

Perbandingan Random Forest dan XGBoost Untuk Prediksi Penjualan Produk E-Commerce Rumah Madu

Destaria Hayatunnisa¹, Permata^{2*}, Adhie Thyo Priandika², Rakhmat Dedi Gunawan²

¹ Fakultas Teknik dan Ilmu Komputer, Program Studi Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

² Fakultas Teknik dan Ilmu Komputer, Program Studi Informatika, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹destaria_hayatunnisa@teknokrat.ac.id, ^{2*}permata@teknokrat.ac.id, ³adhie_thyo@teknokrat.ac.id,

⁴rakhmatdedig@teknokrat.ac.id

Email Penulis Korespondensi: permata@teknokrat.ac.id

Submitted: 08/10/2025; Accepted: 06/12/2025; Published: 08/12/2025

Abstrak—Pengelolaan stok merupakan salah satu tantangan utama bagi UMKM, termasuk Rumah Madu di Bandar Lampung, di mana persediaan madu sering kali ditentukan berdasarkan perkiraan tanpa perhitungan yang tepat. Penelitian ini bertujuan menganalisis dan membandingkan performa algoritma *Random Forest* dan *XGBoost* dalam memprediksi penjualan produk Rumah Madu agar pengelolaan stok lebih terukur. Dataset terdiri dari 1.699 transaksi penjualan madu yang telah melalui proses *cleaning*, transformasi fitur, dan *standardisasi*. Variabel yang digunakan mencakup jenis madu, harga per unit, hari, bulan, status hari libur, dan indikator promosi. Pemodelan dilakukan menggunakan pendekatan *time-series split* dengan data historis sebagai data latih dan data terbaru sebagai data uji. Hasil evaluasi menunjukkan bahwa *Random Forest* menghasilkan MAE sebesar 24,35, RMSE 29,04, dan R^2 -0,9685, sedangkan *XGBoost* menghasilkan MAE 25,50, RMSE 30,58, dan R^2 -1,1825. Nilai R^2 negatif menunjukkan bahwa kedua model belum mampu menjelaskan variasi data secara optimal, bahkan performanya masih di bawah baseline sederhana. Meskipun demikian, hasil *feature importance* mengungkap bahwa variabel harga per unit dan jenis madu merupakan faktor dominan yang memengaruhi penjualan. Penelitian ini menegaskan pentingnya pengembangan model lebih lanjut melalui optimasi parameter dan peningkatan kualitas data agar akurasi prediksi dapat ditingkatkan.

Kata Kunci: Prediksi Penjualan; Random Forest; XGBoost; UMKM; Feature Importance

Abstract—Inventory management is one of the main challenges for small and medium enterprises (SMEs), including Rumah Madu in Bandar Lampung, where honey stock levels are often determined based on estimation rather than precise calculation. This study aims to analyze and compare the performance of the Random Forest and XGBoost algorithms in predicting honey sales to achieve more measurable stock management. The dataset consists of 1,699 honey sales transactions that have undergone cleaning, feature transformation, and standardization processes. The variables used include honey type, unit price, day, month, holiday status, and promotion indicators. Modeling was conducted using a time-series split approach, where historical data served as the training set and recent data as the testing set. The evaluation results show that Random Forest achieved an MAE of 24.35, RMSE of 29.04, and R^2 of -0.9685, while XGBoost achieved an MAE of 25.50, RMSE of 30.58, and R^2 of -1.1825. The negative R^2 values indicate that both models were unable to explain data variation optimally, with performance falling below a simple baseline. Nevertheless, the feature importance analysis revealed that unit price and honey type were the dominant factors influencing sales. This study highlights the need for further model development through parameter optimization and improved data quality to enhance prediction accuracy.

Keywords: Sales Prediction; Random Forest; XGBoost; SMEs; Feature Importance

1. PENDAHULUAN

Madu yang dihasilkan dari aktivitas perlebaran termasuk kategori hasil hutan non-kayu yang memiliki nilai ekonomi tinggi. Di Indonesia, keberadaannya memiliki peluang besar untuk dikembangkan karena didukung oleh iklim tropis serta ketersediaan hutan yang menopang ekosistem budidaya lebah. Selain sebagai produk pangan alami, madu juga memiliki potensi pasar yang luas dan stabil, sehingga banyak UMKM memanfaatkannya sebagai komoditas utama usaha. Salah satunya adalah UMKM Rumah Madu di Kota Bandar Lampung yang berfokus pada penjualan berbagai jenis madu. Namun, dalam praktiknya, usaha ini menghadapi kendala dalam pengelolaan stok dan perencanaan penjualan yang masih dilakukan secara manual dan berdasarkan perkiraan. Oleh karena itu, penelitian ini berfokus pada penerapan metode prediksi penjualan menggunakan algoritma pembelajaran mesin untuk membantu UMKM dalam mengoptimalkan manajemen stok dan meningkatkan efektivitas penjualan produk madu[1].

Seiring derasnya arus digitalisasi, UMKM semakin menyadari bahwa data adalah harta yang sangat bernilai. Digitalisasi yang berkembang begitu cepat telah mengguncang banyak bidang, terutama ranah bisnis. Tren penjualan, pola pasar, hingga perilaku konsumen kini bisa ditangkap lebih jelas lewat kemampuan mengumpulkan, mengolah, dan memanfaatkan data dengan tepat[2]. Dalam persaingan bisnis, umkm-umkm berlomba menghadirkan inovasi sekaligus menyempurnakan produk mereka demi memikat konsumen. Salah satu pelaku industri tersebut adalah Rumah Madu, umkm yang berfokus pada penjualan hasil madu[3].

Berlokasi di Kota Bandar Lampung, tepatnya di Jl. ZA. Pagar Alam No.45, berdiri Rumah Madu yang kini berhadapan dengan persoalan manajemen stok. Persediaan madu sering kali baru dipesan saat hampir habis, karena jumlah kebutuhan hanya ditentukan berdasarkan perkiraan pemilik, bukan lewat perhitungan yang jelas. Kondisi ini menunjukkan lemahnya sistem kontrol pada pengelolaan bahan baku, sehingga menjadi hambatan utama dalam menjalankan usaha penjualan madu tersebut[4].

Dalam menjalankan sebuah usaha, kegiatan prediksi punya peran yang tidak bisa diabaikan. Bagi pemilik bisnis, prediksi menjadi alat bantu penting untuk mengambil keputusan yang lebih tepat, sekaligus berpeluang mendorong peningkatan penjualan produk. Jika dilakukan dengan akurat, prediksi mampu menekan risiko kerugian karena barang menumpuk di gudang atau stok tak terjual, serta menjadikan operasional lebih efisien. Demi mencapai ketepatan tersebut, beragam metode dan teknik telah diuji serta dikembangkan hingga saat ini[5].

Dalam ranah klasifikasi maupun regresi, *Random Forest* dan *XGBoost* dikenal sebagai estimator yang tangguh dengan performa unggul. Kedua algoritma ini memiliki kelebihan mampu mengurangi risiko *overfitting*, sekaligus tetap memberikan prediksi yang cukup tinggi meski dihadapkan pada data hilang maupun distribusi data yang tidak seimbang. *Random Forest* dikenal sebagai salah satu metode klasifikasi sekaligus prediksi yang berakar dari pendekatan CART (*Classification and Regression Trees*). Teknik ini memanfaatkan konsep bagging atau bootstrap aggregating yang dipadukan dengan random feature selection, sehingga hasil prediksi menjadi lebih stabil sekaligus akurat. Di sisi lain, hadir pula *eXtreme Gradient Boosting (XGBoost)*, yakni algoritma yang lahir dari pengembangan gradient tree boosting. Dengan mengandalkan ensemble learning, *XGBoost* terbukti mampu bekerja secara efisien, terutama dalam menangani permasalahan machine learning yang berukuran besar [6].

Berdasarkan penelitian terdahulu yang memprediksi Harga Rumah Dengan *Random Forest*, *Gradient Boosting*, dan *XGBoost* menunjukkan bahwa *XGBoost* memberikan performa terbaik dengan MSE sebesar $1,56 \times 1014$ rupiah², R2 sebesar 0,7746, dan RMSE sekitar 12,5 juta rupiah. *XGBoost* dan *Random Forest* lebih konsisten dibanding *Gradient Boosting*, khususnya pada harga rumah tinggi[7]. Sementara itu pada penelitian yang berjudul Prediksi Tinggi Gelombang Laut di Perairan Semarang – Demak dengan Menggunakan *Random Forest* dan *XGBoost* menunjukkan bahwa untuk variabel SWH, model *XGBoost* secara konsisten memberikan performa yang lebih akurat dibandingkan *Random Forest*, dengan nilai koefisien determinasi (R^2) mencapai 0.9989 dan nilai galat sangat rendah (RMSE = 0.0099 m; MAE = 0.0043 m; MAPE = 1.6%)[8].

Pada Penelitian yang berjudul Analisis *XGBoost* dan *Random Forest* untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla menunjukkan bahwa *XGBoost* memberikan prediksi dengan MSE 14.982,45, RMSE 122,40, dan MAPE 12,35%. *Random Forest* menghasilkan MSE 15.876,21, RMSE 126,00, dan MAPE 13,28% dan Model Gabungan (*weighted average XGBoost 0,6 + Random Forest 0,4*) menunjukkan performa terbaik dengan MSE 14.125,78, RMSE 118,85, dan MAPE 11,87% yang dapat membantu pengambil kebijakan dan pihak terkait dalam perencanaan mitigasi karhutla[9]. Penelitian berikutnya dengan judul Prediksi Keberhasilan Menindaklanjuti Pelanggan pada Dealer Mobil dengan Komparasi Algoritma *Random Forest* dan *XGBoost* menunjukkan *XGBoost* unggul dengan akurasi 91,67%, lebih baik dibanding *Random forest* dengan akurasi 88,89%. Kedua model menunjukkan performa yang seimbang pada kelas positif dan negatif, menandai peningkatan signifikan dari pendekatan sebelumnya[10]. Selain itu, penelitian yang berjudul Analisis Perbandingan Model *Random Forest* dan *XGBoost* dalam Memprediksi Turnover Karyawan menunjukkan bahwa *XGBoosts* sedikit mengungguli *Random Forest*, dengan mencapai akurasi 98,83% dibandingkan 98,80%, yang menunjukkan ketangguhannya dalam tugas prediktif[11].

Tujuan dari penelitian ini adalah untuk menganalisis serta membandingkan performa algoritma *Random Forest* dan *XGBoost* dalam memprediksi penjualan produk Rumah Madu di Bandar Lampung. Penelitian ini diharapkan dapat menghadirkan model prediksi yang akurat dan efisien sehingga mampu membantu pemilik usaha dalam mengelola stok madu secara lebih terukur. Secara teoretis, penelitian ini memberikan kontribusi dalam pengembangan ilmu data mining dan pembelajaran mesin, khususnya pada penerapan algoritma *ensemble learning* dalam konteks prediksi penjualan produk UMKM. Sementara secara praktis, hasil penelitian ini diharapkan dapat menjadi acuan bagi pelaku UMKM dalam memanfaatkan teknologi prediktif untuk mendukung pengambilan keputusan bisnis, mengoptimalkan persediaan, serta meminimalkan risiko kehabisan stok atau penumpukan produk yang tidak terjual[12].

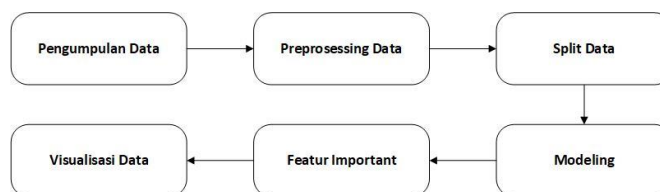
2. METODOLOGI PENELITIAN

Penelitian ini menggunakan data penjualan madu dari UMKM Rumah Madu di Bandar Lampung untuk membangun model prediksi penjualan. Tahapan penelitian meliputi pengumpulan dan pembersihan data, transformasi fitur, pembagian data latih dan uji, pemodelan dengan algoritma *Random Forest* dan *XGBoost*, serta evaluasi menggunakan metrik MAE, RMSE, dan R^2 . Seluruh proses dilakukan menggunakan Python dengan pustaka *Scikit-learn* dan *XGBoost*. Hasil penelitian diharapkan membantu UMKM dalam mengelola stok dan meningkatkan akurasi perencanaan penjualan.

2.1 Tahapan Penelitian

Penelitian ini dilakukan dengan memanfaatkan dataset penjualan madu yang diambil dari UMKM Rumah Madu sebagai sumber data utama. penelitian ini diawali dengan tahap pengumpulan dan persiapan data yang digunakan sebagai dasar dalam proses analisis. Setelah data terkumpul, dilakukan analisis awal berupa perhitungan pendapatan guna memperoleh gambaran umum mengenai pola data yang dimiliki. Tahap berikutnya adalah preprocessing data, yaitu proses pembersihan, normalisasi, dan penanganan data yang hilang agar dataset siap digunakan dalam pemodelan. Setelah data diproses, penelitian dilanjutkan pada tahap prediksi dan evaluasi dengan menerapkan algoritma yang dipilih untuk menilai akurasi model. Hasil dari prediksi tersebut kemudian dianalisis lebih lanjut

melalui *feature importance* untuk mengetahui variabel yang paling berpengaruh dalam model. Terakhir, dilakukan visualisasi hasil agar temuan penelitian dapat disajikan secara lebih jelas dan mudah dipahami. Gambaran lebih rinci terkait metode penelitian ini dapat ditemukan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Data penelitian ini diperoleh dari pencatatan transaksi penjualan madu yang telah tersimpan dalam bentuk tabel Excel diperoleh sebanyak 1700 data. Dataset memuat beberapa variabel penting seperti tanggal transaksi, jumlah terjual, harga per unit, jenis madu, serta total jumlah yang dihasilkan. Informasi tersebut menjadi bahan utama untuk dianalisis lebih lanjut[13].

2.3 Preprocessing Data

Data *pre-processing* merupakan serangkaian tahapan yang dilakukan terhadap data sebelum digunakan dalam proses analisis maupun pemodelan. Tujuan utama dari tahap ini adalah meningkatkan kualitas hasil model dengan melakukan pembersihan data, misalnya memperbaiki atau menghapus data yang rusak maupun tidak relevan. Dengan adanya *pre-processing*, data dapat dipastikan lebih siap dan optimal untuk diolah oleh model machine learning[14]. Tahapan data preparation pada penelitian ini meliputi pembersihan data dengan konversi format harga, transformasi tanggal, serta penghapusan nilai kosong yang didapat sebanyak 1.699 data bersih[15].

2.4 Pembagian Data

Dalam penelitian ini, proses pembagian data dilakukan dengan pendekatan berbasis *time-series*. Dataset dipisahkan berdasarkan periode waktu, di mana data sebelum tanggal 1 Agustus 2024 digunakan sebagai data latih, sedangkan data mulai bulan Agustus 2024 hingga akhir periode penelitian digunakan sebagai data uji. Strategi ini dipilih agar proses pengujian model lebih mencerminkan kondisi nyata, mengingat prediksi penjualan dilakukan berdasarkan urutan waktu. Dengan demikian, baik algoritma *Random Forest* maupun *XGBoost* dilatih menggunakan data historis, kemudian diuji menggunakan data pada periode terbaru untuk menilai performa model[16].

2.5 Pemodelan Data

Penelitian ini bertujuan untuk membangun model prediksi yang dapat membantu UMKM Rumah Madu dalam mengelola stok madu secara lebih terukur. Pada tahap pemodelan, digunakan dua algoritma pembelajaran mesin, yaitu *Random Forest* dan *XGBoost*. Data latih yang digunakan merupakan bagian dari dataset historis sebelum tanggal 1 Agustus 2024, sedangkan data uji diambil dari periode setelah tanggal tersebut. Pembagian ini dilakukan dengan pendekatan *time-series split*, sehingga model dilatih menggunakan data masa lalu dan diuji dengan data terbaru untuk mengevaluasi kinerjanya[17].

2.5.1 Pemodelan *Random Forest*

Random Forest merupakan salah satu algoritma Machine Learning berbasis *Ensemble Learning* yang digunakan dalam proses klasifikasi. Algoritma ini bekerja dengan membangun sejumlah *Decision Tree* secara bersamaan, kemudian hasil prediksi dari setiap pohon digabungkan melalui perhitungan rata-rata untuk memperoleh prediksi akhir yang lebih akurat[18]. Rumus *Random Forest* dapat dilihat pada Persamaan berikut.

$$\text{Regresi: } \hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (1)$$

Rumus 1 menjelaskan bahwa nilai prediksi akhir (\hat{y}) didapat dengan menjumlahkan seluruh hasil prediksi dari masing-masing pohon ($T_i(x)$) terhadap data input x , kemudian dibagi dengan jumlah total pohon (N). Dengan kata lain, setiap pohon memberikan kontribusi berupa hasil prediksi, lalu semua hasil tersebut dirata-ratakan untuk menghasilkan output yang lebih stabil dan akurat.

$$\text{Klasifikasi: } \hat{y} = \text{mode} \{T_1(x), T_2(x), \dots, T_N(x)\} \quad (2)$$

Rumus 2 menunjukkan bahwa hasil prediksi akhir (\hat{y}) diperoleh dengan melihat kelas yang paling sering dipilih oleh seluruh pohon keputusan ($T_i(x)$) dalam hutan. Setiap pohon memberikan "suara" terhadap suatu kelas, lalu kelas dengan jumlah suara terbanyak akan ditetapkan sebagai prediksi akhir. Dengan metode ini, *Random Forest* mampu meminimalkan bias dari pohon individu dan menghasilkan klasifikasi yang lebih handal, karena keputusan didasarkan pada gabungan banyak model, bukan hanya satu.

2.5.2 Pemodelan XGBoost

Extreme Gradient Boosting (XGBoost) merupakan algoritma *machine learning* berbasis *ensemble* yang digunakan dalam penelitian ini untuk memprediksi jumlah penjualan madu. Dengan memanfaatkan data latih yang terdiri dari variabel-variabel seperti harga per unit, jenis madu, promo, hari, bulan, dan status hari libur, *XGBoost* berusaha menemukan parameter optimal (θ) yang mampu menggambarkan hubungan antara fitur-fitur tersebut dengan jumlah madu yang terjual. Melalui proses pelatihan bertahap, algoritma ini membangun sekumpulan pohon keputusan yang secara berurutan memperbaiki kesalahan prediksi model sebelumnya, sehingga dihasilkan prediksi yang lebih akurat terhadap variabel target, yaitu jumlah penjualan[19]. Rumus *XGBoost* dapat dilihat pada Persamaan 3.

$$Obj(0) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^K \Omega(f_x) \quad (3)$$

Fungsi objektif *XGBoost* terdiri dari dua bagian utama. Pertama, fungsi *loss* ($l(y_i, \hat{y}_i)$) yang mengukur selisih antara nilai aktual (jumlah madu terjual) dengan hasil prediksi. Kedua, regularisasi $\Omega(f_x)$ yang mengendalikan kompleksitas model agar tidak overfitting. Dengan cara ini, *XGBoost* mampu menghasilkan prediksi yang lebih akurat dan tetap stabil pada data baru.

2.6 Evaluasi Model

Proses evaluasi dilakukan dengan menerapkan beberapa metrik pengukuran kinerja model, yaitu *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan koefisien determinasi (*R-squared* atau R^2). Ketiga metrik ini digunakan untuk menilai tingkat akurasi model serta menentukan model yang paling optimal dalam melakukan prediksi[20]. Rumus evaluasi model dapat dilihat pada Persamaan berikut.

2.6.1 Mean Squared Error (MSE)

MSE mengukur rata-rata kuadrat selisih antara nilai aktual (y_i) dengan nilai prediksi (\hat{y}_i). Nilai MSE yang lebih kecil menunjukkan bahwa model menghasilkan prediksi yang lebih mendekati data sebenarnya[21]. Rumus dapat dilihat pada Persamaan 4.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

Mean Squared Error (MSE) dihitung dengan menjumlahkan kuadrat selisih antara nilai aktual (y_i) dan nilai prediksi \hat{y}_i , kemudian dibagi dengan jumlah data (n).

2.6.2 Root Mean Squared Error (RMSE)

RMSE merupakan akar kuadrat dari MSE. Metode ini memberikan gambaran besarnya kesalahan dalam satuan yang sama dengan data asli, sehingga lebih mudah diinterpretasikan. Nilai RMSE yang rendah berarti model lebih akurat dalam memprediksi[22]. Rumus dapat dilihat pada Persamaan 5.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

Root Mean Squared Error (RMSE) merupakan akar kuadrat dari MSE, sehingga hasil perhitungannya kembali pada satuan data aslinya.

2.6.3 R-squared (R^2)

R^2 atau koefisien determinasi digunakan untuk menilai seberapa baik model dapat menjelaskan variasi data. Nilainya berkisar antara 0 sampai 1, di mana nilai mendekati 1 menunjukkan model mampu menjelaskan sebagian besar variasi data dengan baik[23]. Rumus dapat dilihat pada Persamaan 6.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

R-squared R^2 atau koefisien determinasi, membandingkan selisih kuadrat antara nilai aktual dan prediksi dengan total variasi data. Nilai R^2 mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data, sedangkan nilai mendekati 0 menandakan model kurang baik dalam memprediksi.

2.6.4 Feature Importance

Feature importance merupakan metode yang digunakan untuk mengetahui seberapa besar pengaruh masing-masing variabel independen terhadap variabel dependen dalam suatu model. Nilai *feature importance* membantu peneliti memahami variabel mana yang paling berkontribusi terhadap hasil prediksi. Dalam algoritma seperti Random Forest dan XGBoost, nilai pentingnya fitur dihitung berdasarkan seberapa besar suatu fitur mampu mengurangi kesalahan prediksi pada setiap pohon keputusan. Dengan mengetahui fitur yang paling berpengaruh, analisis dapat difokuskan pada variabel yang memberikan dampak terbesar terhadap kinerja model, sehingga hasil prediksi menjadi lebih akurat dan interpretatif[24].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pada tahap pengumpulan data, penelitian ini memanfaatkan dataset penjualan madu yang bersumber dari pencatatan transaksi penjualan yang telah tersimpan dalam bentuk tabel Excel dengan jumlah total sebanyak 1.700 data transaksi. Dataset ini memuat beberapa variabel penting seperti tanggal transaksi, jumlah madu terjual, harga per unit, jenis madu, serta total penjualan yang dihasilkan dari setiap transaksi. Proses pengumpulan data dilakukan secara sistematis dengan mengambil data penjualan selama periode waktu tertentu di Rumah Madu. Data tersebut mencerminkan aktivitas penjualan berbagai jenis madu yang dipasarkan, seperti Madu Assalamah, Madu Kids Juara, Madu Batuk, dan Madu Raw Honey Apis mellifera. Setiap transaksi dicatat secara rinci agar dapat digunakan sebagai bahan analisis yang akurat dan relevan dalam penelitian ini. Selain itu, data yang diperoleh juga berfungsi sebagai dasar untuk tahap pengolahan dan analisis berikutnya, seperti *data preprocessing*, eksplorasi data, hingga pemodelan algoritma prediksi penjualan. Dengan adanya data penjualan yang lengkap dan terstruktur, peneliti dapat mengidentifikasi pola pembelian konsumen, tren harga, serta perbandingan performa penjualan antar jenis madu. Data penjualan madu tersebut disajikan pada Tabel 1.

Tabel 1. Data Penjualan Madu

Tanggal Transaksi	Jumlah Terjual	Harga Perunit	Jenis Madu	Jumlah
21/09/2025	4	Rp. 40.000	Madu Assalamah 8 in	160.000
19/09/2025	1	Rp. 100.000	Madu Kids Juara	100.000
18/09/2025	1	Rp. 100.000	Madu Kids Juara	100.000
17/09/2025	1	Rp. 75.000	Madu Angkak	75.000
13/09/2025	2	Rp. 100.000	Madu Raw Honey Apis mellifera	200.000
03/09/2025	1	Rp. 40.000	Madu Batuk	40.000
03/09/2025	1	Rp. 40.000	Madu Assalamah 8 in	40.000
31/08/2025	2	Rp. 40.000	Madu Assalamah 8 in	80.000
30/08/2025	3	Rp. 40.000	Madu Batuk	120.000
30/08/2025	1	Rp. 100.000	Madu Raw Honey Apis mellifera	100.000

3.2 Preprocessing Data

Pada tahap *preprocessing*, dilakukan proses *cleaning data* untuk memastikan kualitas dataset sebelum dianalisis lebih lanjut. Pertama, dilakukan pembersihan pada atribut harga dengan menghilangkan karakter non-numerik sehingga nilai harga dapat dikonversi ke bentuk numerik. Selanjutnya, atribut jumlah terjual diubah menjadi format numerik agar dapat diproses secara konsisten. Kolom tanggal transaksi juga dikonversi ke dalam format *datetime* sehingga dapat digunakan dalam analisis berbasis waktu. Setelah itu, baris data yang tidak valid, kosong, atau memiliki jumlah penjualan nol dihapus dari dataset. Tahapan ini menghasilkan data yang lebih bersih, terstruktur, dan siap digunakan pada proses pemodelan. Berisi hasil implementasi aplikasi ataupun hasil program (yang penting saja), ataupun hasil dari pengujian metode.

3.2.1 Cleaning Data

Setelah dilakukan proses *cleaning*, data yang sebelumnya masih mengandung nilai tidak konsisten berhasil diubah menjadi lebih terstruktur. Kolom harga telah dikonversi ke dalam bentuk numerik sehingga dapat diproses dengan mudah, sementara kolom jumlah terjual juga sudah dipastikan dalam format numerik yang valid. Selain itu, kolom tanggal transaksi berhasil diubah ke dalam format *datetime* agar dapat digunakan pada analisis berbasis waktu. Baris data yang kosong, tidak valid, atau memiliki jumlah penjualan nol juga dihapus. Hasil akhir dari tahap *preprocessing* ini menghasilkan sebanyak 1.699 baris data bersih yang siap digunakan pada tahap analisis dan pemodelan berikutnya. Hasil *Cleaning Data* dapat dilihat pada Gambar 2.

```

Nama kolom: Index(['tanggal transaksi', 'jumlah terjual', 'harga perunit', 'jenis madu',
                  'jumlah'],
                  dtype='object')

Hasil Cleaning Data:
  tanggal_transaksi  jumlah_terjual  harga_perunit \
1      21/09/2025           4.0      Rp40.000
2      19/09/2025           1.0     Rp100.000
3      18/09/2025           1.0     Rp100.000
4      17/09/2025           1.0     Rp75.000
5      13/09/2025           2.0     Rp100.000
6       03/09/2025           1.0     Rp40.000
7       03/09/2025           1.0     Rp40.000
8      31/08/2025           2.0     Rp40.000
9      30/08/2025           3.0     Rp40.000
10     30/08/2025           1.0     Rp100.000

  jenis_madu  jumlah  harga_perunit  jumlah_terjual \
1  Madu Assalamah 8 in  160.000      40000.0           4.0
2  Madu Kids Juara    100.000     100000.0           1.0
3  Madu Kids Juara    100.000     100000.0           1.0
4  Madu Angkak        75.000      75000.0           1.0
5  Madu Raw Honey Apis mellifera  200.000    100000.0           2.0
6  Madu Batuk         40.000      40000.0           1.0
7  Madu Assalamah 8 in  40.000      40000.0           1.0
8  Madu Assalamah 8 in  80.000      40000.0           2.0
9  Madu Batuk        120.000     40000.0           3.0
10 Madu Raw Honey Apis mellifera  100.000    100000.0           1.0
    
```

Gambar 2. Hasil Cleaning data

3.2.2 Feature Engineering

Hasil *feature engineering* pada penelitian ini menghasilkan beberapa variabel baru yang digunakan untuk memperkaya informasi dalam dataset. Variabel tambahan tersebut meliputi *dayofweek* untuk menunjukkan hari terjadinya transaksi, *month* untuk mengidentifikasi bulan penjualan, *holiday* sebagai penanda transaksi pada akhir pekan, serta *promo_flag* yang merepresentasikan kondisi harga produk di bawah kuartil 25% sebagai indikasi adanya promosi. Selain itu, variabel *jenis_madu* juga telah dibersihkan agar siap diproses lebih lanjut sebagai data kategorik. Dengan adanya penambahan fitur ini, dataset yang berjumlah 1.699 entri dengan 13 kolom menjadi lebih informatif, sehingga diharapkan dapat meningkatkan kemampuan model dalam memprediksi pola penjualan madu berdasarkan faktor harga, waktu, maupun jenis produk. Hasil *Feature Engineering* dapat dilihat pada Gambar 3.

```

Hasil Feature Engineering (5 baris pertama):
  tanggal transaksi  jumlah terjual  harga perunit  \
0      01/01/2023      57.0      Rp170.000
1      02/01/2023      62.0      Rp100.000
2      03/01/2023      58.0      Rp100.000
3      04/01/2023       9.0      Rp40.000
4      05/01/2023      60.0      Rp40.000

      jenis_madu  jumlah  harga_perunit  jumlah_terjual  \
0  Madu Hutan Apis mellipera 1 KG  9.690.000      170000.0      57.0
1      Madu Kids Juaraa  6.200.000      100000.0      62.0
2      Madu Kids Juaraa  5.800.000      100000.0      58.0
3      Madu Assalamah 8 in  360.000      40000.0       9.0
4      Madu Assalamah 8 in  2.400.000      40000.0      60.0

      tanggal  dayofweek  month  holiday  promo_flag  \
0  2023-01-01         6      1      1         0
1  2023-01-02         0      1      0         0
2  2023-01-03         1      1      0         0
3  2023-01-04         2      1      0         1
4  2023-01-05         3      1      0         1

      jenis_madu
0  Madu Hutan Apis mellipera 1 KG
1      Madu Kids Juaraa
2      Madu Kids Juaraa
3      Madu Assalamah 8 in
4      Madu Assalamah 8 in

Ringkasan Dataset (info kolom):
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1699 entries, 0 to 1698
Data columns (total 13 columns):
#   Column              Non-Null Count  Dtype
---  -
0   tanggal transaksi  1699 non-null   object
1   jumlah terjual     1699 non-null   float64
2   harga perunit      1699 non-null   object
3   jenis_madu         1699 non-null   object
4   jumlah             1699 non-null   object
5   harga_perunit      1699 non-null   float64
6   jumlah_terjual     1699 non-null   float64
    
```

Gambar 3. Hasil *Feature Engineering*

3.2.3 Pipeline

Hasil *pipeline preprocessing* menunjukkan bahwa data telah berhasil diubah dari bentuk awal yang masih terdiri atas enam fitur menjadi lebih kaya dengan total tiga puluh tiga fitur. Transformasi ini dilakukan melalui proses *one-hot encoding* pada variabel kategorik, yaitu *jenis_madu*, sehingga setiap jenis madu direpresentasikan dalam bentuk kolom biner (0 atau 1). Dengan cara ini, model dapat membedakan setiap kategori madu secara numerik tanpa kehilangan makna kategorinya. Selain itu, variabel numerik seperti harga per unit, *dayofweek*, *month*, *holiday*, dan *promo_flag* telah melalui proses standarisasi menggunakan *StandardScaler*, sehingga nilai-nilainya terdistribusi dalam skala yang seragam. Proses ini penting untuk menghindari dominasi satu fitur terhadap fitur lainnya dan memastikan bahwa semua variabel memiliki pengaruh yang seimbang dalam proses pelatihan model. Selain transformasi dan standarisasi, tahap *pipeline preprocessing* juga mencakup pembersihan data dari nilai kosong atau anomali yang dapat memengaruhi hasil analisis, serta penyesuaian format data agar sesuai dengan kebutuhan algoritma yang digunakan. Dengan hasil ini, dataset menjadi lebih siap untuk dimodelkan oleh algoritma *machine learning* karena seluruh fitur telah berada dalam bentuk yang optimal dan konsisten untuk analisis prediktif. Tahapan ini juga berperan penting dalam meningkatkan efisiensi proses pelatihan model, mengurangi risiko kesalahan perhitungan, dan memperkuat keakuratan hasil prediksi yang dihasilkan. Hasil *Pipeline* secara keseluruhan dapat dilihat pada Gambar 4

```

Shape X sebelum preprocessing: (1699, 6)
Shape X setelah preprocessing : (1699, 33)
Hasil Preprocessing (5 baris pertama):
jenis_madu_Madu Al-syifa jenis_madu_Madu Angkak \
0 0.0 0.0
1 0.0 0.0
2 0.0 0.0
3 0.0 0.0
4 0.0 0.0

jenis_madu_Madu Angkak Trombofit jenis_madu_Madu Apis Trigona \
0 0.0 0.0
1 0.0 0.0
2 0.0 0.0
3 0.0 0.0
4 0.0 0.0

jenis_madu_Madu Assalamah jenis_madu_Madu Assalamah 8 in \
0 0.0 0.0
1 0.0 0.0
2 0.0 0.0
3 0.0 1.0
4 0.0 1.0

jenis_madu_Madu Batuk jenis_madu_Madu Dorsata jenis_madu_Madu Gholibah \
0 0.0 0.0 0.0
1 0.0 0.0 0.0
2 0.0 0.0 0.0
3 0.0 0.0 0.0
4 0.0 0.0 0.0

jenis_madu_Madu Hitam ... jenis_madu_Madu Sarang jenis_madu_Madu Sehat \
0 0.0 ... 0.0 0.0
1 0.0 ... 0.0 0.0
2 0.0 ... 0.0 0.0
3 0.0 ... 0.0 0.0
4 0.0 ... 0.0 0.0

jenis_madu_Madu Syifa Kids jenis_madu_Madu Zuriat \
0 0.0 0.0
1 0.0 0.0
2 0.0 0.0
3 0.0 0.0
4 0.0 0.0

jenis_madu_Madu gholibah harga_perunit dayofweek month holiday \
0 0.0 1.040027 1.474729 -1.581287 1.570779
1 0.0 -0.049161 -1.515800 -1.581287 -0.636627
2 0.0 -0.049161 -1.017378 -1.581287 -0.636627
3 0.0 -0.982752 -0.518957 -1.581287 -0.636627
4 0.0 -0.982752 -0.020535 -1.581287 -0.636627

promo_flag
0 -0.694657
1 -0.694657
2 -0.694657
3 1.439560
4 1.439560
    
```

Gambar 4. Hasil pipeline preprocessing

3.3 Pembagian Data

Hasil pembagian data menunjukkan bahwa dataset telah dipisahkan menjadi data latih (train) dan data uji (test) berdasarkan periode waktu tertentu. Data latih terdiri dari 1.062 baris dengan rentang tanggal mulai 1 Januari 2023 hingga 31 Juli 2024, sedangkan data uji terdiri dari 637 baris dengan periode mulai 1 Agustus 2024 hingga 21 September 2025. Pembagian ini dilakukan dengan pendekatan time series split, di mana data historis digunakan untuk melatih model, sedangkan data terbaru digunakan untuk menguji performa model. Dari contoh data yang ditampilkan, baik data latih maupun data uji memiliki atribut yang sama, yaitu harga per unit, hari dalam minggu (*dayofweek*), bulan, status hari libur, promo flag, serta jenis madu. Hal ini memastikan bahwa struktur data pada train dan test tetap konsisten sehingga model dapat belajar dari pola pada data latih dan kemudian diuji pada data terbaru dengan kondisi serupa. Berisi hasil pembahasan dan bisa perbandingan dari hasil penelitian sebelumnya. Hasil Pembagian data dapat dilihat pada Gambar 5.

```

Train: (1062, 6) Test: (637, 6)

Periode Train: 2023-01-01 00:00:00 s.d 2024-07-31 00:00:00
Periode Test : 2024-08-01 00:00:00 s.d 2025-09-21 00:00:00

Contoh data Train:
harga_perunit dayofweek month holiday promo_flag \
0 170000.0 6 1 1 0
1 100000.0 0 1 0 0
2 100000.0 1 1 0 0

jenis_madu
0 Madu Hutan Apis mellipera 1 KG
1 Madu Kids Juaraa
2 Madu Kids Juaraa

Contoh data Test:
harga_perunit dayofweek month holiday promo_flag jenis_madu
1062 100000.0 3 8 0 0 Madu Kids Juaraa
1063 100000.0 3 8 0 0 Madu Kids Juaraa
1064 175000.0 3 8 0 0 Madu Apis Trigona
    
```

Gambar 5. Hasil Pembagian Data

3.4 Pemodelan Klasifikasi

Hasil modeling menunjukkan bahwa performa kedua algoritma, yaitu *Random Forest* dan *XGBoost*, masih belum optimal dalam memprediksi data penjualan madu. Pada model *Random Forest*, nilai MAE (*Mean Absolute Error*) sebesar 24,23 dan RMSE (*Root Mean Squared Error*) sebesar 28,57, dengan nilai R^2 sebesar -0,90. Sementara itu, model *XGBoost* menghasilkan MAE sebesar 25,42, RMSE sebesar 30,61, dan R^2 sebesar -1,19. Nilai MAE dan RMSE yang relatif tinggi menunjukkan bahwa terdapat perbedaan yang cukup besar antara nilai prediksi dan nilai aktual. Selain itu, nilai R^2 yang negatif pada kedua model mengindikasikan bahwa model tidak mampu menjelaskan variabilitas data dengan baik, bahkan performanya lebih buruk dibandingkan dengan model sederhana yang hanya menggunakan rata-rata sebagai prediksi. Dengan demikian, diperlukan langkah perbaikan seperti tuning parameter, penambahan fitur yang lebih relevan, atau metode praproses data yang lebih baik agar model dapat memberikan hasil prediksi yang lebih akurat. Hasil prediksi dapat dilihat pada Tabel 2.

Tabel 2. Hasil Prediksi *Random Forest* dan *XGBoost*

	Random Forest	XGBoost
MEA	24.23	25.42
RMSE	28.57	30.61
R^2	-0.90	-1.19

3.5 Evaluasi Model Klasifikasi

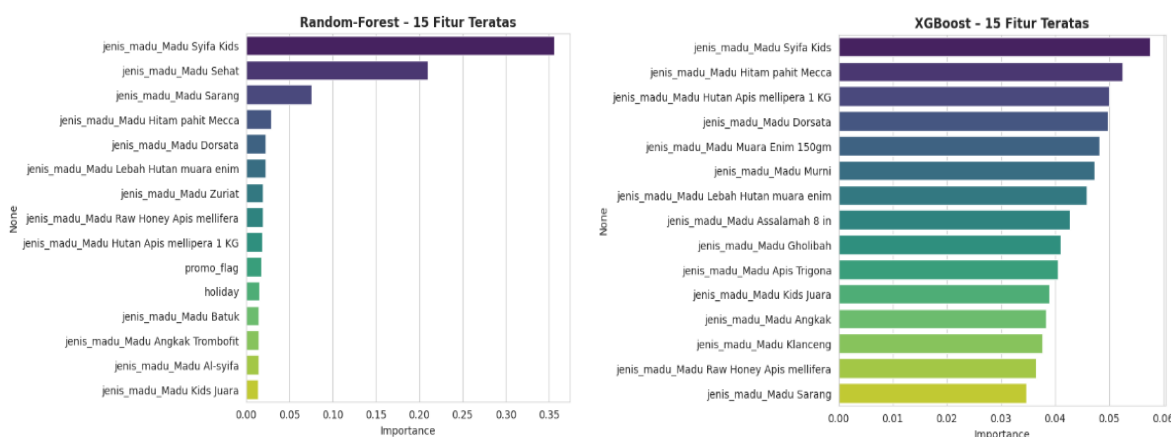
Berdasarkan hasil evaluasi model, baik *Random Forest* maupun *XGBoost* masih menunjukkan performa yang kurang baik dalam memprediksi data, yang terlihat dari nilai R^2 negatif pada keduanya. Model *Random Forest* menghasilkan nilai MAE sebesar 24,35 dan RMSE sebesar 29,04 dengan $R^2 = -0,9685$, sedangkan model *XGBoost* memiliki MAE sebesar 25,50, RMSE sebesar 30,58, dan $R^2 = -1,1825$. Nilai MAE dan RMSE yang lebih kecil pada *Random Forest* dibandingkan dengan *XGBoost* menandakan bahwa *Random Forest* sedikit lebih akurat dalam memprediksi dibandingkan *XGBoost*. Namun, keduanya masih belum mampu menjelaskan variasi data dengan baik, karena nilai R^2 negatif berarti model lebih buruk dibandingkan prediksi sederhana menggunakan rata-rata.

Tabel 3. Hasil Evaluasi Model *Random Forest* dan *XGBoost*

	Random Forest	XGBoost
MEA	24.23	25,50
RMSE	28.57	30,58
R^2	-0.905	-1,1825

3.6 Visualisasi Data

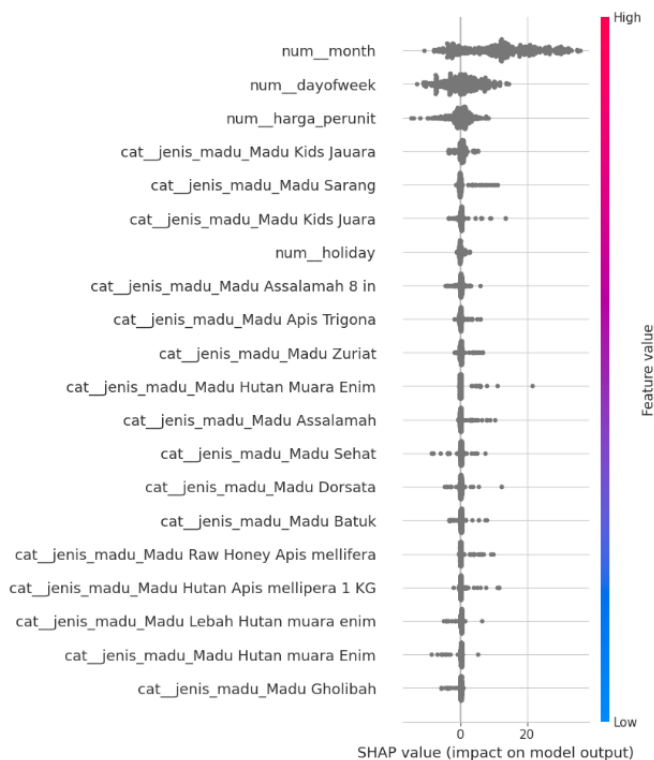
Visualisasi data digunakan untuk memahami pola, tren, dan hubungan antarvariabel secara lebih jelas. Pada penelitian ini, visualisasi membantu mengenal karakteristik data, mendeteksi outlier, serta menganalisis keterkaitan antarvariabel. Jenis visualisasi yang digunakan meliputi histogram, boxplot, diagram batang, scatter plot, dan heatmap korelasi, serta *feature importance* dari model *Random Forest* dan *XGBoost* untuk menilai variabel paling berpengaruh. Dengan demikian, visualisasi mempermudah analisis sekaligus memudahkan penyampaian temuan secara intuitif.



Gambar 6. Hasil *Feature Importance*

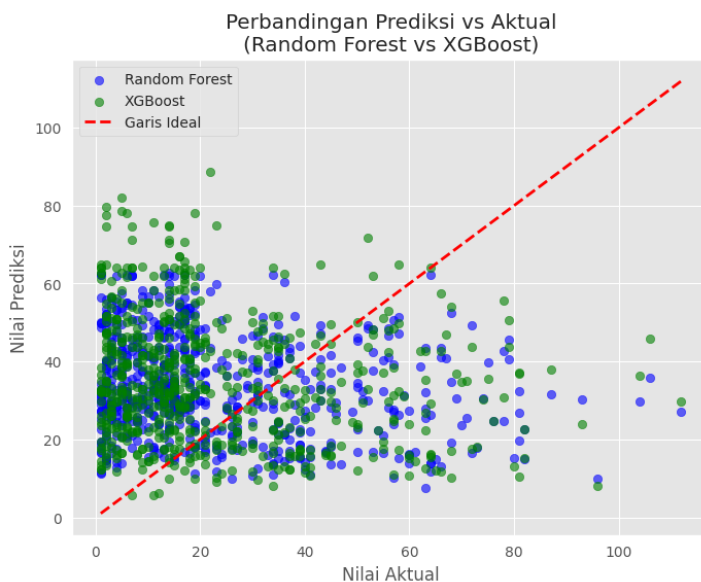
Gambar 6 menunjukkan hasil analisis *feature importance* yang menggambarkan besarnya pengaruh setiap fitur terhadap hasil prediksi penjualan madu. Berdasarkan hasil tersebut, faktor jenis madu menjadi variabel yang paling berpengaruh terhadap model. Pada algoritma *Random Forest*, penjualan madu sangat dipengaruhi oleh beberapa jenis madu tertentu seperti Madu Syifa Kids dan Madu Sehat. Sementara itu, pada algoritma *XGBoost*, kontribusi fitur

terlihat lebih merata dengan jenis madu seperti Syifa Kids, Hitam Pahit Mecca, dan Dorasta yang menunjukkan nilai penting yang cukup tinggi. Selain faktor jenis madu, fitur lain seperti *harga per unit*, *promo flag*, serta atribut waktu seperti *month* dan *holiday* juga memiliki pengaruh meskipun relatif lebih kecil. Hasil ini mengindikasikan bahwa jenis madu merupakan faktor utama yang menentukan tingkat penjualan, sedangkan faktor lain hanya berperan sebagai pendukung dalam memengaruhi keputusan pembelian konsumen.



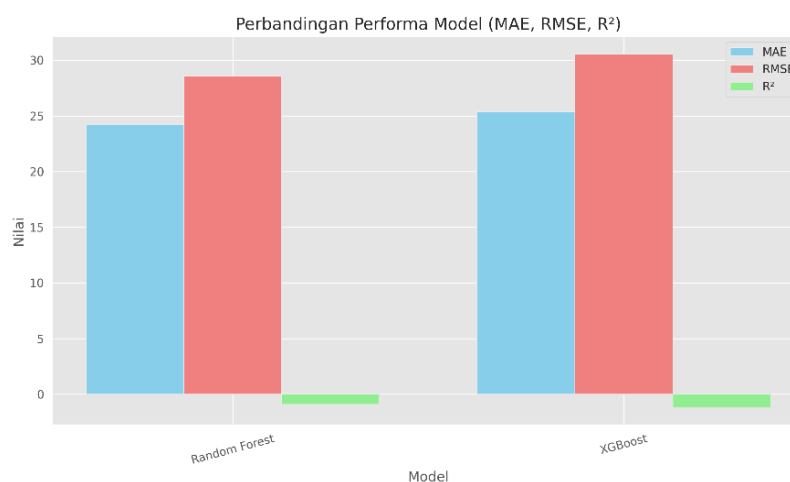
Gambar 7. Hasil SHAP Value

Pada Gambar 7 merupakan hasil visualisasi SHAP menunjukkan kontribusi setiap fitur terhadap prediksi model. Terlihat bahwa fitur *num_month*, *num_dayofweek*, dan *num_harga_perunit* memiliki pengaruh paling dominan dalam memengaruhi hasil prediksi. Nilai SHAP yang positif mengindikasikan bahwa fitur tersebut mendorong prediksi ke arah lebih tinggi, sedangkan nilai negatif mendorong ke arah lebih rendah. Warna pada grafik merepresentasikan nilai asli dari fitur, di mana warna merah menandakan nilai tinggi dan biru menandakan nilai rendah. Dengan demikian, grafik ini membantu menjelaskan bahwa faktor waktu (bulan dan hari) serta harga per unit madu menjadi variabel utama yang berkontribusi terhadap variasi hasil prediksi penjualan, sementara jenis madu tertentu memberikan pengaruh tambahan namun relatif lebih kecil.



Gambar 8. Hasil Perbandingan Prediksi vs Aktual

Pada Gambar 8 hasil grafik menunjukkan perbandingan antara nilai aktual dan hasil prediksi dari dua model, yaitu Random Forest (titik biru) dan XGBoost (titik hijau), dengan garis merah putus-putus sebagai garis ideal di mana prediksi sama persis dengan nilai aktual. Jika model bekerja dengan baik, sebaran titik seharusnya mengikuti garis merah tersebut. Namun, pada grafik terlihat bahwa sebagian besar titik menyebar cukup jauh dari garis ideal, yang menandakan bahwa kedua model masih memiliki kesalahan prediksi yang cukup besar. Random Forest cenderung memberikan prediksi yang lebih terkonsentrasi pada rentang tengah (sekitar 20–60), sehingga terlihat lebih stabil, sementara XGBoost menghasilkan prediksi yang lebih bervariasi, bahkan hingga nilai ekstrem, sehingga lebih rentan menghasilkan kesalahan besar. Selain itu, kedua model tampak sering melakukan underestimation pada nilai aktual yang tinggi, karena banyak prediksi berada di bawah garis ideal ketika nilai aktual semakin besar. Secara keseluruhan, baik Random Forest maupun XGBoost masih belum mampu memberikan prediksi yang mendekati nilai sebenarnya secara konsisten, sehingga diperlukan perbaikan seperti tuning hyperparameter atau feature engineering agar performa model lebih optimal.



Gambar 9. Grafik Perbandingan Performa Model (MEA, RMSE, R²)

Gambar 8 memperlihatkan perbandingan performa model Random Forest dan XGBoost berdasarkan tiga metrik evaluasi, yaitu MAE (*Mean Absolute Error*), RMSE (*Root Mean Square Error*), dan R² (*Coefficient of Determination*). Nilai MAE dan RMSE menunjukkan rata-rata besar kesalahan prediksi, sedangkan R² menunjukkan seberapa baik model mampu menjelaskan variasi data. Berdasarkan grafik, model Random Forest memiliki nilai MAE dan RMSE yang lebih rendah dibandingkan XGBoost, sehingga dapat disimpulkan bahwa Random Forest menghasilkan prediksi yang relatif lebih akurat dengan kesalahan lebih kecil. Sementara itu, nilai R² pada kedua model terlihat negatif, yang berarti kedua model belum mampu menjelaskan hubungan variabel dengan baik, bahkan performanya lebih buruk dibandingkan baseline sederhana. Namun, secara komparatif, Random Forest masih menunjukkan performa yang lebih baik dibandingkan XGBoost dalam penelitian ini.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa performa algoritma Random Forest dan XGBoost dalam memprediksi penjualan produk UMKM Rumah Madu di Bandar Lampung masih menunjukkan keterbatasan. Hasil evaluasi menunjukkan bahwa Random Forest menghasilkan nilai MAE sebesar 24,35, RMSE sebesar 29,04, dan R² sebesar -0,9685, sementara XGBoost menghasilkan MAE sebesar 25,50, RMSE sebesar 30,58, dan R² sebesar -1,1825. Nilai R² negatif pada kedua model menandakan bahwa kemampuan kedua algoritma dalam menjelaskan variasi data penjualan madu masih kurang optimal, bahkan lebih buruk dibandingkan prediksi sederhana menggunakan rata-rata. Meskipun demikian, Random Forest secara komparatif memberikan prediksi yang lebih stabil dan memiliki kesalahan yang lebih kecil dibandingkan XGBoost, sehingga cenderung lebih akurat dalam kondisi dataset saat ini. Analisis feature importance dan SHAP menunjukkan bahwa faktor jenis madu menjadi variabel paling dominan dalam memengaruhi penjualan, diikuti oleh variabel harga per unit, hari, dan bulan transaksi. Visualisasi prediksi versus data aktual juga menegaskan bahwa kedua model masih sering melakukan underestimation pada nilai penjualan yang tinggi, serta sebaran prediksi cenderung menyimpang dari garis ideal, menunjukkan perlunya penyempurnaan model. Keterbatasan penelitian ini terletak pada ukuran dataset yang relatif terbatas, fitur yang mungkin belum sepenuhnya representatif, serta model yang belum dilakukan tuning hyperparameter secara mendalam. Untuk penelitian selanjutnya, disarankan untuk menambah jumlah data, melakukan eksplorasi fitur tambahan yang relevan, serta melakukan optimasi parameter model agar prediksi dapat lebih akurat dan mampu mendukung pengelolaan stok madu secara lebih efektif. Dengan demikian, meskipun Random Forest menunjukkan performa sedikit lebih baik, penelitian ini menekankan pentingnya pengembangan model lebih lanjut untuk mendukung pengambilan keputusan UMKM Rumah Madu secara lebih tepat dan efisien.

REFERENCES

- [1] D. Sarah, R. N. Suryana, dan K. Kirbrandoko, “Strategi Bersaing Industri Madu (Studi Kasus: CV Madu Apiari Mutiara),” *J. Apl. Bisnis dan Manaj.*, vol. 5, no. 1, hal. 71–83, 2019, doi: 10.17358/jabm.5.1.71.
- [2] G. Ramadhan, A. Faqih, dan F. Sandy Eka Permana, “Analisis Tren Penjualan Menu Seafood Dengan Algoritma Random Forest Untuk Meningkatkan Strategi Pemasaran,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 4, hal. 5942–5949, 2025, doi: 10.36040/jati.v9i4.13911.
- [3] B. W. D. Nugroho, N. J. K. Jakti, M. A. N. Rochman, dan A. J. Nugroho, “Analisis Pengendalian Kualitas Produk Gula Dan Biaya Kualitas Dalam Menunjang Efektivitas Produksi (Studi Kasus: PT Madu Baru Pg Madukismo),” *J. Teknol. dan Manaj. Ind. Terap.*, vol. 2, no. 2, hal. 72–81, 2023.
- [4] Y. M. N. S. W. I. F. Marisa, “Prediksi Penjualan Warung Kopi Oi Menggunakan Metode Random Forest Dan Xgboost,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, hal. 1–7, 2025.
- [5] M. S. E. S. A. K. Zyen, “Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk,” *Rekayasa Tek. Inform. dan Inf.*, vol. 5, no. September, hal. 1–9, 2024.
- [6] S. M. P. Negeri dan T. Barat, “Perbandingan Metode Random Forest Dan Xgboost Dalam Prediksi Harga Jual Rumah Di Wilayah Jabodetabek,” *J. Multidisiplin Saintek*, vol. 10, no. September, hal. 1–19, 2025, doi: 10.8734/Kohesi.v1i2.365.
- [7] D. P. Sari, Bety Wulan, “View of Analisis Perbandingan Prediksi Harga Rumah Dengan Random Forest, Gradient Boosting, dan XGBoost.pdf,” *Indones. J. Learn. Technol. Innov.*, vol. 04, no. juni, hal. 1–10, 2025.
- [8] G. Erutjahjo dan A. Supriyanto, “Jurnal Informatika : Jurnal pengembangan IT Prediksi Tinggi Gelombang Laut di Perairan Semarang – Demak dengan Menggunakan Random Forest dan XGBoost,” vol. 10, no. 4, hal. 869–881, 2025, doi: 10.30591/jpit.v10i4.9315.
- [9] D. Sangaji dan T. Sutabri, “Analisis XGBoost dan Random Forest untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla,” *J. Pustaka AI (Pusat Akses Kaji. Teknol. Artif. Intell.)*, vol. 5, no. 1, hal. 13–18, 2025, doi: 10.55382/jurnalpustakaai.v5i1.905.
- [10] H. N. Heidy *et al.*, “Prediksi Keberhasilan Menindaklanjuti Pelanggan pada Dealer Mobil dengan Komparasi Algoritma Random Forest dan XGBoost,” *J. Ilm. Komput.*, vol. 21, no. Agustus, hal. 933–939, 2025.
- [11] M. B. Setiawan dan A. Rahmatulloh, “Analisis Perbandingan Model Random Forest dan XGBoost dalam Memprediksi Turnover Karyawan,” *J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 15, no. 2, hal. 393–400, 2025, [Daring]. Tersedia pada: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [12] R. Nurhidayat dan N. Hendrastuty, “Analisis Sentimen Komentar Media Sosial Twitter Terhadap Tes CPNS dengan Algoritma Naive Bayes,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 3, hal. 1477–1489, 2024, doi: 10.47065/bits.v6i3.6148.
- [13] M. L. T. Alfianti dan R. Supriyanto, “Perbandingan Kinerja Algoritma Random Forest, AdaBoost, dan XGBoost Dalam Memprediksi Resiko Penyakit Osteoporosis,” *J. Ilmu Komput. dan Agri-Informatika*, vol. 11, no. 2, hal. 172–184, 2024, doi: 10.29244/jika.11.2.172-184.
- [14] R. Hidayat *et al.*, “Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket,” *Simkom*, vol. 10, no. 1, hal. 101–109, 2025, doi: 10.51717/simkom.v10i1.703.
- [15] M. B. Prayogi, F. Apriani, dan Nirma, “Prediksi Angka Harapan Hidup Menggunakan Random Forest dan XGBoost Regression,” *J. Inform. dan Teknol.*, vol. 2, no. 1, hal. 112–121, 2025.
- [16] H. A. N. E. W. Pamungkas, “Perbandingan Algoritma Machine Learning: Svm, Random Forest, Dan Xgboost Untuk Prediksi Stroke,” *J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, hal. 1098–1110, 2025.
- [17] A. Ananda Surya, D. Rizki Darmawan, dan A. Solichin, “Prediksi Kapabilitas Calon Debitur Menggunakan Analisis Data Machine Learning Dengan Metode Random Forest,” *J. Algoritm.*, vol. 22, no. 1, hal. 777–788, 2025, doi: 10.33364/algoritma/v.22-1.1929.
- [18] E. Danuarta dan D. Alita, “Perbandingan Algoritma Naive Bayes dan Random Forest untuk Melakukan Analisis Sentimen Cyberbullying Generasi Z Pada Twitter,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 4, hal. 2448–2458, 2025, doi: 10.47065/bits.v6i4.6909.
- [19] R. W. D. E. Ratnawati, “Peramalan Penjualan Produk Menggunakan Extreme Gradient Boosting (Xgboost) Dan Kerangka Kerja Crisp-Dm Untuk Pengoptimalan Manajemen Persediaan (Studi Kasus: Ub Mart),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 2, hal. 1–12, 2025.
- [20] I. G. Ayu, R. Astarani, dan I. G. Surya, “Analisis Perbandingan XGBoost dan LightGBM dalam Prediksi Penjualan Ritel Walmart Store Sales,” *J. Nas. Teknol. Inf. dan Apl.*, vol. 3, no. MI, hal. 717–728, 2025.
- [21] I. M. Sianturi dan D. Harinto, “Prediksi Penetapan Tarif Penerbangan Menggunakan Auto-MI Dengan Algoritma Random Forest,” *J. Sist. Informasi, Tek. Inform. dan Teknol. Pendidik.*, vol. 2, no. 1, hal. 40–48, 2022, doi: 10.55338/justikpen.v2i1.37.
- [22] A. Prayuda dan I. Pratama, “Prediksi Jumlah Kedatangan Wisatawan Mancanegara Di Indonesia Berdasarkan Pintu Masuk Kedatangan Udara,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 9, no. 2, hal. 232–241, 2024, doi: 10.36341/rabit.v9i2.4787.
- [23] A. Ruliff Brahmantyo, “Perbandingan Hasil Prediksi Harga Properti Di Daerah Brooklyn Menggunakan Metode XGBoost, Random Forest, dan Linear Regression,” *J. Komput. dan Inform.*, vol. 18, no. 2, hal. 124–132, 2023.
- [24] S. Fitriani, E. Budiman, M. Fadli, M. Surono, dan H. Sulistiani, “Optimalisasi Metode Random Forest menggunakan Particle Swarm Optimization dalam Prediksi Prestasi Mahasiswa,” *Pros. Semin. Nas. Teknol. Komput. dan Sains*, vol. 3, no. 1, hal. 406–415, 2025.