



IMPLEMENTASI ALGORITMA *RANDOM FOREST REGRESSION* UNTUK MEMPREDIKSI PENJUALAN PRODUK DI SUPERMARKET

Rahmat Hidayat¹, Haris Tri Saputra², Mirdatul Husnah³, Nabila⁴, M Bintang Hidayatullah⁵, Muhammad Naufal Nazhmi⁶, Jauzaa Azra⁷, Astri Rana⁸

¹rahmat.hidayat@lecturer.unri.ac.id, ²haristrisaputra@lecturer.unri.ac.id,

³mirdatulhusnah@lecturer.unri.ac.id, ⁴nabila3905@student.unri.ac.id,

⁵m.bintang3908@student.unri.ac.id, ⁶muhammad.naufal5934@student.unri.ac.id,

⁷jauzaa.azra5942@student.unri.ac.id, ⁸astri.rana5943@student.unri.ac.id

¹Prodi Manajemen Informatika, FMIPA, Universitas Riau

^{2,3,4,5,6,7,8}Prodi Sistem Informasi, FMIPA, Universitas Riau

Abstrak

Prediksi penjualan merupakan aspek penting dalam pengelolaan operasional Supermarket. Algoritma *Machine Learning* dapat dimanfaatkan untuk menghasilkan prediksi yang akurat. Supermarket mencatat transaksi penjualan setiap harinya, data transaksi tersebut dapat dimanfaatkan untuk mendapatkan informasi dalam memperoleh keuntungan. Pada penelitian ini, algoritma *Random Forest Regression* digunakan dengan tujuan untuk memprediksi penjualan di Supermarket. Data dalam dataset berjumlah 1000 data. Proses *preprocessing* data dimulai dengan menentukan fitur yang paling relevan sebagai variabel independen dan variabel dependen. Selanjutnya, entri kosong pada data numerik diisi dengan nilai rata-rata (mean), sedangkan pada data kategori, entri kosong diisi dengan nilai modus. Evaluasi kinerja model algoritma diukur dengan menggunakan beberapa metrik, yaitu *Out-of-Bag (OOB) score* sebesar 0.9999, *Mean Squared Error (MSE)* sebesar 2.4899, *R-squared* atau koefisien determinasi mencapai 0.9999, dan *Mean Absolute Error (MAE)* bernilai 0.9305. Secara keseluruhan, nilai-nilai metrik menunjukkan model *Random Forest Regression* sangat akurat untuk prediksi penjualan di supermarket.

Kata kunci: Supermarket, Prediksi, Produk, *Random Forest Regression*.

Abstract

Sales prediction is an important aspect in managing supermarket operations. Machine Learning algorithms can be used to produce accurate predictions. Supermarkets record sales transactions every day, this transaction data can be used to obtain information to gain profits. In this research, the Random Forest Regression algorithm is used with the aim of predicting sales in supermarkets. The dataset used was obtained from the Kaggle website and contains individual transaction records from a supermarket chain in Myanmar. The data in the dataset contains 1000 data. The data preprocessing process begins by determining the most relevant features as independent variables and dependent variables. Next, empty entries in numeric data are filled with the average value (mean), while in category data, empty entries are filled with the mode value. Evaluation of the performance of the algorithm model is measured using several metrics, namely Out-of-Bag (OOB) Score of 0.9999, Mean Squared Error (MSE) of 2.4899, R-squared or coefficient of determination reaching 0.9999, and Mean Absolute Error (MAE) of 0.9305.

Keyword: Supermarket, Prediction, Product *Random Forest Regression*.

1. Pendahuluan

Dunia bisnis mengalami perubahan signifikan di tengah pesatnya perkembangan teknologi dan transformasi digital yang mengharuskan perusahaan terus berinovasi demi mempertahankan dan meningkatkan persaingan [1]. Seiring dengan kemajuan zaman, persaingan dalam dunia bisnis dan teknologi semakin meningkat. Perusahaan atau pelaku bisnis membutuhkan teknologi untuk mengolah data yang ada agar dapat mendukung jalannya bisnis. Perusahaan harus memiliki kemampuan untuk melakukan analisis dan meramalkan berbagai kemungkinan yang akan terjadi di masa depan [2]. Dengan meramalkan kemungkinan yang akan terjadi, perusahaan dapat membuat keputusan yang efektif dan efisien.

Supermarket atau pasar swalayan disebut juga toko ritel yang memberikan konsep belanja lengkap dengan menyediakan beragam produk dan barang kebutuhan sehari-hari di satu lokasi [3]. Berbeda dengan toko ritel tradisional, supermarket umumnya memiliki jumlah stok yang lebih besar dan variasi produk yang lebih luas. Supermarket melakukan transaksi penjualan setiap harinya, banyaknya data transaksi yang diperoleh membuat data tersebut dapat dimanfaatkan untuk mendapatkan informasi dalam memperoleh keuntungan [4].

Data *history* (data masa lalu) yang ada dapat dimanfaatkan menggunakan metode analisis sehingga menghasilkan pola penjualan tertentu untuk peramalan atau prediksi [5]. Tujuan dari penentuan pola ini adalah untuk menentukan perkiraan volume dalam penjualan serta mengidentifikasi pasar potensial yang akan didominasi dimasa depan. Prediksi penjualan yang tepat memungkinkan pihak manajemen untuk mengoptimalkan stok produk, mengurangi risiko kerugian akibat kelebihan atau kekurangan persediaan, serta merencanakan promosi dengan lebih baik. Prediksi telah digunakan sebagai alat atau salah satu bahan pertimbangan dari pengambilan keputusan, terutama di sektor bisnis atau ekonomi dimana kerugian diminimalisir sebisa mungkin dan keuntungan atau laba diraih semaksimal mungkin [2]. Prediksi merupakan kegiatan mengidentifikasi suatu hal pada periode mendatang berdasarkan pada deskripsi orang lain/benda atau hubungan antara suatu hal yang dapat diketahui saat ini [6].

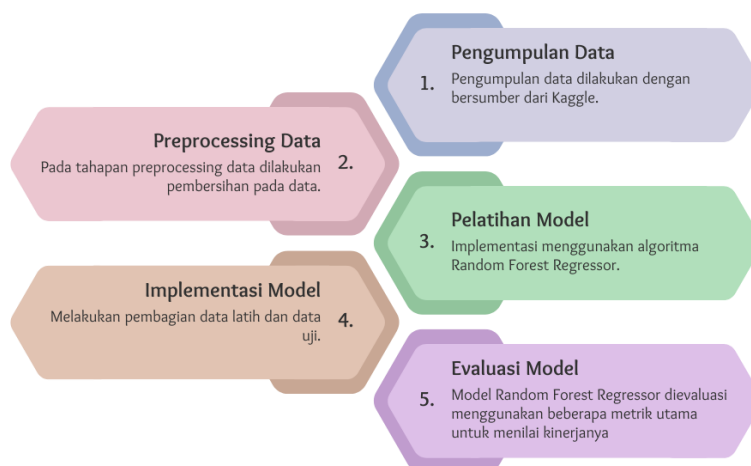
Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* [7]. Penggunaan metode data mining dapat membantu dalam menganalisa prediksi penjualan di Supermarket dengan menggunakan metode *Random Forest Regression* untuk melakukan peramalan atau prediksi. *Random Forest Regression* merupakan salah satu model regresi yang didasarkan pada teknik *supervised learning* menggunakan *history* data [8]. Metode ini merupakan jenis *ensemble learning*, yang berarti menggabungkan beberapa model untuk mendapatkan hasil prediksi yang lebih akurat. Pada dasarnya, model *Random Forest Regression* dibuat dengan menggabungkan beberapa Pohon Keputusan (*Decision Tree*), dimana setiap pohon (*tree*) bertindak sebagai pengambil keputusan independen [9]. Kelebihan dari *Random Forest* adalah kemampuannya dalam menangani ketergantungan dan interaksi antara variabel yang kompleks serta mampu mengatasi *overfitting* [10].

Model yang dilatih menggunakan algoritma random forest dapat dievaluasi keakuratannya menggunakan beberapa metrik evaluasi seperti *Mean Square Error* (MSE), *Mean Absolute Error* (MAE), dan R-Squared. MSE adalah pengestimasian nilai *error* pada persamaan dengan menjumlahkan kuadrat seluruh kesalahan atau error peramalan dari setiap periode dan membaginya dengan jumlah periode peramalan yang dijadikan data uji [5]. MAE (*Mean Absolute Error*) adalah salah satu metrik evaluasi dalam *machine learning* yang digunakan untuk mengukur seberapa dekat prediksi model dengan nilai aktual. Metrik ini menghitung rata-rata dari nilai absolut selisih antara prediksi dan nilai sebenarnya, tanpa mempertimbangkan arah kesalahan (positif atau negatif) [11]. R-squared (R^2) mengukur seberapa baik model dapat menjelaskan variabilitas dalam data. Nilai R^2 berada dalam rentang 0 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa model sangat baik dalam menjelaskan variasi data [12].

Penelitian peramalan atau prediksi dengan menggunakan algoritma *Random Forest Regression* sebelumnya pernah dilakukan pada penelitian yang berjudul “Model *Random Forest Regression* Untuk Peramalan Penyebaran Covid-19 di Indonesia”. Hasil perhitungan evaluasi MAE pada model *Random Forest Regression* penelitian tersebut tidak besar, yaitu sebesar 6.477 untuk peramalan *New Cases* dan 0.2469 untuk peramalan *New Deaths* yang artinya hasil nilai yang diramalkan tidak berbeda jauh dari nilai aktual [8]. Pada Penelitian Green Arter Sandag yang membahas mengenai prediksi rating aplikasi App Store dengan menggunakan metode *Random Forest* yang menggunakan dataset 7198 rows dan 16 attributes. Hasil dari pengujian yang didapatkan nilai *accuracy* 86.27%, *preccission* 84.64% dan nilai RMSE 0.313 [13]. Penelitian lain juga pernah dilakukan oleh Evita Fitri dengan melakukan pengujian metode *Linear Regression*, *Random Forest Regression* dan *Gradient Boosted Trees Regression* pada dataset harga rumah, perbandingan prediksi dengan *error rate* lebih kecil terdapat pada hasil pada metode *Random Forest Regression*. Evaluasi prediksi dilakukan dengan melihat hasil error pada *Root Mean Squared Error* (RMSE) setiap model, yaitu pada *Random Forest Regression* didapat nilai error 0.440, *Linear Regression* didapat nilai error terkecil di range 0.515 dan pada *Gradient Boosted Trees Regression* sebesar 0.508. Nilai akurasi prediksi harga rumah menggunakan metode *Random Forest Regression* menghasilkan akurasi tertinggi sebesar 81,5% dibandingkan dengan metode *Linear Regression* dan *Gradient Boosted Trees Regression* [14].

Berdasarkan dari penelitian sebelumnya, algoritma *Random Forest Regression* diterapkan untuk memprediksi penjualan pada supermarket. Penggunaan algoritma *Random Forest Regression* ini didasarkan pada hasil dari beberapa penelitian terkait algoritma yang memberikan performa bagus untuk mengatasi permasalahan prediksi atau regresi lain. Untuk menghitung skor prediksi supermarket, digunakan evaluasi MAE, *OOB Score*, MSE, dan *R-squared*.

2. Metode Penelitian



Gambar 1. Alur Metode Penelitian

Penelitian ini memiliki 5 alur metode penelitian yang dimulai dari pengumpulan data, *preprocessing* data, pelatihan model, implementasi model, dan evaluasi model. Penjelasan masing-masing alur akan dijabarkan pada bab ini. Gambar 1 merupakan alur metode penelitian. Pada penelitian ini beberapa tahapan dilakukan untuk memprediksi total penjualan supermarket di Myanmar. Data ini diambil dari *Supermarket Sales Dataset* oleh Fares Ashraf pada tahun 2023 di Kaggle dan berisi catatan transaksi individu dari sebuah jaringan supermarket di Myanmar dengan rincian untuk setiap penjualan. Sebelum data diolah, dilakukan tahap *preprocessing* agar data dapat dilatih menjadi sebuah model prediksi. Algoritma yang digunakan dalam model ini adalah *Random Forest Regressor*, dan model yang telah diimplementasikan dievaluasi menggunakan beberapa metrik, yaitu *OOB Score*, MSE, R^2 , serta MAE, guna mengukur kinerja prediksi model terlebih dahulu dengan beberapa detail atribut.

2.1. Pengumpulan Data

Seleksi atau pemilihan data perlu dilakukan dari himpunan data operasional sebelum tahapan penggalian informasi dalam *Knowledge Discovery in Databases* dimulai [15]. Data dari penelitian ini diambil dari situs *Kaggle* pada bulan September 2024. Dataset ini berisi data transaksi individu dari sebuah jaringan Supermarket di Myanmar. Dataset mencakup berbagai informasi terkait transaksi, seperti identitas faktur, lokasi cabang, jenis pelanggan, kategori produk, jumlah dan harga barang, pajak, total transaksi, metode pembayaran, serta indikator biaya dan keuntungan. Contoh dataset yang digunakan pada penelitian ini dapat dilihat pada Tabel 1. Selain itu, terdapat pula penilaian kepuasan pelanggan pada setiap transaksi. Dataset transaksi tersebut selanjutnya akan diolah untuk dilakukan proses *preprocessing* data pada langkah berikutnya.

2.2. Preprocessing Data

Data *pre-processing* adalah berbagai langkah atau proses yang dilakukan pada data sebelum data tersebut digunakan dalam analisis atau pemodelan [16]. Tujuan dari *pre-processing* ini adalah untuk meningkatkan kualitas *output* model dengan melakukan *cleaning* pada data, seperti memperbaiki atau menghapus data yang rusak atau tidak relevan [17]. Tahap ini dilakukan untuk memastikan data dapat diolah secara optimal oleh model *machine learning*. Proses dimulai dengan analisis untuk menentukan fitur

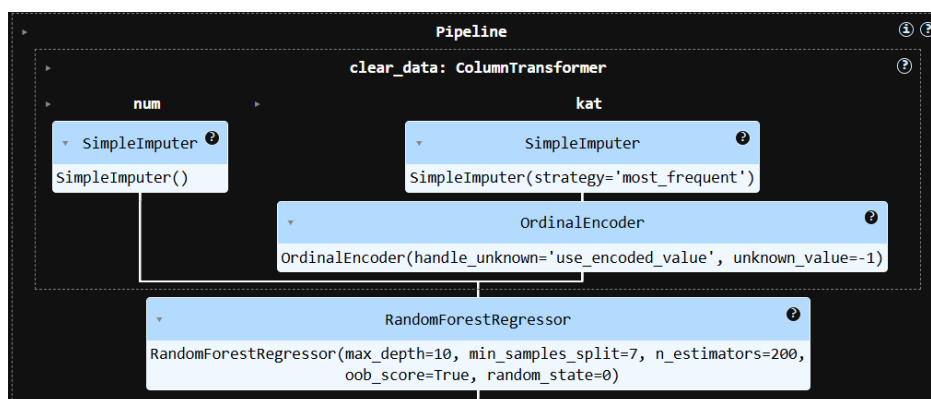
yang paling relevan sebagai variable independen dan variable dependen. Selanjutnya, data dipisah berdasarkan tipe, yaitu data dengan tipe objek dan data dengan tipe numerik. Pada data numerik, entri kosong diisi dengan nilai rata-rata (*mean*), sedangkan pada data kategori, entri kosong diisi dengan nilai modus, dan data kategori dikonversi menjadi numerik menggunakan *Ordinal Encoder*, kemudian semua proses tersebut disatukan dalam *pipeline* dengan variabel *preprocessor* menggunakan *column transformer*. Untuk melakukan prediksi model data dibagi menjadi dua yaitu data latih dan data uji, karena untuk memprediksi diperlukan data baru yang tidak dikenali oleh model [18].

Tabel 1. Contoh Dataset

Branch	City	Customer Type	Gender	Product Line	Unit Price	Quantity	Tax 5%
Alex	Yangon	Member	Male	Home and lifestyle	63.56	10	31.7800
Cairo	Mandalay	Member	Male	Sports and travel	73.97	1	3.6985
Alex	Yangon	Normal	Male	Fashion accessories	77.02	5	19.2550
Cairo	Mandalay	Normal	Female	Food and beverages	67.77	1	3.3885
Alex	Yangon	Normal	Male	Food and beverages	33.88	8	13.5520

2.3. Pelatihan Model

Algoritma yang digunakan dalam penelitian ini adalah *Random Forest Regressor*. Beberapa pengaturan parameter yang digunakan yaitu, yaitu *n_estimators*, yang menunjukkan jumlah pohon keputusan; *max_depth*, yang membatasi kedalaman maksimum pohon untuk mencegah *overfitting*; dan *min_samples_split*, yang mengatur jumlah minimum sampel yang diperlukan untuk memisahkan simpul. Model ini juga diatur dengan *random_state* untuk konsistensi hasil serta *oob_score* untuk menghitung skor prediksi menggunakan sampel *out-of-bag* sebagai bagian dari validasi model. Proses *preprocessing* dan pembentukan model digabungkan dalam sebuah *pipeline* untuk memastikan semua langkah diterapkan konsisten pada data latih dan data uji. Gambar 2 merupakan *pipeline preprocessing* dan model yang telah diterapkan.



Gambar 2. Pipeline Preprocessing dan Model.

2.4. Implementasi Model

Setelah model dilatih menggunakan data latih (*x_train* dan *y_train*), langkah selanjutnya adalah menguji model tersebut dengan data uji (*x_test*). Tujuan dari pengujian ini adalah untuk menilai kemampuan model dalam memprediksi hasil pada data yang tidak digunakan selama pelatihan, sehingga dapat dievaluasi sejauh mana model mampu menggeneralisasi dan menghasilkan prediksi yang akurat di luar data latih.

2.5. Evaluasi Model

Model *Random Forest Regressor* dievaluasi menggunakan beberapa metrik utama untuk menilai kinerjanya. Tujuannya adalah menarik kesimpulan dari hasil penerapan algoritma untuk memahami kinerja, keberhasilan, serta mengidentifikasi pola-pola atau karakteristik yang muncul selama proses learning [19]. Metrik pertama, *OOB Score*, mengukur akurasi model dengan data sampel yang tidak terlibat dalam pelatihan, memberikan indikasi kemampuan generalisasi model. MSE digunakan untuk mengidentifikasi rata-rata kuadrat kesalahan antara prediksi dan nilai aktual, memberikan bobot lebih pada kesalahan besar. *R-squared* mengukur seberapa baik model dapat menjelaskan variabilitas dalam data target, yang menggambarkan kekuatan hubungan antara variabel independen dan dependen. Terakhir, MAE mengukur rata-rata kesalahan absolut pada prediksi, yang secara langsung menggambarkan tingkat kesalahan dalam satuan yang sama dengan data target.

3. Hasil dan Pembahasan

3.1. Hasil

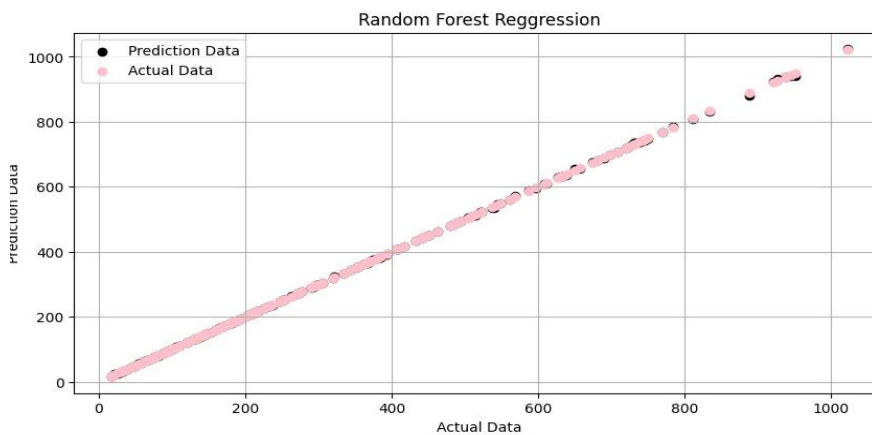
Kode program menggunakan bahasa Python diterapkan pada tahapan ini. Tabel 2 merupakan sebuah baris kode yang menghasilkan model prediksi penjualan supermarket yang terstruktur dengan baik, mulai dari pemilihan fitur hingga penerapan algoritma *Random Forest Regression*. Proses yang dilakukan memastikan data diolah secara optimal, sehingga model mampu menghasilkan prediksi yang akurat dan andal. Dimulai dengan memilih fitur-fitur yang relevan untuk model, seperti informasi tentang cabang, kota, jenis pelanggan, harga produk, dan lainnya, yang kemudian dijadikan variabel independen (x), sementara variabel dependen (y) adalah total penjualan. Selanjutnya, dataset dibagi menjadi data latih (80%) dan data uji (20%) untuk memastikan model dilatih dengan data yang cukup dan dievaluasi dengan data yang terpisah. Proses imputasi juga diterapkan pada data numerik yang memiliki nilai hilang dengan menggunakan strategi imputasi rata-rata untuk mengisi kekosongan tersebut. Selain itu, fitur kategori juga diproses dengan menggunakan imputasi modus dan encoding ordinal.

Tabel 2. Kode Program Phyton

Kode	Deskripsi
<pre>features = ['Branch', 'City', 'Customer type', 'Gender','Product line', 'Unit price', 'Quantity', 'Tax 5%', 'gross income', 'Date', 'Payment', 'cogs', 'Rating'] x = data[features] y = data['Sales'] from sklearn.model_selection import train_test_split</pre>	<p>Memilih variabel independen (x) dan variabel dependen (y).</p> <p>Membagi dataset menjadi set pelatihan dan pengujian dengan proporsi 80:20.</p>
<pre>x_train, x_test, y_train, y_test = train_test_split(x, y, train_size=0.8, test_size=0.2, random_state=0) kol_ubah_num = SimpleImputer(strategy='mean') kol_ubah_kat = Pipeline(steps=[('imputer', SimpleImputer(strategy='most_frequent')), ('ordinal', OrdinalEncoder(...))])</pre>	<p>Mendefinisikan strategi imputasi mean untuk kolom numerik.</p> <p>Membuat pipeline untuk mengatasi kolom kategori dengan imputasi modus dan encoding ordinal.</p>

<pre>from sklearn.ensemble import RandomForestRegressor</pre>	Mendefinisikan model Random Forest Regressor sebagai model prediksi.
<pre>model = RandomForestRegressor (n_estimators=200, max_depth=10, min_samples_split=7, random_state=0, oob_score=True) pipeline.fit(x_train, y_train)</pre>	Melatih model menggunakan data latih.
<pre>predik = pipeline.predict(x_test)</pre>	Memprediksi nilai target untuk data uji.

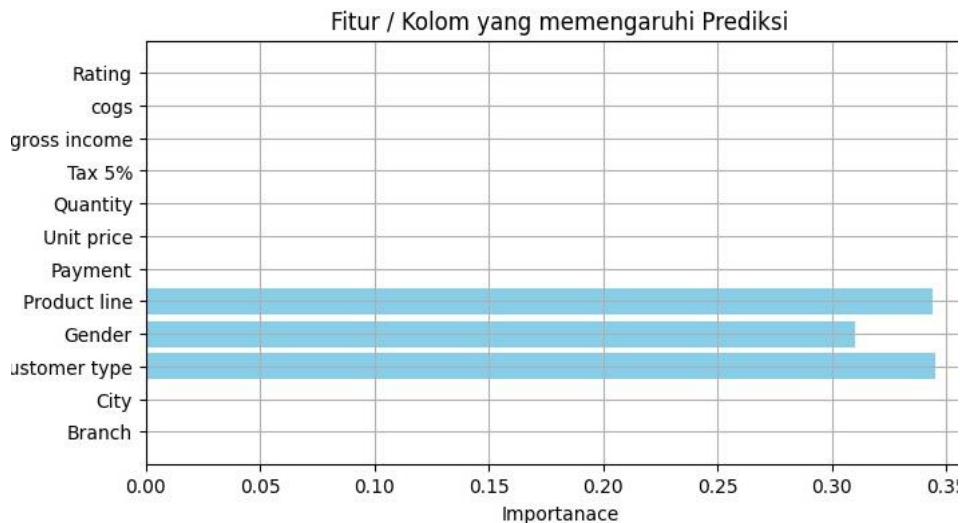
Setelah tahap persiapan data, model *Random Forest Regression* diterapkan dengan sejumlah parameter yang dioptimalkan untuk meningkatkan akurasi prediksi. Jumlah estimasi pohon (*n_estimators*) diatur sebanyak 200 pohon, yang memungkinkan model untuk belajar dari berbagai subset data dan meningkatkan stabilitas prediksi. Kedalaman maksimum pohon (*max_depth*) dibatasi hingga 10, untuk mencegah model terlalu mendalam yang bisa menyebabkan *overfitting*, serta memastikan model tetap generalisasi dengan baik. Selain itu, parameter *min_samples_split* diatur pada nilai 7, yang berarti pohon akan membagi node hanya jika ada minimal 7 sampel dalam satu node. Hal ini membantu mengontrol kompleksitas pohon dan mencegah model terlalu rumit. Pengaturan parameter ini bersama dengan *OOB scoring* memungkinkan model untuk melakukan evaluasi internal selama pelatihan tanpa memerlukan data uji terpisah, memberikan gambaran tentang performa model tanpa risiko *overfitting*. Model ini kemudian dilatih dengan data latih dan diuji dengan data uji, menghasilkan prediksi yang dapat dievaluasi menggunakan metrik seperti *OOB*, *MSE*, dan *R-squared*.



Gambar 3. Grafik Perbandingan Hasil Prediksi Penjualan Dengan Nilai Sebenarnya.

Gambar 3 menunjukkan perbandingan antara nilai aktual (*Actual Data*) dan nilai prediksi (*Prediction Data*) yang dihasilkan oleh model *Random Forest Regression*. Sumbu horizontal merepresentasikan data aktual, sementara sumbu vertikal menunjukkan data prediksi. Titik-titik pada grafik merepresentasikan pasangan nilai aktual dan prediksi untuk setiap data uji. Terlihat bahwa sebagian besar titik berada sangat dekat dengan garis diagonal, yang menunjukkan bahwa nilai prediksi sangat mendekati nilai aktual. Hal ini mengindikasikan bahwa model memiliki tingkat akurasi yang tinggi dalam memprediksi target. Sebaran yang konsisten sepanjang garis diagonal juga mencerminkan bahwa model bekerja dengan baik di berbagai rentang data. Kesimpulannya, model *Random Forest Regression* yang

digunakan sangat andal dalam melakukan prediksi penjualan supermarket. Selain itu visualisasi *features important* juga diperlukan untuk mengetahui seberapa besar pengaruh masing-masing *feature* variabel independen terhadap variabel dependen.



Gambar 4. Grafik *Feature Importance* Dari Model.

Gambar 4 menunjukkan pentingnya fitur atau kolom dalam memengaruhi prediksi model, yang diukur menggunakan metode *feature importance* dari algoritma *Random Forest Regression*. Pada grafik tersebut, sumbu horizontal merepresentasikan tingkat *importance* atau kontribusi setiap fitur terhadap prediksi, sedangkan sumbu vertikal menunjukkan nama fitur. Dari grafik, terlihat bahwa fitur dengan kontribusi terbesar terhadap prediksi adalah *Product line*, diikuti oleh *Gender* dan *Customer type*. Sementara itu, fitur seperti *Rating*, *cogs*, dan *gross income* memiliki kontribusi yang relatif kecil terhadap hasil prediksi. Informasi ini dapat digunakan untuk memahami faktor mana yang paling memengaruhi penjualan supermarket dan membantu dalam pengambilan keputusan yang lebih tepat.

3.2. Pengujian

Pengujian model dilakukan untuk mengevaluasi kinerja model dalam memprediksi data yang belum pernah dilihat sebelumnya. Berbagai metrik evaluasi, seperti OOB, MSE, dan R-squared, digunakan untuk mengukur akurasi, kesalahan prediksi, dan sejauh mana model dapat menjelaskan variabilitas dalam data, sehingga memberikan gambaran menyeluruh tentang kemampuan prediksi model yang dibangun.

Tabel 3. Kode Evaluasi Python

Kode	Deskripsi
<pre>from sklearn.metrics import mean_squared_error, mean_absolute_error oob_score = model.oob_score_ print(f'Out-of-Bag Score: {oob_score}')</pre>	<p>Mengimpor metrik evaluasi MSE, R-squared, dan MAE.</p> <p>Menampilkan <i>Out-of-Bag Score</i> untuk mengevaluasi performa model tanpa menggunakan data uji. Berguna jika ada data hilang pada set data pelatihan.</p>
<pre>mse = mean_squared_error(y_test, predik) print(f'Mean Squared Error: {mse}')</pre>	<p>Menghitung <i>Mean Squared Error</i> untuk mengukur perbedaan antara nilai prediksi dan nilai aktual dalam data uji.</p>

<code>r2 = r2_score(y_test, predik)</code> <code>print(f'R-squared: {r2}')</code>	Menghitung koefisien determinasi R-squared, yang menunjukkan seberapa baik model dapat menjelaskan variasi data.
<code>skor_error = mean_absolute_error(y_test, predik)</code> <code>print(f'Mean Absolute Error: {skor_error}')</code>	Menghitung <i>Mean Absolute Error</i> untuk mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual.

Evaluasi model yang dilakukan pada Tabel 3 menggunakan beberapa metrik untuk menilai kinerja prediksi *Random Forest Regressor*. Model menunjukkan *OOB Score* sebesar 0.9999, yang mencerminkan tingkat keakuratan model dalam memprediksi sampel yang tidak digunakan dalam pelatihan, memberikan indikasi kemampuan generalisasi yang sangat baik. Metrik MSE sebesar 2.4899, yang menunjukkan rata-rata kuadrat kesalahan prediksi yang kecil. Selanjutnya, *R-squared* atau koefisien determinasi mencapai 0.9999 (mendekati 1), mengindikasikan bahwa model mampu menjelaskan hampir semua variabilitas data dalam variabel target. Terakhir, MAE bernilai 0.9305, yang menggambarkan kesalahan absolut rata-rata dari prediksi, juga relatif kecil. Hasil ini menunjukkan bahwa model *Random Forest Regressor* memiliki performa yang sangat baik dalam memprediksi total penjualan berdasarkan data yang tersedia.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Random Forest Regression sangat efektif dalam memprediksi penjualan supermarket. Proses *preprocessing* data, seperti penanganan data numerik yang hilang dengan nilai rata-rata (*mean*) dan data kategori dengan nilai modus, telah berhasil meningkatkan kualitas data untuk pelatihan model. Dengan memanfaatkan 13 fitur yang relevan dan membagi data menjadi 80% data latih serta 20% data uji, model menunjukkan performa yang sangat baik. Evaluasi kinerja menghasilkan nilai *Out-of-Bag* (OOB) sebesar 99,99%, *Mean Squared Error* (MSE) sebesar 2,4899, *R-squared* mendekati 1 (0,99996), serta *Mean Absolute Error* (MAE) hanya 0,9305. Hasil ini menunjukkan bahwa model memiliki akurasi yang sangat tinggi, *error* yang rendah, serta kemampuan generalisasi yang optimal terhadap data baru. Algoritma ini terbukti mampu memanfaatkan data transaksi supermarket untuk menghasilkan prediksi penjualan yang andal dan akurat, sehingga mendukung pengelolaan operasional dengan lebih baik. Keandalan model ini menunjukkan potensinya dalam membantu pengambilan keputusan berbasis data di sektor ritel khususnya dalam memprediksi penjualan supermarket.

4.2. Saran

Sebagai langkah untuk mendukung pengembangan penelitian, disarankan mencoba algoritma lain seperti *Gradient Boosting*, *Support Vector Regression*, atau *Neural Networks*. Perbandingan dengan algoritma tersebut dapat membantu mengidentifikasi metode terbaik untuk dataset yang digunakan. Selain itu, penting untuk menguji model pada dataset yang lebih besar dan bervariasi, seperti data dari periode waktu yang lebih panjang atau transaksi dari cabang supermarket lainnya. Memperluas cakupan data ini tidak hanya meningkatkan kemampuan generalisasi model, tetapi juga memberikan wawasan yang lebih mendalam serta tantangan yang lebih bermakna bagi algoritma *Random Forest Regression*. Selanjutnya, menambahkan fitur-fitur relevan seperti informasi promosi, musim atau hari libur, dan tren penjualan dapat membantu model mengenali pola yang lebih kompleks sehingga menghasilkan prediksi yang lebih akurat dan kompleks.

Daftar Pustaka

- [1] A. R. Hapriyanto, "Multidisciplinary Science Strategi Inovatif dalam Meningkatkan Daya Saing Bisnis di Era Digital," *Nusant. J. Multidiscip. Sci.*, vol. 2, no. 1, pp. 108–117, 2024.
- [2] H. Herman, I. Riadi, and Y. Kurniawan, "Vulnerability Detection With K-Nearest Neighbor and Naïve Bayes Method using Machine Learning," *Int. J. Artif. Intell. Res.*, vol. 7, no. 1, p. 10, 2023, doi: 10.29099/ijair.v7i1.795.
- [3] R. Nisrina and C. Devina, "Analisis Penerapan Sistem Informasi Akuntansi Persediaan Pada Supermarket Suzuya Di Rantau Prapat," *J. Ekon. Manaj. Dan Bisnis*, vol. 1, no. 3, pp. 103–112, 2024.
- [4] V. Rahayu, "Analisis Algoritma Apriori dan FP-Growth Dalam Menemukan Pola Frequent Item

-
- Data Association Rule Pada Supermarket,” *Explore*, vol. 11, no. 2, p. 20, 2021, doi: 10.35200/explore.v11i2.436.
- [5] F. N. Adnan, “Optimasi Analisis Peramalan dengan Metode Regresi Weighted Moving Average,” *JOINS (Journal Inf. Syst.)*, vol. 4, no. 2, pp. 119–128, 2019, doi: 10.33633/joins.v4i2.2265.
- [6] I. Darmayanti, P. Subarkah, L. R. Anunggilarsa, and J. Suhaman, “Prediksi Potensi Siswa Putus Sekolah Akibat Pandemi Covid-19 Menggunakan Algoritme K-Nearest Neighbor,” *JST (Jurnal Sains dan Teknol.)*, vol. 10, no. 2, pp. 230–238, 2021.
- [7] A. Muzakir and R. A. Wulandari, “Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree,” *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016, doi: 10.15294/sji.v3i1.4610.
- [8] D. Yulianto and A. S. Nugraheni, “Efektivitas Pembelajaran Daring Dalam Pembelajaran Bahasa Indonesia,” *Decod. J. Pendidik. Teknol. Inf.*, vol. 1, no. 1, pp. 33–42, 2021, doi: 10.51454/decode.v1i1.5.
- [9] M. Aditya Pratama, M. Munawaroh, W. Joko Pranoto, P. Studi Teknik Informatika, F. Sains dan Teknologi, and U. Muhammadiyah Kalimantan Timur, “Perbandingan Performa Algoritma Linear Regresi dan Random Forest untuk Prediksi Harga Bawang Merah di Kota Samarinda,” *J. Ilmu Tek.*, vol. 1, no. 2, pp. 172–182, 2024, [Online]. Available: <https://doi.org/10.62017/tekonik>
- [10] H. Tantyoko, D. K. Sari, and A. R. Wijaya, “Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection,” *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 83–89, 2023, doi: 10.36080/idealis.v6i2.3036.
- [11] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, “Prediksi Penjualan Kendaraan Menggunakan Regresi Linear : Studi Kasus pada Industri Otomotif di Indonesia,” vol. 4, pp. 1199–1216, 2024, [Online]. Available: <https://j-innovative.org/index.php/Innovative%0APrediksi>
- [12] Malik, H. Hidayat, A. H. Muhammad, and Kusnawi, “Penerapan Algoritma Monte Carlo Untuk Memprediksi IPS Dan IPK Berdasarkan Karakteristik Mahasiswa Perguruan Tinggi X Di Kota Cirebon,” vol. 2, no. 4, pp. 81–96, 2024.
- [13] G. A. Sandag, “Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest,” *CogITO Smart J.*, vol. 6, no. 2, pp. 167–178, 2020, doi: 10.31154/cogito.v6i2.270.167-178.
- [14] E. Fitri, “Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah,” *J. Appl. Comput. Sci. Technol.*, vol. 4, no. 1, pp. 58–64, 2023, doi: 10.52158/jacost.v4i1.491.
- [15] N. L. W. S. R. Ginantra *et al.*, *Data Mining dan Penerapan Algoritma*. 2021.
- [16] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmadden, and L. Efrizoni, “Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 273–281, 2024, doi: 10.57152/malcom.v4i1.1085.
- [17] I. P. Teddy, D. Wijaya, and I. Bagus, “Uji Performansi Algoritma LR dan RFR pada Implementasi Sistem Prediksi Harga Rumah,” vol. 1, no. 2016, pp. 917–924, 2023.
- [18] R. J. Beckman and S. Weisberg, *Applied Linear Regression*, vol. 29, no. 1. 1987. doi: 10.2307/1269895.
- [19] A. Hurifiani, A. Irma Purnamasari, and I. Ali, “Penerapan Algoritma Regresi Linear Untuk Prediksi Penjualan Alat Tulis Kantor (Atk) Di Bumdes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 266–273, 2024, doi: 10.36040/jati.v8i1.8305.