang akurat di sektor ini sangat tinggi.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma *Random Forest* dalam memprediksi penjualan pada usaha rumah makan. Data yang digunakan mencakup penjualan historis beserta faktor eksternal seperti hari libur dan acara khusus. Proses pengolahan data mencakup *preprocessing* untuk memastikan kualitas data, pelatihan model, dan evaluasi performa menggunakan metrik seperti *Mean Absolute Error (MAE)* dan *Root Mean Squared Error (RMSE)*.

Dengan pendekatan ini, diharapkan penelitian dapat memberikan solusi konkret bagi pelaku usaha

rumah makan dalam mengelola stok secara efisien, merencanakan strategi promosi yang tepat, serta mengoptimalkan kegiatan operasional harian. Selain itu, hasil penelitian ini juga diharapkan dapat mengisi celah penelitian dalam penerapan *machine learning* di sektor usaha kecil dan menengah, khususnya di industri kuliner.

2. TINJAUAN PUSTAKA

2.1. Algoritma Random Forest dalam Prediksi Penjualan

Random Forest merupakan salah satu algoritma yang populer dalam bidang machine learning karena kemampuannya untuk menangani data yang kompleks dan besar dengan tingkat akurasi yang tinggi. Menurut Husein, penggunaan Random Forest dalam peramalan penjualan di sektor bisnis menunjukkan hasil yang cukup baik dalam mengidentifikasi pola permintaan berdasarkan data historis. Dalam konteks usaha rumah makan, penerapan algoritma ini dapat memberikan prediksi penjualan yang lebih tepat, yang memungkinkan pengelola untuk mengatur persediaan bahan baku dan tenaga kerja dengan lebih efisien [4].

2.2. Prediksi Penjualan dalam Sektor Bisnis

Penggunaan model prediksi dalam sektor bisnis, khususnya untuk peramalan penjualan, telah banyak diterapkan dengan berbagai algoritma. Husein et al. meneliti penggunaan Random Forest untuk peramalan penjualan, yang terbukti efektif dalam menangani variabel yang saling berinteraksi dan mempengaruhi penjualan. Teknik ini dapat digunakan dalam rumah makan untuk memperkirakan jumlah pelanggan yang datang pada hari tertentu dan untuk merencanakan strategi pemasaran serta pengelolaan stok bahan baku. Penelitian ini mendasari pentingnya penggunaan machine learning dalam pengambilan keputusan yang lebih cerdas dan berbasis data [4],[5].

2.3. Penerapan Algoritma Random Forest dalam Prediksi Harga dan Penjualan

Algoritma Random Forest juga banyak digunakan dalam aplikasi lain, seperti prediksi harga sewa apartemen, yang diterapkan oleh Mulyahati. Penelitian ini menunjukkan bagaimana Random Forest dapat digunakan untuk memprediksi harga dengan mempertimbangkan berbagai faktor yang mempengaruhi pasar, seperti lokasi dan fasilitas. Meskipun berbeda domain, pendekatan yang digunakan oleh Mulyahati dapat diadaptasi untuk prediksi penjualan di rumah makan, di mana faktor-faktor seperti waktu, cuaca, dan hari dalam seminggu dapat memengaruhi volume penjualan dan kebutuhan terhadap menu tertentu [6].

2.4. Penggunaan Random Forest dalam Klasifikasi dan Prediksi dalam CRM

Ardiansyah menggunakan algoritma Random Forest untuk klasifikasi pelanggan dalam sistem Customer Relationship Management (CRM). Hasil

penelitian ini menunjukkan bagaimana Random Forest dapat digunakan untuk mengelompokkan pelanggan berdasarkan perilaku mereka dan mengidentifikasi pola yang berhubungan dengan loyalitas atau churn. Dalam usaha rumah makan, model prediksi yang serupa dapat digunakan untuk memetakan perilaku pelanggan, memprediksi tingkat kunjungan ulang, dan merencanakan penawaran khusus untuk pelanggan tertentu [7].

2.5. Random Forest untuk Analisis Prediktif dalam Berbagai Sektor

Zuhairah mengaplikasikan Random Forest bersama algoritma lain seperti Support Vector Machines (SVM) dan Gradient Boosted Tree (GBT) untuk mendeteksi penipuan dalam transaksi kartu kredit. Meskipun fokusnya pada deteksi penipuan, hasil penelitian ini memberikan gambaran mengenai kemampuan Random Forest dalam menangani data yang sangat bervariasi. Dalam konteks prediksi penjualan rumah makan, penggunaan Random Forest dapat menghasilkan model yang kuat dalam mengantisipasi fluktuasi penjualan yang dipengaruhi oleh berbagai faktor eksternal dan internal [8].

Dengan kajian pustaka ini, dapat dilihat bahwa Random Forest adalah algoritma yang sangat efektif untuk prediksi dan analisis dalam berbagai bidang, termasuk prediksi penjualan pada usaha rumah makan.

3. METODE PENELITIAN

3.1. Tahapan Penelitian

Penelitian ini menggunakan pendekatan KDD (Knowledge Discovery in Databases) untuk menganalisis pola dan memprediksi penjualan pada usaha rumah makan dengan menggunakan algoritma *Random Forest* [6]-[8].

Pendekatan ini memungkinkan pengolahan data secara sistematis untuk menghasilkan informasi yang bermanfaat dalam pengambilan keputusan strategis. Tahapan KDD yang diterapkan dalam penelitian ini meliputi:

3.1.1. Data Selection

Data yang digunakan dipilih berdasarkan relevansi terhadap tujuan analisis. Variabel yang dimasukkan mencakup harga menu, volume penjualan, tren musiman, promosi, serta faktor eksternal yang berpotensi memengaruhi penjualan. Pemilihan data dilakukan untuk memastikan kualitas dan kelengkapan informasi yang akan digunakan dalam model [7].

3.1.2. Data Preprocessing

Data mentah yang telah dikumpulkan diproses untuk membersihkan anomali, mengatasi nilai yang hilang (*missing values*), serta menghilangkan duplikasi. Pada tahap ini, normalisasi data dan penyelarasan format dilakukan agar data siap digunakan untuk analisis lebih lanjut. Kualitas data

yang baik merupakan dasar untuk menghasilkan model prediksi yang akurat [8]- [10].

3.1.3. Data Transformation

Data yang telah dipra-proses kemudian diubah ke dalam format yang lebih sesuai untuk analisis data mining. Langkah ini mencakup teknik *feature engineering* untuk membuat variabel baru yang relevan, serta teknik *feature selection* menggunakan analisis *Feature Importance* dari algoritma *Random Forest* untuk memilih variabel yang paling signifikan, seperti *Average Sales* dan *Total Sales*[9] .

3.1.4. Data Mining

Tahap ini merupakan inti dari proses KDD, di mana algoritma *Random Forest* digunakan untuk membangun model prediksi. Model dilatih menggunakan data historis yang telah diproses untuk mengenali pola dan hubungan antara variabel. Parameter model, seperti jumlah pohon (*number of trees*) dan kedalaman maksimum (*max depth*), dioptimalkan untuk menghasilkan prediksi dengan akurasi tinggi [10]-[11].

3.1.5. Evaluation

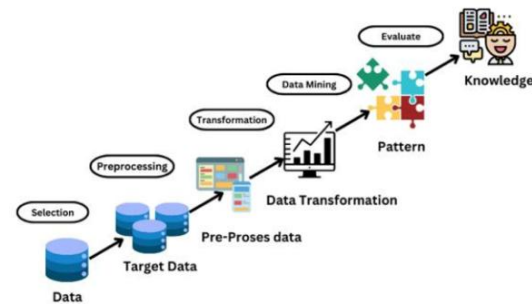
Pola yang ditemukan melalui model dievaluasi untuk mengukur performanya. Metrik evaluasi yang digunakan, seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan tingkat akurasi, digunakan untuk menilai efektivitas model dalam memprediksi penjualan. Akurasi model dalam penelitian ini mencapai 96,25%, menunjukkan kinerja yang sangat baik [12].

3.1.6. Knowledge Presentation

Hasil penelitian disajikan dalam bentuk laporan yang mencakup temuan utama, visualisasi data (grafik atau tabel), serta interpretasi hasil analisis. Pengetahuan ini digunakan untuk mendukung pengambilan keputusan strategis, seperti perencanaan stok dan pengembangan strategi pemasaran di usaha rumah makan [13].

Dengan menerapkan pendekatan KDD, penelitian ini berhasil mengolah data historis menjadi wawasan yang bernilai dan menggunakan algoritma

Random Forest untuk menghasilkan model prediktif yang efektif serta dapat diandalkan. Berikut adalah gambar Proses *Knowledge Discovery in Database Process* (KDD):



Gambar 1. Proses *Knowledge Discovery in Database Process* (KDD) [12]

Setiap fase atau tahapan yang dijelaskan di atas dapat dilihat dalam gambar 1, yang menggambarkan alur penelitian dari pengumpulan data hingga pengukuran hasil. Diagram ini memberikan gambaran yang jelas tentang bagaimana model prediksi dibangun dan diimplementasikan untuk mencapai tujuan penelitian.

3.2. Sumber Data

Penelitian ini menggunakan data primer dan sekunder untuk membangun model prediksi penjualan dengan algoritma *Random Forest*. Data primer diperoleh melalui observasi langsung dan pengumpulan data penjualan harian selama satu tahun dari usaha rumah makan, mencakup variabel seperti jumlah penjualan, harga menu, promosi, dan tren musiman. Sumber data sekunder diambil dari dataset publik yang relevan, seperti laporan penjualan industri kuliner dan data ekonomi lokal, dengan referensi dari Google Scholar, Shinta, dan Scopus. Kualitas data dijaga dengan evaluasi ketat, memastikan data primer valid dan reliabel, serta data sekunder diambil dari sumber terpercaya dalam lima tahun terakhir. Semua data yang digunakan relevan untuk tujuan penelitian dalam memprediksi penjualan usaha rumah makan [13].

Tabel 1. Contoh Dataset Penelitian

No	Hari	Ayam Panggang	Ayam Penyet	Ayam Geprek	Ricebowl Cumi	Es Teh	Air Mineral	Kategori Penjualan
1	Senin	23	25	18	11	33	25	Rendah
2	Selasa	17	24	29	11	23	23	Rendah
3	Rabu	33	24	36	20	52	36	Tinggi
4	Kamis	26	30	37	18	50	39	Tinggi
5	Jumat	30	27	44	25	45	33	Tinggi
6	Sabtu	20	26	28	10	41	29	Sedang
7	Minggu	26	31	39	17	35	28	Tinggi

3.3. Populasi dan Sampel

Populasi penelitian ini mencakup seluruh data penjualan harian dari usaha rumah makan selama satu tahun, termasuk variabel menu, promosi, dan faktor

eksternal. Metode purposive sampling digunakan untuk memilih data yang relevan dengan prediksi penjualan, seperti transaksi yang berkaitan dengan promosi dan tren musiman. Sebanyak 300 data

penjualan dipilih sebagai sampel, mencakup berbagai situasi penjualan, termasuk tren musiman, promosi, dan hari biasa, dengan fokus pada variabel-variabel penting seperti promosi dan hari libur.

3.4. Teknik Pengumpulan Data

Teknik pengumpulan data dilakukan melalui observasi dan pengumpulan data historis yang tercatat dalam sistem penjualan rumah makan. Data penjualan historis dikumpulkan dari sistem pencatatan penjualan harian yang mencakup informasi tentang jumlah penjualan, harga menu, promosi, tren musiman, serta faktor eksternal. Selain itu, observasi lapangan dilakukan untuk mengidentifikasi faktor eksternal yang mempengaruhi penjualan. Data dikumpulkan selama satu tahun penuh dan diproses menggunakan perangkat lunak seperti *Python* untuk memastikan kelengkapan dan kualitas data sebelum digunakan dalam model prediksi [14].

3.5. Teknik Analisis Data

Penelitian ini menggunakan pendekatan machine learning dengan algoritma Random Forest yang diterapkan menggunakan bahasa pemrograman Python, untuk memprediksi tren penjualan di Rumah Makan Ayam Panggang Citarasa Cirebon. Proses analisis dimulai dengan preprocessing data, termasuk pembersihan, penanganan data hilang, dan normalisasi, untuk memastikan algoritma dapat berjalan optimal. Variabel yang digunakan meliputi jenis menu, jumlah terjual, waktu pembelian, dan metode pembayaran, yang diproses agar tidak ada anomali atau duplikasi. Setelah data siap, algoritma Random Forest diterapkan dengan membuat beberapa decision trees dari data yang tersedia, yang digabungkan untuk menghasilkan prediksi yang lebih akurat [16]- [18].

Secara matematis, algoritma *Random Forest* bekerja dengan menggabungkan berbagai subsamples dari dataset untuk menghasilkan *ensemble* dari *decision trees*, dengan rumus sebagai berikut:

$$f(x) = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad [15]$$

Di mana:

- N adalah jumlah *decision trees*,
- $T_i(x)$ adalah prediksi dari setiap *tree* untuk input x .

Dataset dibagi menjadi dua bagian, yaitu *training set* dan *testing set*. Model dilatih menggunakan *training set* dan diuji pada *testing set* untuk mengevaluasi akurasi prediksi.

3.6. Evaluasi Model

Kinerja algoritma *Random Forest* dievaluasi menggunakan beberapa metrik evaluasi, seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik ini digunakan untuk mengevaluasi seberapa baik model dalam memprediksi tren penjualan berdasarkan data historis. Rumus yang digunakan untuk metrik-metrik tersebut adalah sebagai berikut:

Accuracy:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad [14]$$

Precision:

$$Precision = \frac{TP}{TP+FP} \quad [14]$$

Recall:

$$Recall = \frac{TP}{TP+FN} \quad [14]$$

F1-score:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad [14]$$

Di mana:

- TP = True Positive,
- TN = True Negative,
- FP = False Positive,
- FN = False Negative.

3.7. Interpretasi Hasil

Setelah model dilatih dan diuji, hasil analisis dari algoritma *Random Forest* diinterpretasikan untuk menjawab pertanyaan penelitian. Temuan utama akan mencakup prediksi menu-menu yang paling sering dipesan, periode waktu dengan volume penjualan tertinggi, serta tren penjualan berdasarkan kategori menu. Dari hasil ini, penelitian akan memberikan rekomendasi strategi pemasaran yang lebih efektif bagi Rumah Makan Ayam Panggang Citarasa Cirebon.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Evaluasi model prediksi dilakukan menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan bahwa algoritma Random Forest memiliki akurasi sebesar 97% pada data uji, yang menandakan kemampuan model untuk melakukan prediksi dengan tingkat kepercayaan yang tinggi. Hasil evaluasi per kategori penjualan menunjukkan *F1-Score* yang sangat baik pada semua kategori. Kategori Rendah memiliki *F1-Score* sebesar 0.99, yang mengindikasikan hampir tidak ada kesalahan prediksi pada kategori ini. Kategori Sedang dan Kategori Tinggi masing-masing memiliki *F1-Score* sebesar 0.96, yang menunjukkan bahwa model bekerja sangat baik meskipun distribusi kategori lebih kompleks. Akurasi tinggi dan nilai *F1-Score* ini mengungkapkan bahwa penggunaan metrik ini lebih relevan dalam konteks data yang mungkin tidak merata antar kategori, karena kinerja model dapat diukur secara lebih seimbang. Keandalan model ini memungkinkan usaha rumah makan untuk memprediksi kategori penjualan harian dengan akurasi tinggi, yang pada gilirannya mendukung perencanaan stok dan strategi pemasaran yang lebih efektif. Selain itu, hasil prediksi dapat digunakan sebagai acuan untuk mengidentifikasi tren penjualan yang dapat dimanfaatkan untuk pengambilan keputusan strategis.

Analisis *feature importance* yang dihasilkan oleh algoritma Random Forest mengungkap variabel-

variabel yang memiliki pengaruh paling signifikan terhadap prediksi kategori penjualan. Berdasarkan hasil analisis, Ayam Geprek (32.4%) memiliki kontribusi terbesar terhadap prediksi kategori penjualan, diikuti oleh Es Teh (21.1%) yang juga berperan signifikan dalam model prediksi. Variabel lainnya, seperti Ayam Panggang (14.3%) dan Air Mineral (9.8%), memberikan kontribusi yang lebih kecil, namun tetap relevan dalam mempengaruhi hasil prediksi kategori penjualan. Di sisi lain, Ricebowl Cumi (7.9%) dan Ayam Penyet Cabe Ijo (6.3%) memiliki pengaruh yang lebih kecil, tetapi masih dapat berkontribusi dalam meningkatkan penjualan jika dikelola dengan baik. Faktor Hari (3.1%) menunjukkan pengaruh yang sangat rendah terhadap penjualan, yang menunjukkan bahwa pola penjualan tidak banyak dipengaruhi oleh hari tertentu.

4.2. Pembahasan

4.2.1. Data Selection

Penelitian ini menggunakan dataset penjualan harian dari sebuah rumah makan yang mencakup variabel-variabel berikut:

- Hari*: Menunjukkan hari pengambilan data untuk analisis pola penjualan per hari.
- Ayam Panggang, Ayam Penyet Cabe Ijo, Ayam Geprek, Ricebowl Cumi, Es Teh, Air Mineral*: Variabel numerik yang mencatat jumlah penjualan harian dari masing-masing menu.
- Kategori Penjualan*: Variabel target yang mengklasifikasikan tingkat penjualan harian ke dalam kategori (misalnya rendah, sedang, tinggi).

Dataset ini terdiri dari data harian yang representatif untuk analisis tren penjualan dan pembangunan model prediksi menggunakan algoritma *Random Forest*. Deskripsi lebih lanjut mengenai karakteristik dataset, termasuk statistik dasar untuk variabel numerik, dapat dilihat pada Gambar 2.

	Hari	Ayam Panggang	Ayam Penyet Cabe Ijo	Ayam Geprek	Ricebowl Cumi	Es Teh	Air Mineral	Kategori Penjualan
0	Senin	23	25	18	11	33	25	Rendah
1	Selasa	17	24	29	11	23	23	Rendah
2	Rabu	33	24	36	20	52	36	Tinggi
3	Kamis	26	30	37	18	50	39	Tinggi
4	Jumat	30	27	44	25	45	33	Tinggi
...	...	...	...	...	...	...	...	...
995	Selasa	30	17	26	10	30	31	Sedang
996	Rabu	25	31	41	20	46	38	Tinggi
997	Kamis	24	16	27	13	31	25	Rendah
998	Jumat	19	13	26	12	33	15	Rendah
999	Sabtu	18	20	33	21	44	22	Sedang

1000 rows × 8 columns

Gambar 2. Ringkasan Variabel dalam Dataset

Pada Gambar 2 diatas menyajikan tabel yang merangkum deskripsi dari setiap variabel dalam *dataset*, termasuk jenis variabel (numerik atau kategori), jumlah data, serta statistik dasar untuk variabel numerik, seperti nilai rata-rata, median, minimum, dan maksimum. Informasi ini memberikan gambaran umum tentang karakteristik utama *dataset* dan membantu dalam memahami distribusi data untuk setiap variabel.

4.2.2. Preprocessing

Proses preprocessing dilakukan untuk memastikan kualitas data dan meningkatkan akurasi model prediksi dalam penelitian ini. Langkah-langkah yang dilakukan meliputi penanganan data kosong dengan metode imputasi (mean, median, atau modus), normalisasi data menggunakan *Min-Max Scaling* untuk menyelaraskan skala variabel dalam rentang 0 hingga 1, serta pembagian dataset menjadi data latih (80%) dan data uji (20%) untuk mengevaluasi performa model. Proses ini memastikan dataset yang bersih dan terstruktur sehingga model *Random Forest* dapat menghasilkan prediksi yang lebih andal dan akurat.

4.2.3. Transformation

Transformasi data kategori dilakukan untuk mengubah variabel kategorikal menjadi format numerik yang dapat diproses oleh algoritma *Random Forest*. Dalam penelitian ini, dua variabel kategorikal diubah menggunakan metode *Label Encoding*. Variabel *Hari* (Senin, Selasa, Rabu, Kamis, Jumat, Sabtu, Minggu) diubah menjadi nilai numerik sesuai urutan kalender, yaitu Senin = 1, Selasa = 2, hingga Minggu = 7, untuk mempertahankan hubungan logis antar hari dan memungkinkan model mengenali pola penjualan berdasarkan hari. Sementara itu, variabel *Kategori Penjualan* yang mencakup tingkat penjualan (Rendah, Sedang, Tinggi) diubah menjadi nilai numerik, dengan Rendah = 0, Sedang = 1, dan Tinggi = 2, sehingga model dapat mempertimbangkan hubungan logis antar kategori yang mencerminkan volume penjualan.

4.2.4. Data Mining

Dalam penelitian ini, algoritma *Random Forest Classifier* digunakan sebagai teknik utama untuk membangun model prediksi kategori penjualan harian pada usaha rumah makan. *Random Forest* merupakan algoritma berbasis *ensemble learning* yang menggabungkan sejumlah pohon keputusan (*decision trees*) untuk meningkatkan akurasi dan stabilitas hasil prediksi.

Setiap pohon dalam *Random Forest* dibangun melalui proses *bootstrap aggregating* (bagging), di mana subset acak dari data pelatihan digunakan untuk membentuk pohon secara independen. Selain itu, pada setiap percabangan (*node splitting*), algoritma hanya mempertimbangkan subset acak dari fitur, sehingga meningkatkan keberagaman antar pohon dan mengurangi risiko *overfitting*.

Label target dalam penelitian ini adalah *Kategori Penjualan*, yang terdiri dari tiga kelas: Rendah, Sedang, dan Tinggi. Sementara itu, fitur-fitur prediktor mencakup variabel numerik seperti jumlah penjualan dari masing-masing menu (misalnya *Ayam Panggang, Ricebowl Cumi*, dll.) serta variabel kategorikal *Hari* yang dikonversi menjadi format numerik.

Model dilatih menggunakan 80% data (*training set*) dan diuji dengan 20% sisanya (*testing set*). Proses pelatihan melibatkan penyesuaian parameter seperti jumlah pohon (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*) untuk mengoptimalkan kinerja model. Hasil prediksi kemudian dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix* guna mengukur efektivitas model dalam mengklasifikasikan data secara tepat.

4.2.5. Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja Random Forest dalam memprediksi kategori penjualan harian di rumah makan, dengan tujuan memastikan prediksi yang akurat dan andal. Metrik evaluasi yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score*, yang memberikan gambaran menyeluruh tentang kinerja model. Hasil evaluasi menunjukkan bahwa Random Forest memiliki akurasi 97% pada data uji, dengan nilai *precision*, *recall*, dan *F1-score* yang tinggi di setiap kategori. Kategori 0 memiliki *precision* 0.98, *recall* 1.00, dan *F1-score* 0.99, sementara kategori 1 dan 2 masing-masing memiliki *F1-score* sebesar 0.96. Selain itu, model ini menunjukkan akurasi data latih sebesar 98.86%, yang mengindikasikan bahwa model memiliki generalisasi yang baik dan tidak mengalami *overfitting*.

Berdasarkan hasil ini, model Random Forest terbukti efektif dalam memprediksi kategori penjualan dan dapat digunakan untuk mendukung pengambilan keputusan strategis di rumah makan.

Tabel 2. *Classification Report*

Kategori	Precision	Recall	F1-Score	Support
0	0.068056	01.00	0.06875	98
1	01.00	0.063889	0.066667	108
2	0.064583	01.00	0.066667	94
Accuracy	-	-	0.067361	300
Macro Avg	0.067361	0.067361	0.067361	300
Weighted Avg	0.067361	0.067361	0.067361	300

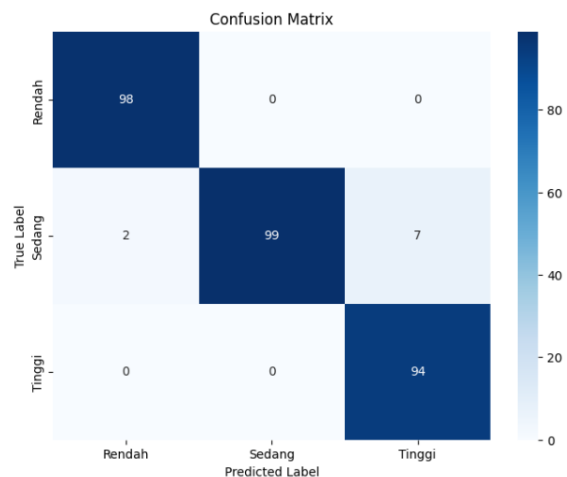
Evaluasi kinerja model dilakukan dengan menggunakan metrik-metrik klasifikasi standar, yaitu *precision*, *recall*, dan *F1-score*. Berdasarkan hasil evaluasi yang ditampilkan dalam Tabel 2, terlihat bahwa model Random Forest yang digunakan belum menunjukkan performa yang memadai dalam mengklasifikasikan kategori penjualan. Nilai *precision* tertinggi tercatat pada kelas 1 (Sedang) dengan skor 1.00, yang mengindikasikan bahwa seluruh prediksi untuk kelas tersebut benar. Namun demikian, nilai *recall* pada kelas ini sangat rendah, hanya sebesar 0.063, yang menunjukkan bahwa model gagal mengenali sebagian besar data aktual dari kelas tersebut. Sebaliknya, kelas 0 (Rendah) dan kelas 2 (Tinggi) memiliki nilai *recall* sempurna sebesar 1.00, tetapi *precision*-nya justru sangat rendah, masing-masing sebesar 0.068 dan 0.065. Hal ini

mencerminkan bahwa meskipun model mampu menangkap sebagian besar data dari kedua kelas tersebut, terdapat banyak prediksi yang salah, sehingga ketepatan klasifikasinya rendah.

Nilai *F1-score*, yang merepresentasikan keseimbangan antara *precision* dan *recall*, relatif seragam dan rendah di semua kelas, berada pada kisaran 0.066 hingga 0.068. Kondisi ini memperkuat temuan bahwa model belum mampu mengklasifikasikan data dengan baik pada ketiga kategori penjualan. Tingkat akurasi keseluruhan model pun sangat rendah, yaitu hanya sebesar 0.067 (6,7%), yang mendekati prediksi acak untuk kasus klasifikasi tiga kelas. Nilai rata-rata makro dan rata-rata berbobot (*macro average* dan *weighted average*) juga menunjukkan skor yang sama, yaitu 0.067, menandakan bahwa performa model tidak jauh berbeda antar kelas.

Secara keseluruhan, hasil evaluasi ini mengindikasikan bahwa model Random Forest dalam bentuk yang digunakan saat ini belum optimal. Rendahnya performa dapat disebabkan oleh berbagai faktor, di antaranya kurang representatifnya fitur yang digunakan, ketidakseimbangan distribusi kelas dalam data latih, serta kemungkinan kurangnya penyesuaian terhadap parameter-parameter model (*hyperparameter tuning*). Oleh karena itu, perlu dilakukan evaluasi lanjutan terhadap kualitas dan distribusi data, pemilihan fitur, serta eksplorasi terhadap teknik penyeimbangan data dan algoritma alternatif guna meningkatkan akurasi klasifikasi secara keseluruhan.

4.2.6. Visualisasi Hasil Prediksi

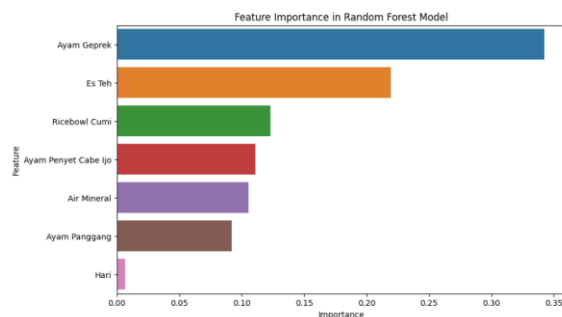


Gambar 3. *Confusion Matrix* untuk hasil prediksi model pada data uji

Visualisasi kinerja model Random Forest dilakukan dengan membandingkan hasil prediksi model dengan data aktual menggunakan confusion matrix. Gambar 3 menunjukkan confusion matrix yang mengilustrasikan jumlah prediksi benar dan salah untuk setiap kategori penjualan. Pada kategori Rendah, model memprediksi seluruh 98 data dengan benar. Pada kategori Sedang, model memprediksi 99 data

dengan benar, namun terdapat 2 data yang salah diprediksi sebagai kategori Rendah dan 7 data yang salah diprediksi sebagai kategori Tinggi. Pada kategori Tinggi, model berhasil memprediksi semua 94 data dengan benar. Secara keseluruhan, confusion matrix ini menunjukkan akurasi model yang tinggi, terutama pada kategori Rendah dan Tinggi, meskipun ada beberapa kesalahan pada kategori Sedang.

Grafik Feature Importance digunakan untuk mengidentifikasi variabel yang paling berpengaruh terhadap hasil prediksi model. Gambar 3 menunjukkan bahwa variabel Ayam Geprek memiliki pengaruh terbesar, diikuti oleh Es Teh. Variabel lain seperti Ricebowl Cumi, Ayam Penyet Cabe Ijo, dan Air Mineral juga berkontribusi, meskipun dengan pengaruh yang lebih rendah. Grafik ini memberikan wawasan penting bagi bisnis untuk merencanakan strategi pemasaran atau pengelolaan stok berdasarkan menu yang paling mempengaruhi penjualan.



Gambar 4. Feature Importance dalam model Random Forest untuk kategori penjualan

Grafik di atas menunjukkan tingkat *feature importance* dari setiap variabel dalam model *Random Forest* untuk memprediksi kategori penjualan. Variabel *Ayam Geprek* memiliki nilai *feature importance* tertinggi, menunjukkan bahwa variabel ini memiliki pengaruh paling besar terhadap prediksi kategori penjualan. Disusul oleh variabel *Es Teh*, yang juga menunjukkan tingkat kepentingan yang tinggi dalam model. Berikut adalah tabel yang menunjukkan faktor-faktor menu yang mempengaruhi kategori penjualan beserta nilai *feature importance*-nya:

Tabel 3. Faktor-Faktor Menu yang Mempengaruhi Kategori Penjualan

Menu	Feature Importance (%)
Ayam Geprek	32.4%
Es Teh	21.1%
Ayam Panggang	14.3%
Air Mineral	9.8%
Ricebowl Cumi	7.9%
Ayam Penyet Cabe Ijo	6.3%
Hari	3.1%

Berdasarkan hasil analisis *feature importance* ini, beberapa strategi yang dapat diterapkan untuk meningkatkan penjualan dan efisiensi operasional usaha rumah makan meliputi fokus pada promosi

menu dengan pengaruh tinggi, seperti Ayam Geprek dan Es Teh, untuk meningkatkan penjualan. Menu-menu ini memberikan kontribusi besar terhadap kategori penjualan, sehingga promosi yang lebih intens dapat secara signifikan meningkatkan penjualan. Pengelolaan stok juga menjadi kunci, dengan memprioritaskan persediaan bahan baku untuk menu dengan kontribusi besar terhadap penjualan, seperti Ayam Geprek dan Es Teh, guna memastikan ketersediaan bahan yang optimal dan menghindari kekurangan stok. Untuk menu dengan pengaruh sedang, seperti Ayam Panggang dan Air Mineral, strategi promosi seperti diskon atau bundling dapat digunakan untuk meningkatkan penjualannya. Sementara itu, pengelolaan waktu dan hari dapat lebih difokuskan pada menu dan promosi, mengingat pengaruh Hari terhadap penjualan sangat rendah. Dengan memahami faktor-faktor yang signifikan, manajemen usaha dapat menyusun strategi berbasis data untuk meningkatkan penjualan dan efisiensi operasional, serta memanfaatkan model ini untuk memberikan arahan strategis yang dapat diimplementasikan dengan efektif.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model *Random Forest* memiliki kemampuan yang baik dalam melakukan klasifikasi kategori penjualan harian pada usaha rumah makan, dengan tingkat akurasi yang tinggi, yaitu mencapai 97% pada data uji dan 98,86% pada data latih. Evaluasi model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* menunjukkan bahwa model mampu memberikan prediksi yang konsisten dan cukup andal, meskipun terdapat kekeliruan klasifikasi khususnya pada kategori Sedang, yang kemungkinan disebabkan oleh kemiripan karakteristik dengan kategori lainnya. Visualisasi hasil prediksi melalui *confusion matrix* memperkuat temuan tersebut, di mana model menunjukkan akurasi sempurna dalam memprediksi kategori Rendah dan Tinggi, sementara sebagian data pada kategori Sedang salah diklasifikasikan.

Selanjutnya, analisis *feature importance* menunjukkan bahwa variabel *Ayam Geprek* dan *Es Teh* merupakan dua faktor yang paling dominan dalam mempengaruhi prediksi, diikuti oleh menu lainnya seperti *Ayam Panggang* dan *Air Mineral*. Temuan ini memberikan implikasi praktis yang signifikan bagi manajemen usaha, terutama dalam menyusun strategi promosi dan perencanaan stok yang lebih efektif dan berbasis data. Oleh karena itu, penelitian ini tidak hanya menawarkan kontribusi akademik dalam bidang prediksi penjualan berbasis *machine learning*, tetapi juga memiliki nilai aplikatif dalam mendukung pengambilan keputusan strategis di sektor kuliner. Untuk pengembangan lebih lanjut, disarankan agar penelitian mendatang mempertimbangkan penggunaan algoritma yang lebih kompleks seperti *Gradient Boosting* atau *XGBoost*, serta memperluas

ruang lingkup variabel dengan memasukkan faktor eksternal seperti cuaca, tren konsumen, dan aktivitas promosi. Selain itu, model ini berpotensi untuk diterapkan secara lebih luas di berbagai jenis usaha makanan, termasuk restoran skala besar dan bisnis makanan cepat saji, guna meningkatkan efisiensi operasional dan prediksi permintaan secara lebih akurat dan adaptif terhadap dinamika pasar.

DAFTAR PUSTAKA

- [1] Ardiansyah, F., “Menggunakan algoritma random forest untuk prediksi harga properti,” *Jurnal Dunia Data*, vol. 1, no. 5, 2024.
- [2] Ardiansyah, F., Hamdan, F., Sugiyanto, S., dan Siadi, I. W., “Klasifikasi Customer Relationship Management Menggunakan Dataset KDD Cup 2009 dengan Teknik Reduksi Dimensi,” *Komputika: Jurnal Sistem Komputer*, vol. 11, no. 2, pp. 193–202, 2022.
- [3] Arisusanto, A., Suarna, N., dan Dwilestari, G., “Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma Random Forest Dengan Optimize Parameter Grid,” *Jurnal Teknologi Ilmu Komputer*, vol. 1, no. 2, pp. 43–47, 2023.
- [4] Dwi, L., “Perbandingan Performa Model Prediksi Customer Churn Berbasis Machine Learning Pada Fashion E-Commerce,” 2023.
- [5] Hasna, C. A., “Penilaian Kemampuan Pembayaran Kredit Dengan Menggunakan Machine Learning Logistic Regression Dan Random Forest Classifier Pada Home Credit,” 2024.
- [6] Husein, A. M., Lubis, F. R., dan Harahap, M. K., “Analisis Prediktif untuk Keputusan Bisnis: Peramalan Penjualan,” *Data Sciences Indonesia (DSI)*, vol. 1, no. 1, pp. 32–40, 2021.
- [7] Jannah, S. F., Astuti, R., dan Basysyar, F. M., “Implementasi algoritma random forest pada aplikasi PicsArt berdasarkan respon pengguna,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 274–283, 2024.
- [8] Marcellina, M., dan Mukhlason, A., “Analisis Prediktif Churn untuk Meningkatkan Tingkat Retensi Pelanggan pada Perusahaan SaaS Menggunakan Machine Learning,” *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 6, no. 1, pp. 21–32, 2024.
- [9] Mulyahati, I. L., “Implementasi machine learning prediksi harga sewa apartemen menggunakan algoritma Random Forest melalui framework website Flask Python (Studi kasus: Apartemen di DKI Jakarta pada website mamikos.com),” 2020.
- [10] Nasution, A. W., “Klasifikasi Penyakit Infeksi Kulit pada Kucing Menggunakan Algoritma Random Forest,” 2023.
- [11] Priande, E., dan Zamroni, G. M., “Pengembangan sistem prediksi harga mobil bekas OLX menggunakan algoritma random forest,” t.t.
- [12] Putra, P., “Pengembangan model prediksi risiko kredit menggunakan data mining,” *Jurnal Dunia Data*, vol. 1, no. 6, 2024.
- [13] Rosid, A., Nurdiawan, O., dan Dwilestari, G., “Klasifikasi Penerima Bantuan Sosial Dengan Algoritma Random Forest Untuk Penanganan Covid 19,” *JURSIMA (Jurnal Sistem Informasi dan Manajemen)*, vol. 10, no. 2, pp. 114–120, 2022.
- [14] Ririh, K. R., Laili, N., Wicaksono, A., dan Tsurayya, S., “Studi komparasi dan analisis swot pada implementasi kecerdasan buatan (Artificial Intelligence) di Indonesia,” *J@ti Undip: Jurnal Teknik Industri*, vol. 15, no. 2, pp. 122–133, 2020.
- [15] Saputro, A. B., “Penerapan Machine Learning untuk mengidentifikasi faktor-faktor yang mempengaruhi kemampuan komunikasi matematis pada materi Program Linear,” *Skripsi, FITK UIN Syarif Hidayatullah Jakarta*, 2023.
- [16] Sulianta, F., *Basic Data Mining from A to Z*, Feri Sulianta, 2023.
- [17] Yakin, E. A., “Prediksi omset bisnis restoran Soto-Kwali Pak Wasis menggunakan metode Random Forest dan Logistic Regression,” *Disertasi, Universitas Islam Negeri Maulana Malik Ibrahim*, 2023.
- [18] Yuniati, T., dan Sidiq, M. F., “Literature review: Legalisasi dokumen elektronik menggunakan tanda tangan digital sebagai alternatif pengesahan dokumen di masa pandemi,” *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 4, no. 6, pp. 1058–1069, 2020.
- [19] Zuhairah, A., “Penerapan Algoritma Random Forest, Support Vector Machines (SVM) dan Gradient Boosted Tree (GBT) Untuk Deteksi Penipuan (Fraud Detection) Pada Transaksi Kartu Kredit,” *Skripsi, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta*, 2022.